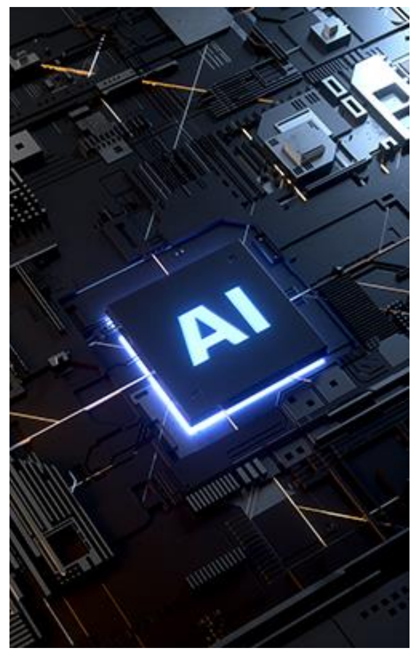
www.pipelinepub.com
Volume 18, Issue 6

Gearing Up for Industry 4.0 with 5G and Native AI

By: Subhankar Pal



The Internet of Things (IoT) is a major business opportunity for mobile operators. That's why 5G is designed from the ground up to support demanding IoT use cases and requirements—including ones that 4G and Wi-Fi find challenging to handle.

For example, 5G has three feature sets that are ideal for Industry 4.0 applications. These include Ultra-Reliable Low-Latency Communications (URLLC), which reduces latency to as little as one millisecond, key for mission-critical use cases such as time-sensitive networking (TSN), and Massive Machine-Type Communications (mMTC), which enables 5G networks to support up to one million IoT devices per square kilometer, such as autonomous material handlers, industrial robots and sensors. The last feature set is Enhanced Mobile Broadband (eMBB), which supports bandwidth-intensive applications such as 4K video cameras in and around a plant to monitor production, employee safety and physical security.

Even with these and other advanced capabilities, mobile operators will need additional tools to ensure that their 5G networks can take full advantage of the IoT opportunity. This is also the case for operators of private 5G networks, which 76 percent of manufacturers plan to deploy by 2024, according to an Analysys Mason survey.

Beyond baseline 5G

Artificial intelligence and machine learning (AI/ML) provide the foundation for Industry 4.0 and the benefits they enable. The mission-critical use cases in Industry 4.0 demand ultra-real-time AI/ML performance. So, 5G can't just be AI enabled. It must be AI native.

AI native is when AI/ML capabilities are built inside 5G network functions. One example of this has been to integrate AI/ML to enhance the radio access network (RAN) medium access control (MAC) scheduler. Thanks to this capability, the network can intelligently boost quality of experience (QoE) with ultra-real-time predictive analytics. QoE is critical for use cases like smart manufacturing where technologies such as digital twins can improve shop floor performance and enhance safety by providing remote human assistance through immersive experiences.

What's more, this kind of AI/ML enhancement also does not have to lock the operator into a proprietary ecosystem, and suffer from vendor lock-in. Operators can benefit from Open RAN (O-RAN) guidelines with more choice and flexibility to maximize spectrum efficiency. This allows them to overcome one of the major limitations of today's RAN architecture: radio resource management can be static and unable to adapt quickly to changing traffic conditions and subscriber behavior.

This example also highlights how AI/ML will fundamentally transform 5G communication systems and how they will be designed and deployed in the future. AI/ML will have the capability to even control parts of the physical (PHY) and medium access control (MAC) layer functions. All of this directly benefits IoT, a plethora of Industry 4.0 use cases and more—now and well into the future.

Optimizing efficiency and performance

Many manufacturing IoT applications are designed to maximize efficiency. Just as how AI/ML can transform 5G networks, it can do the same for manufacturing IoT applications. Here it can reduce the amount of radio resources, electricity and spectrum required. Of course, the savings in electricity can help operators boost their green credentials and deliver CAPEX and OPEX savings. In the case of a public 5G network, for example, CAPEX and OPEX savings can help the operator price its services competitively yet profitably—a major benefit in the notoriously cost-sensitive IoT market.

These benefits are possible even in demanding use cases. For instance, when IoT applications such as live 4K video require ultra-fast gigabit speeds, they will not requiremore radio resources, spectrum or power than low-bandwidth applications. How is this possible?

It is because RAN functions such as admission control, radio-resource scheduling, mobility management, and radio-resource management are all currently rules-based. They cannot adapt to the dynamics of the network, or the services accessed by each IoT device. When ML-infused intelligent automation is introduced to power RAN functions, especially in lower protocol layers, it can make decisions dynamically for efficient network resource allocation and improve the subscriber experience. Therefore, it can dynamically boost use cases such as live 4K video.

Critical considerations when infusing AI/ML

The key challenge when introducing ML into the lower RAN layers is the limited time window available for scheduling decisions, which take about one millisecond. For IoT applications where every fraction of a second counts, ML algorithms cannot introduce any delays. Even a few seconds of processing delay can severely impact the overall QoE and undermine spectral efficiencies.

It's also critical to ensure that ML-based updates can run on existing 5G-gNodeB platforms without requiring hardware-accelerators or GPUs. This can be achieved by introducing intelligence into the most critical and time-sensitive Layer 2 MAC scheduler functions, which play a key role in radio resource allocation. In lab environments, this approach has boosted spectral efficiency by 15 percent while increasing cell throughput by 11.76 percent. This is what empowers operators—public and private—to boost each IoT application's QoE dynamically. That's a remarkable capability.

In the MAC layer, AI-powered predictive analytics can forecast and then assign the appropriate modulation and coding scheme (MCS) values for signal transmission based on each IoT device's signal quality and mobility patterns. The RAN can then intelligently assign MAC resources, across different network layers to meet each bespoke need. For example, it can deliver more accurate MCS predictions, achieve better spectrum efficiency and intelligently handle different devices that have diverse requirements. This automation can efficiently increase the amount of traffic on each cell, giving operators the ability to manage more subscribers and launch new IoT applications successfully.

Powerful, flexible, and greener frameworks

The framework for native AI and 5G utilizes fully containerized architecture. It is cloud-native and runs as a collection of independent and loosely coupled microservices. They are highly scalable and can run on top of public, private, and hybrid clouds—giving operators more flexibility and choice. They use service mesh architecture for managed, observable, and secure communications. And by using APIs and the service mesh to abstract the microservice network's complexity, this powerful framework allows operators and enterprises to introduce and scale new services fast.

As both the hardware layer and the AI layer are abstracted, manufacturers and other enterprises have the agility, performance, and savings they need to stay connected, competitive and profitable. Connectivity with native AI is critical for the digitization of industries and to accelerate Industry 4.0 initiatives. The AI native approach optimizes the radio resources allocated to users with intelligent link adaptation. This improves QoE and supports mission-critical operations. And improved QoE is one of the key attributes of O-RAN. O-RAN with native AI is a powerful combination.

One of the main benefits of native AI and 5G O-RAN is especially timely. Integrated together, they can deliver outstanding QoE for Industry 4.0 by optimizing radio resources and higher bandwidths. Importantly, it can do so without consuming more radio resources. That is less power consumption to transmit radio signals. Put simply, this improves energy efficiency and enhances sustainability. Sustainability is said to be at the top of telco agendas after research revealed that the industry's CO_2 emissions are nearly twice that of the aviation industry. Clearly native AI on mobile networks can't come soon enough.