Intelligent RAN congestion management could help operators to get more from their 4G networks

December 2021

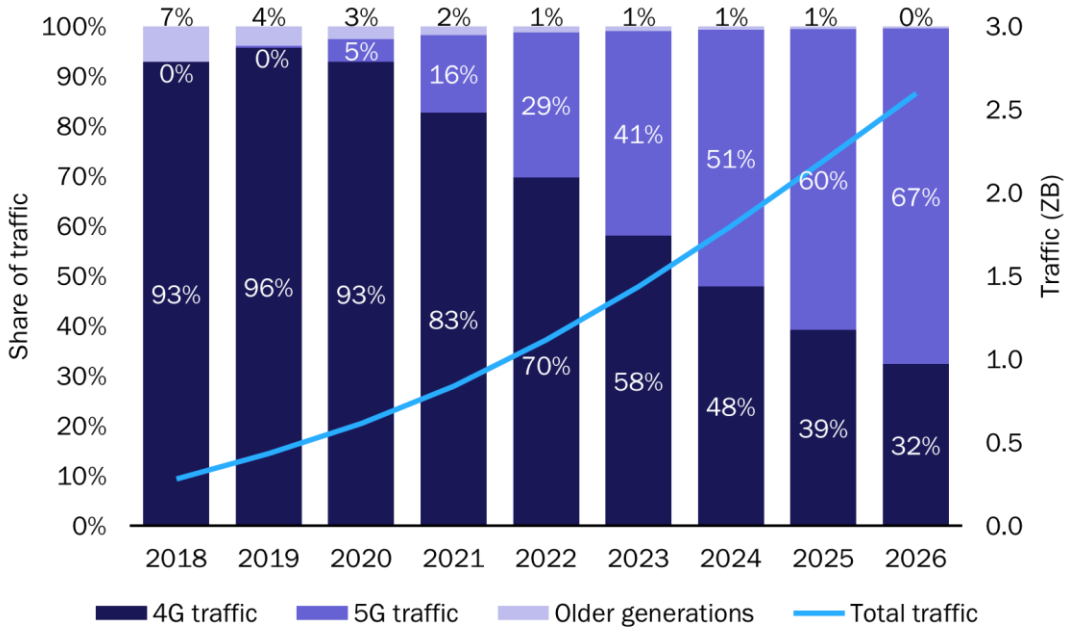
Ameer Gaili and Gorkem Yigit

Managing network congestion and quality of experience (QoE) is becoming increasingly challenging for mobile operators as the shift to new video delivery protocols accelerates, and consumers' mobile data consumption patterns change due to the COVID-19 pandemic. Mobile operators need to ensure that their network and RAN optimisation techniques are able to cope with these trends, particularly for their existing 4G infrastructure, in order to maintain a high level of quality of service (QoS) and customer satisfaction. In this article, we consider the growing importance of intelligent RAN optimisation and congestion management to enable operators to maximise the return from their existing 4G network assets without the need for significant additional capex/opex.

4G networks will still carry more than half of all mobile traffic worldwide in 2024

5G will provide major additional capacity to mobile networks, but the availability and coverage of 5G networks around the world will remain limited for the next few years. Large portions of the population, particularly in emerging markets, will depend mainly on 4G networks to access mobile services. Existing 4G networks will therefore continue to play a critical role in consumer connectivity services as operators invest in 5G roll-outs. Indeed, Analysys Mason forecasts that mobile data traffic worldwide will grow at a CAGR of 27% between 2020 and 2026, with no major 5G-related surge, and that 5G traffic will only overtake 4G traffic in 2024 (Figure 1). OTT mobile video will continue to be a key driver of mobile data consumption, and mobile/cloud gaming will also account for a growing share of traffic.

