NEXT-GENERATION METRO NETWORK AND EDGE COMPUTING ARCHITECTURES

Peter Fetterolf, Ph.D.

EXECUTIVE SUMMARY

Cloud computing, edge computing, 5G, and metro networks are converging with the promise of offering new categories of consumer and enterprise services. Network operators expect to generate new revenues with edge services that could help accelerate growth throughout the decade. However, for next-generation 5G and edge services to become a success, network operators need to transform their networks and their operating models.

Many edge applications, such as Open RAN, MEC, and video CDN, have strict constraints around bandwidth, latency, and security. Network traffic is continuing to grow, and a significant share of traffic will be terminated in regional and metro edge compute nodes. This means that there will be fundamental shifts in metro network requirements and architectures. Metro topologies will include both ring and spine-leaf architectures; access and aggregation bandwidth will increase from 1 Gbps to 10/25/100/400 Gbps; network slicing will become a critical requirement in metro networks.

Future metro networks will need high levels of scalability, flexibility, and service intelligence baked into all components of the network and edge data centers.

Service and application intelligence are critical component of next-generation metro networks. End-to-end orchestration and automation are necessary to ensure network scalability, elasticity, and security. This paper explains how trends in 5G, edge computing, and metro network traffic are driving the requirements, topologies, and architecture of future metro networks.

Key Trends in Metro Networks and Edge Computing

Over the past decade cloud computing has disrupted information technology. Over the next decade we expect to see a similar disruption as cloud computing moves from regional data centers to edge and far edge data centers. There are three primary drivers for edge computing:

1. Applications and services require low latency that can only be delivered by edge computing solutions
2. Edge computing can reduce regional and national network traffic by moving processing to the edge
3. Applications require edge security

In the longer term we expect an ecosystem of edge services and applications to emerge around industrial IoT, remote medicine, connected vehicles, and other use cases. In the short term edge applications being deployed today are driving new network architectures. Some of the current edge applications are depicted in Figure 1 and described in the following paragraphs.

Video Caching and CDN Networks

Content delivery networks have been around for over 20 years, and video caching is a key technology that has enabled video streaming services. Video streaming continues to grow and with the advent of 5G mobile communications we expect mobile video streaming services on smart phones will continue to accelerate. Additionally, 4K and 8K ultra HD-TV will contribute to the growing needs of the network to exponentially increase bandwidth. The growth in demand for video streaming combined with the additional bandwidth required by 4K and 8K video means that video caches will need to move from regional data centers to metro edge data centers. Moving the video cache closer to the source of demand reduces the bandwidth required in regional and metro networks and improves video quality for end users. This is especially important as more high-quality video is delivered to smart phones across the 5G network.

4G/5G CUPS Architectures

Mobile packet core technology is used to control and manage mobile data traffic. Recently, the control plane and the user plane in the packet core has been separated using Control User Plane Separation (CUPS). Although CUPS is used in some 4G LTE networks, it is a standard part of 5G architecture. The User Plane Function (UPF) is the packet core function responsible for processing and forwarding IP packets in the mobile network. The requirement for edge applications and the uncertainty around network traffic patterns drives the need for distributing UPFs to the metro edge. This allows packets to terminate on edge computing nodes or be forwarded to internet peering points without needing to return to a regional data center where the packet core control plane is hosted. Service providers are deploying UPFs to the metro edge today, and we expect UPFs to be deployed further out to the edge as new edge applications emerge.

Open RAN and Cloud RAN Architectures

The emergence of Open RAN (O-RAN) and Cloud RAN (C-RAN) is transforming the requirements of metro networks. C-RAN networks centralize broadband baseband functions in edge nodes, which contain DU/CU pools. O-RAN and C-RAN architectures have significant impact on metro networks because they require fronthaul services, which have strict requirements for bandwidth, latency, and timing. Essentially, the radio is being split into two components; the network connecting these components must deliver radio signals that require high bandwidth, low latency, and synchronized timing. We expect 100GE or higher bandwidths are required for fronthaul, and network latency must be 100 microseconds or less. C-RAN could be one of the first latency-sensitive edge applications.

MEC Services

Multiaccess Edge Computing (MEC) is a category for emerging edge services that will be provided by service providers at the network edge. MEC nodes could coexist in edge data centers with C-RAN CU/DU servers or be placed at the customer's enterprise edge locations. MEC services are expected to include a variety of latency-sensitive edge applications. Examples of MEC applications are:

- IoT applications
- Real-time control applications for drones and robots
- Connected vehicle applications
- Cloud video gaming

MEC applications typically have requirements for low latency, security, and high bandwidth. These applications will drive requirements for bandwidth, latency, and QoS on metro networks.

