www.pipelinepub.com
Volume 19, Issue 6

How Hyperscalers Fit in a 5G-Connected World

By: **Sameh Yaman**

With our cars, devices, cities, and utilities becoming ever more connected, hyperscalers have an increasingly important role in realizing the full potential of 5G and tied to that, open radio access network (RAN) (and continuing beyond 5G)—even if that's outside the traditional comfort zone of telecom operators.

The role of hyperscalers in 5G

Hyperscalers will be key to these models by providing the cloud infrastructure for which neutral hosts are used to support all network operations.

Technology is no longer accessed only through endpoints like computers or smartphones; it mediates everything we do. A little over a decade ago, technologists were increasingly talking up the Internet of Things (IoT) era—how small, connected devices could be embedded in everyday objects that we take for granted—and subsequently, these tiny machines grew by the billions.

A world with 175 zettabytes of data

During the COVID-19 crisis, technology was key to managing, tracking, and monitoring population health. It brought remote or hybrid ways of working into our everyday lives, and data usage accelerated accordingly, with Internet usage not only soaring to record levels but also with many consumers using data in whole new ways. This, though, was only the thin end of the wedge. As smart cities and connected vehicles increasingly become the default model, legacy infrastructure is made digital, and industry adopts more intelligent systems like robotics, IoT, and predictive maintenance—it's evident that today, we really are on the cusp of a totally connected planet. It's little surprise that by 2025, the world may have generated a staggering 175 zettabytes of data—and perhaps, with the explosion of data post-pandemic, even more.

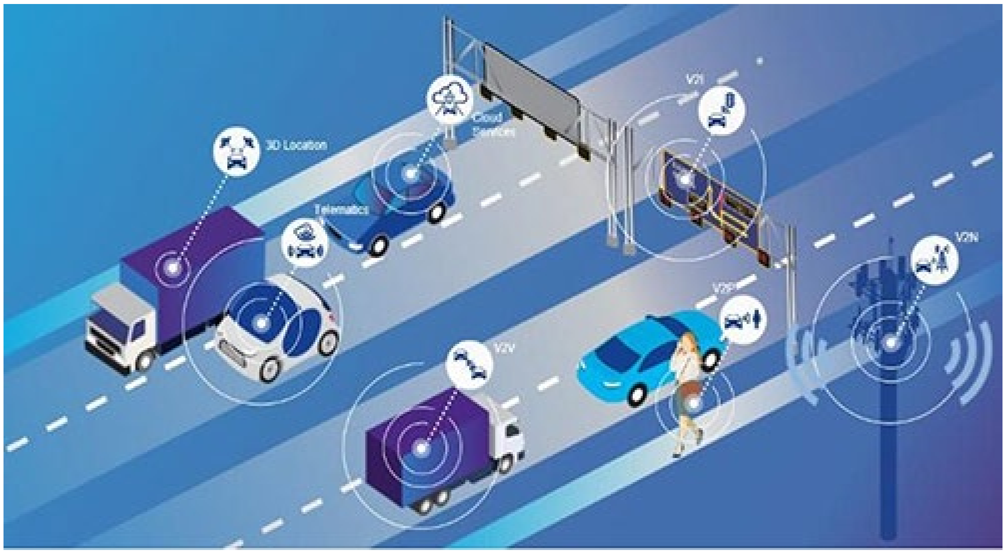

Figure 1: Enabling the Connected world

The tech nervous system depends on fast responses

Underpinning all of this are advances in connectivity like cloud, edge computing, and, crucially, 5G, which due to its high speed and ultra-low latency, is the fuel for this technological nervous system. The future of connectivity will not only be defined by frontier technologies like high-band 5G, says McKinsey, but also by expanding existing technologies like fiber, low- to mid-band 5G, and other long and short-range solutions.

Cloud-native economics and availability

Although traditionally telecom operators have favored a longer-term approach to data centers than turning to the cloud providers (including Google Cloud, Microsoft Azure Cloud, and AWS)—to achieve 'five nines' or 99.999 percent availability—there are now tremendous opportunities for telcos to tap into the hyperscale ecosystem. While 5G and Open RAN allow data transfer at low latency and with high availability, the information they relay still needs to be crunched, processed, and turned into actionable data at fixed locations, whether at a hyperscaler or an edge computing base. While International Data Corporation (IDC) defines a hyperscale deployment as five thousand or more servers on a ten thousand square foot or larger footprint, we believe the true definition is the ability to scale rapidly in response to demand—a growing requirement in the new era of 5G, Open RAN, and beyond 5G—and the data demands of the traffic that it allows.