Figure 1: Cellular data traffic by generation, worldwide, 2018–2026[1]



Source: Analysys Mason, 2021

The COVID-19 pandemic caused some operators to delay their 5G roll-out plans and concentrate on getting the most out of their existing network infrastructure to cope with the traffic surge. This delay in 5G deployment progress has extended 4G's lifespan in developed markets, and those operators in emerging markets that do not have immediate plans for 5G have been forced to enhance their 4G networks to handle the growing number of users and traffic.

All these trends show that mobile operators will need to balance 4G networks improvements with 5G investments, at least in the short term. Extending the performance and lifespan of existing 4G networks will be critical to address the mobile traffic demands of today and to avoid any slowdown in 5G roll-outs.

Legacy methods for RAN congestion management are becoming less effective due to new consumption patterns and delivery protocols

There are two major challenges for operators regarding traffic optimisation management in the RAN: the shift in peak data usage as a result of the pandemic and the evolving encryption protocols for video streaming. These two factors are increasingly rendering legacy RAN congestion management solutions ineffective.

Most of the RAN congestion management methods that are in place today are based on predefined policies that rely on fixed locations (such as business districts and city centres) and historical traffic patterns (for example, peak traffic during commuting hours in the morning and evening). However, the pandemic has disrupted working patterns worldwide. Peak times and congestion locations have changed dramatically (for example, traffic in commercial districts has declined, while that in suburban/residential areas has increased) and have

1 For more information, see Analysys Mason's Wireless network data traffic: worldwide trends and forecasts.

become much less predictable. Existing static and reactive RAN optimisation and congestion techniques cannot fully address mobile users' changing consumption behaviour.

The second challenge is the growing adoption of new content delivery protocols and encrypted traffic. Major OTT service providers such as Google and Meta dictate how video is delivered over the internet by using their own, proprietary variants of standardised protocols (Google QUIC and Meta mvfst, respectively); other OTT content providers are following suit. In addition, the level of encryption used in these video delivery mechanisms and protocols continues to increase. For example, the new TLS 1.3 protocol, which is already supported by Google, Meta and Netflix, obfuscates the server name indication (SNI) data that operators use to identify traffic and manage network congestion. Unpredictable OTT protocols and behaviour will therefore reduce operators' ability to optimise capacity because their current RAN optimisation policies are highly content-specific and traditional packet inspection will not be able to identify and categorise these encrypted traffic flows effectively.

Together, these two factors present a major barrier to delivering a high QoS/QoE to mobile users. Addressing them by adding more capacity in the 4G RAN (for example, via densification or MIMO) is an expensive option for operators and is not ideal given that many are focusing their network investments on 5G.

Intelligent RAN optimisation solutions are needed in order to cost-effectively manage 4G RAN congestion

Operators need more-sophisticated network traffic and RAN congestion management tools and techniques that can effectively classify encrypted traffic and provide near-real-time monitoring and immediate action in order to cope with changing consumer behaviour and content delivery mechanisms. Some key capabilities that operators should look to obtain include:

- the ability to handle the continuous evolution of encryption protocols through AI/ML-based heuristic models and reporting structures
- the ability to detect and predict congestion (for example, by using ML algorithms/model training) and ensure that quality of delivery and quality of picture thresholds are met consistently
- the ability to manage and optimise traffic at an individual user session level in an automated manner, rather than by using static parameters, in order to take pre-emptive action before a user experiences congestion. Some operators have reported being able to use this capability to reduce the number of congested cells by 15% and create additional RAN capacity to deliver 25% more video traffic.[2]

Video/content QoE (determined by the start time, buffering time, lags and streaming quality) has a big impact on customer satisfaction and operators strive to be the best or the fastest at delivering mobile services. However, RAN congestion in 4G networks remains a challenge. 4G networks will remain crucial for mobile connectivity, so intelligent RAN optimisation and congestion management will continue to play a key role in enabling the delivery of a high QoE and freeing up resources that can be put into 5G.

2 ENEA Openwave, *Session Congestion Manager: get more out of your 4G network*. Available at: https://owmobility.com/traffic-management/session-congestion-manager-scm/.