Residential Network Bandwidth

In residential broadband networks we continue to see demand for bandwidth increasing. This is in part due to remote offices requiring more bandwidth and driving the demand to an all-time high in 2020 because of COVID. We expect these COVID trends of working remotely to continue after the pandemic is over. Fiber to the home and DOCSIS 3.1 are increasing the speed of home internet access to 1 Gbps and higher and driving more video, virtual meetings, gaming, and other high-bandwidth and latency-sensitive applications. Residential broadband will continue to drive the requirements for metro networks.

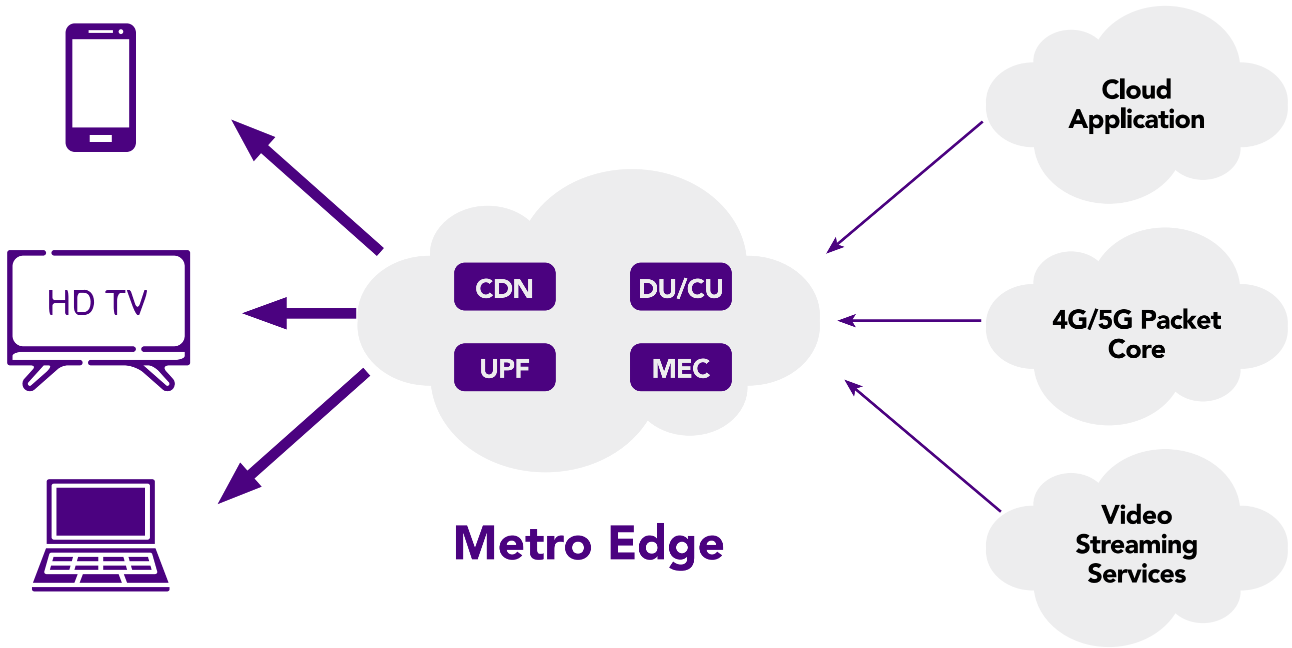

Figure 1. Metro Edge Applications

Changes in Network Traffic

The emergence of edge applications is having serious implications on network traffic patterns. More traffic is staying in the metro because of the distribution of virtual service instantiations across an increasingly distributed cloud architecture. This leads to more traffic terminating in far edge or regional edge data centers while at the same time traffic patterns are becoming more bursty and unpredictable. ACG Research has forecasted network traffic at the metro, regional, and national level. Specifically, we have forecasted mobile traffic per user and residential broadband traffic per household for metro, regional, and national networks. These monthly traffic projections are presented in Figure 2 and Figure 3. This analysis shows that over the next five years traffic will continue to grow, but metro traffic will grow at a higher rate than regional traffic, and regional traffic will grow at a higher rate than national traffic. This is because the key drivers of network traffic are the edge applications. These changing traffic dynamics have serious implications to metro edge network architectures.