Different demands, different performance metrics

In the next few years, this kind of connectivity will transform whole industries. Take healthcare: online consultations are already proving popular and are predicted to reach $836 million in market value by 2024. But more imaginative use cases like augmented reality-based training for doctors and nurses are also on the horizon. And remote surgeries for minimally invasive procedures may be worth $13 billion by 2025, with an installed surgical robot base of around 10,000 units. Each of these procedures has different requirements in latency, bandwidth, privacy, and mobility to support artificial intelligence. While routine health check-ups may be more forgiving regarding latency and reliability, remote surgical procedures underscore the need for end-to-end network monitoring and assurance.

In the transport and logistics sector, businesses will soon be able to take advantage of a wealth of new, better data at every stage of the delivery process. 5G is poised to enable the more thorough monitoring of not only the location of goods but also the condition of the goods, too—a continuous process unlike at present, where tracking is restricted to key checkpoints in the chain. This extra level of detail in every aspect of the supply chain opens new efficiency gains for logistics companies, who will be able to identify or reduce issues throughout the delivery process, like delays or cold storage faults, chipping away the potential for human error.

Figure 2: Unmanned data centers are the future

In addition, the manufacturing sector is increasingly turning to private 5G edge cloud computing to build a responsive 'nervous system' that can connect robotics with material handling and predictive maintenance applications, and 5G has become critical to realizing next-generation factory automation. But high bandwidth is required here to support smart factory rates from 1-20 GB/second—the real-time IoT sensors can often run into the millions for a single deployment, for instance.

Observability of everything is a 5G dream

Meanwhile, hand tools as simple as screwdrivers are now being turned into digital assets that stream torque, position, and calibration data back through the network. Datacenter software gathers, coordinates, and analyzes this sensor data, alerting facility managers to possible issues with predictive maintenance.

The shift to ultra low-latency edge computing

Smart cities are already enjoying infrastructure rejuvenation for the digital age, making street lighting, utilities, and traffic more efficient and greener. The emergence of more complex, better Advanced Driver-Assistance Systems (ADAS) means that, with vehicle-to-infrastructure communication, drivers and passengers will be able to optimize their routes based on real-time information from roads. In contrast, cities and road operators can build portfolios of valuable, real-time data about how thoroughfares and streets are used, better informing public safety measures and road maintenance. Here, system latency in the 1-2 millisecond range is crucial as a gating requirement for performance and driver safety.

In data centers, too, 5G-powered sensors and remote automation capabilities will manage and maintain the next generation, particularly at the edge—with these 'lights out' facilities surveyed by robots or drones and ultra-reliable low latency communications (URLLC) linking to service dispatch centers. In contrast, temperature and humidity sensors can feed environmental data back to automate hardware and HVAC adjustments. Because humans aren't required to maintain them, unmanned data centers can also be in cooler, less hospitable climates—or even underwater—to make them naturally more efficient.

Distributed and disaggregated 5G networks

Use cases like these are poised to reshape industries for the better, allowing businesses to dive more granularly into data and chart new paths for efficiency. However, with costs and complexities only set to rise, some of the biggest challenges for hyperscalers will be driven by the intricacies of distributed, disaggregated, cloud-native 5G networks.

With technologies like virtualized RAN, massive multiple input – multiple output (MIMO), and antenna beamforming further complicating radio frequency (RF) and network performance testing, new challenges will likely emerge around spectrum analysis, demodulation, and service-level agreement (SLA) conformance.

This end-to-end network complexity is a challenge in a world where operators have to meet higher customer demands for performance, efficiency, and reliability while also managing their own stack.

To overcome this and fully realize the power of 5G and Open RAN, infrastructure must be equipped with seamless end-to-end network slicing orchestrated for the needs of each unique vertical it serves. This requires a departure from legacy methods of data center and network testing and assurance because critical 5G IoT use cases leave no margin for error around reliability and SLA conformance.

Stress testing lowers costs, outages, and boosts performance

New developments in network stress testing can assist in safeguarding or even preventing potential real-world failures. In contrast, pre-deployment testing for 5G verticals could be broadened into RF testing and certification, RAN transport or Xhaul connection calibration and verification, and application, slicing analytics, and emulation.

By emphasizing upstream testing and validation, operators can reduce unplanned outages, troubleshooting, or the need for updates after deployment while also boosting performance. Cloud-enabled test automation can run high-performance throughput testing between data centers to verify network slice integrity. Self-healing is a key attribute of autonomous networks, while data center intelligence can automatically adjust cell bandwidth, location, and tilt. Meanwhile, the emergence of Open RAN architecture offers huge opportunities in low latency verticals, partly due to the complete disaggregation and multivendor interoperability inherent to Open RAN.

Dark, zero-touch networks are feasible

As data centers become a more active part of the telco network, they must meet the challenges required by telecom operators for reliability in the 5G era—where self-managed, zero-touch networks governed by automation and robotics are finally feasible.