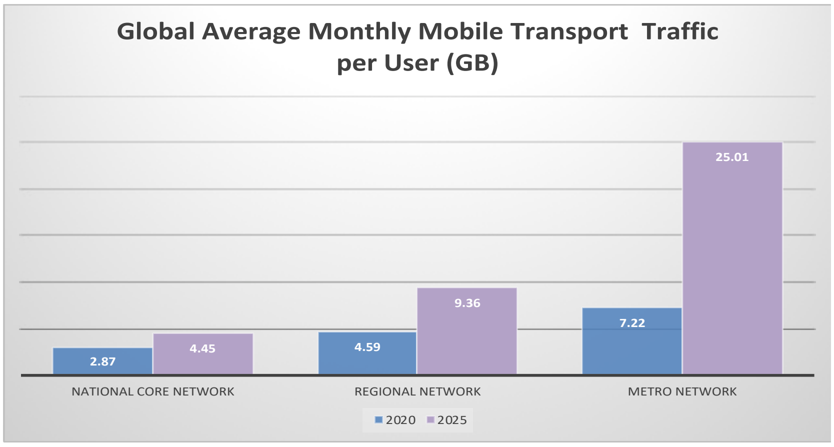

Figure 2. Global Average Monthly Mobile Traffic per User (GB)

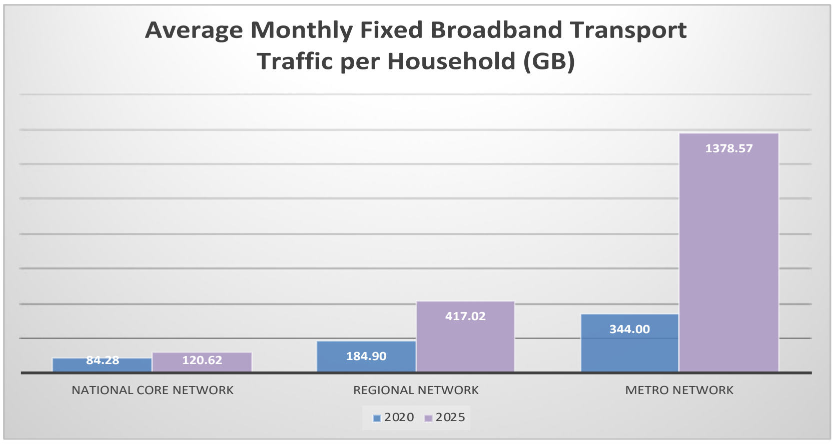

Figure 3. Average Monthly Fixed Broadband Traffic per Household (GB)

Implications for the Metro Network

The emergence of edge applications and services and the changing dynamics of network traffic are driving new requirements for metro networks. Future metro networks will need:

- High levels of scalability and flexibility
- New capacities and architectures
- Service and application intelligence
- End-to-end automation
- Convergence of legacy metro networks

Most metro networks are built using rings and are designed to carry traffic from access nodes, across an aggregation network, to regional data centers. Traffic is typically best-effort internet or MPLS business traffic. Traffic is sometimes throttled when monthly traffic volume limits are exceeded, but QoS is not typically applied at the application and service levels.

Future metro networks need to evolve from fixed, ridged networks to flexible, service-aware cloud networks. Metro networks need to have cloud-scale service agility. Edge data centers support resource pooling for multiple edge applications: vDU, vCU, UPF, MEC, and other edge applications. Networks need to have elastic scalability to support flexible edge resource pools. This elastic scalability means that new routers and network architectures with scalable bandwidth, real-time monitoring and control, and end-to-end automation must be implemented.

In addition to the demands of evolving edge applications, the demand for network bandwidth is continuing to increase, and most of that demand will result in the need for higher metro network capacity. For example:

- Residential broadband is growing to 1 Gbps and beyond
- Business services are growing from T1 speeds to 1 Gbps for branch offices and 10–100 Gbps for headquarters or data centers
- Mobile 5G service will require 10 Gbps for backhaul and 100 Gbps for fronthaul

Today, most access rings are 1 Gbps. Clearly, access rings need to grow to 10/25/100 Gbps, and aggregation nodes need to support 100–400 Gbps. This is a dramatic change from the metro networks in place today. To support this dramatic growth in traffic we see two types of architectures coexisting:

1. Metro ring networks
2. Metro spine-leaf networks

These network architectures are depicted in Figure 4. In some cases, rings will provide adequate bandwidth and scalability to meet emerging edge and traffic requirements; however, in some dense urban metro areas network traffic will drive the need for a metro network with a spine-leaf topology. The spine-leaf architecture provides a higher level of scalability, flexibility, and resiliency—an approach that has been used in cloud data centers for many years. These network topologies can and will coexist, and service providers need to use routers that have flexibility to support different network topologies.

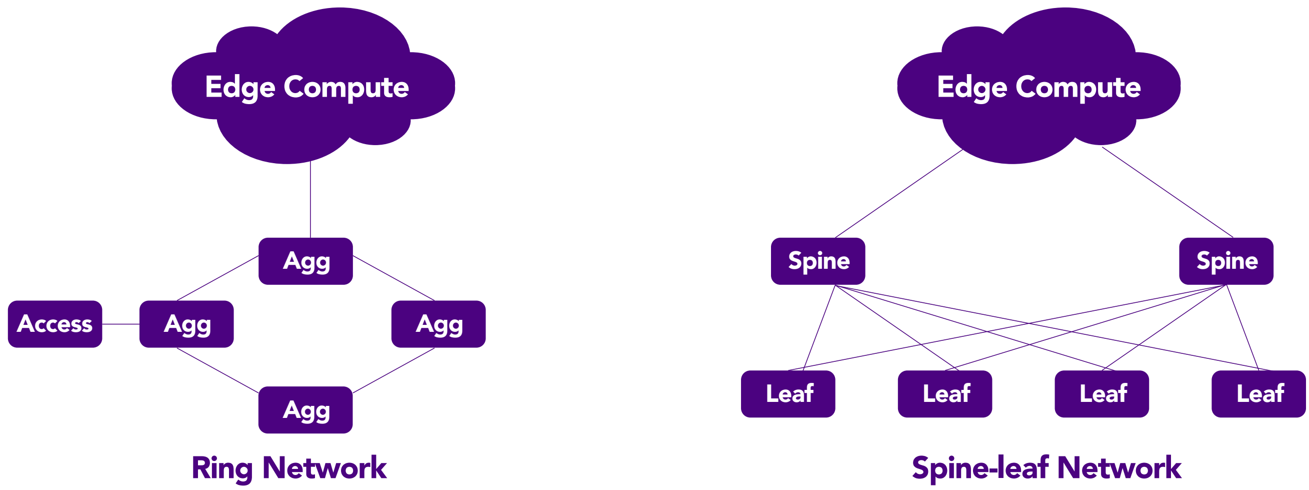

Figure 4. Two Network Topologies: Rings and Spine Leaf

In addition to supporting different network topologies, routers must be able to support network slicing, which requires a combination of VPNs and QoS for different services and applications (Figure 5). For example, fronthaul service between a radio unit on a cell site tower and a DU in a far edge node will require guaranteed bandwidth and latency. Similarly, real-time control applications used for robots and drones will also have strict latency requirements. Premium video traffic will need guaranteed bandwidth; many other internet applications will use best effort. Some of the key routing technologies required to support these requirements are:

- Segment routing
- IPV6
- L2VPN
- L3VPN
- Weighted fair queuing

5G RANs are architected and designed to support network slicing and QoS; however, the metro IP network must also be designed to support slicing end to end to guarantee end-to-end service level agreements (SLA).

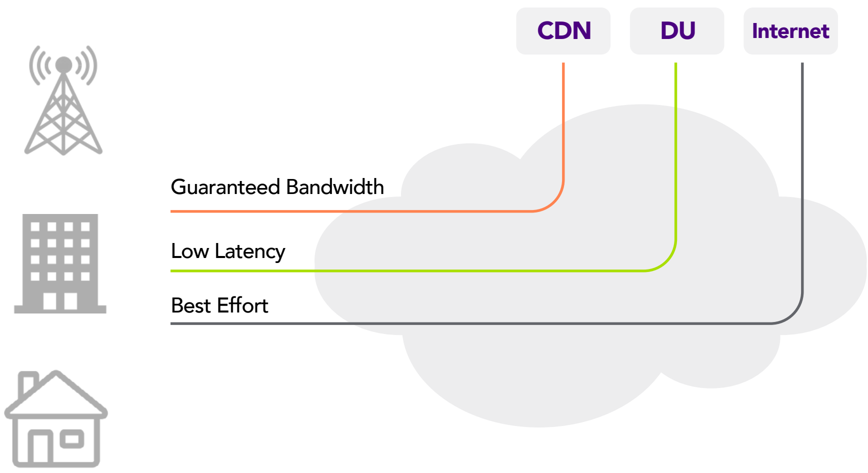

Figure 5. Network Slicing Architecture

Service and Application Intelligence

Network slicing also requires service provisioning, service assurance, and WAN control. Without service assurance it is impossible to guarantee SLAs for individual slices. This drives the requirement for end-to-end automation and service intelligence. Cloud computing could not work without service orchestration and automation. Future metro networks must have similar levels of orchestration and automation to support next-generation services and network slicing. Network orchestration must be integrated with edge data center orchestration such that services can be provisioned and managed end to end. A high-level view of a service and application intelligence architecture is presented in Figure 6. The key functions that need to be integrated using automation and network orchestration are service provisioning, service assurance, and WAN control.

Service provisioning is the process of instantiating a network service. Services could be end-customer services or network services provided to applications. Regardless of the nature of the service, it is critical to have an automated process of instantiating a service. New services require planning, design, deployment, and testing. Services can use multiple network resources that include network connections, QoS, bandwidth, and latency constraints. Services use data center resources that include virtual machines, containers, servers, and data center SDNs. Manual provisioning and service chaining of VNFs and network resources would be extremely labor intensive and prone to error. It is critical that future metro edge networks use automation and orchestration to automatically provision and test services end to end.

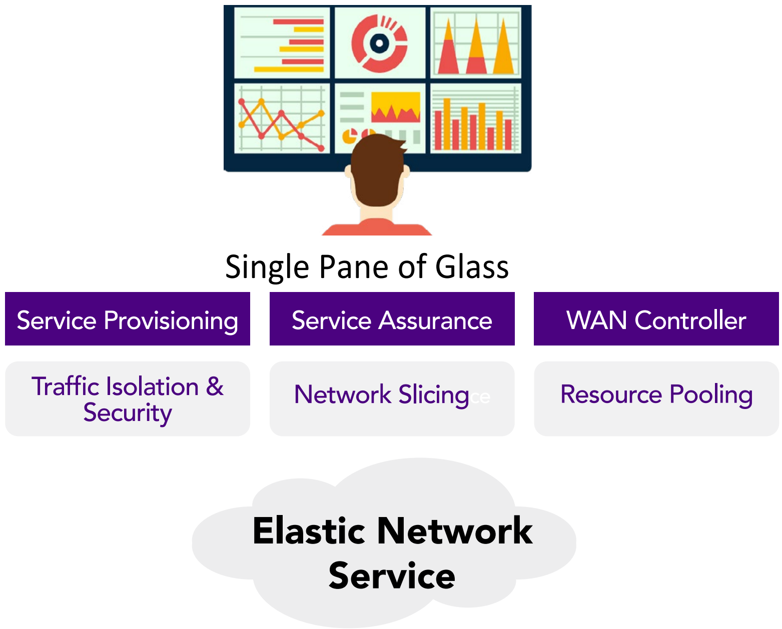

Figure 6. Service and Application Intelligence

Service assurance includes all activities required to validate QoS prior to activation, keeping the service running, and maintaining SLAs. This includes troubleshooting, repair, and performance management. Service assurance must also be applied to individual network slices to guarantee SLAs to end-customers. One of the problems with many service assurance solutions is that they do not measure the performance and quality of services in real time. A new paradigm has emerged that is known as active assurance. Unlike passive monitoring systems, active assurance uses active, synthetic traffic to verify application and service performance at the time of service delivery and throughout the life of the service. This is critical to ensuring a good user experience and meeting strict SLAs. Service and application intelligence combined with orchestration and automation are necessary tools for next-generation networks.

Conclusion and Summary

The emergence of 5G and edge computing will transform service offerings for network operators. This will drive a technology and architecture transformation in metro networks. Some of the key changes from legacy metro networks to next-generation networks are summarized in Table 1.

Legacy Metro Network	Next-Generation Metro Network
Ridged, fixed networks	Elastic networks with cloud-scale service agility
Ring topologies	Both ring and leaf-spine topologies
Mostly best effort with limited QoS	Network slicing with multiple services and VPNs; strict bandwidth, latency, and security constraints
Skilled network engineers designing and operating networks	Network automation assists engineers and reduces complexity
Silos and network element management systems and some managers of managers	Service assurance of slices, end-to-end orchestration, automation, and service chaining

Table 1. Legacy vs Next-Generation Metro Networks