How does edge computing architecture impact latency

This article answers that question mapping out the potential “level of edge” and where different operators have announced they are testing or planned to deploy edge nodes.

Dalia Adib, Edge computing practice lead

December 2019

© STL Partners

The edge computing latency promise

One of edge computing's promises is reducing latency to sub XYZ milliseconds thanks to the benefits of an edge computing architecture. The issue is that different applications have different levels of latency they can tolerate and, since distributed compute architecture is a broad spectrum of potential locations, these applications should be optimised to run at the optimal location.

The question of "where is the edge?" is difficult to answer and is often lost in translation depending on who you are talking to. For gamers, the edge is the end-device or the console, for manufacturers, it is more likely to be on-site within the production facility, for a CDN provider this may be at internet exchange points, yet for the telco community it is within the operator's network. We have termed this the network edge.

At a high level, the network edge refers to resources for storage and compute closer to the end user somewhere on the telecoms network. The challenge is that each operator has a different topology, which means that although each mobile network is made up of an access, transport and core network, the number of hops it takes for traffic to get through the network is not equivalent across operators. So, it's not as simple as saying a use case that requires 70ms should reside at a network aggregation point.

Network edge computing architecture considerations

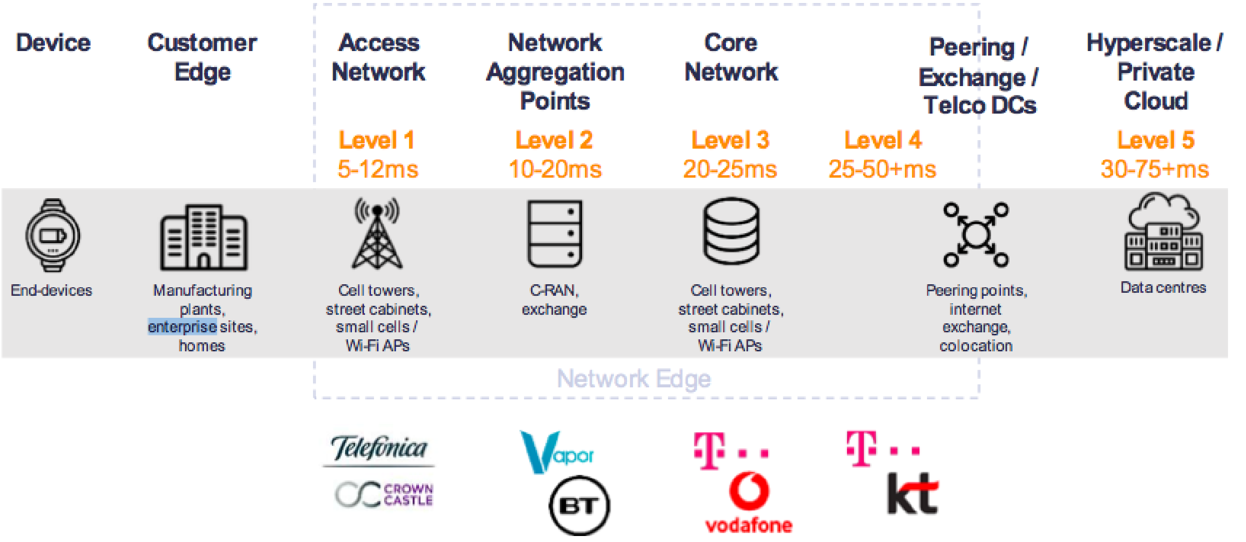
Mobile operators are evaluating a range of factors when determining how they should build out their edge computing infrastructure:

1. Application latency needs: most applications can tolerate latency of 100ms or more, but there are some that have a sub-50ms requirement, for example streaming virtual reality or for mission critical applications
2. Network topology: as mentioned above, operators need to be able evaluate where a potential edge site is in terms of hops, as well as which part of the network it sits in, to determine if it can meet the latency requirement for a particular use case
3. Site and maintenance costs: the cloud computing economies of scale may diminish as the number of sites proliferate at the edge, for example there is a significant difference in maintaining 1-2 large data centres to maintaining 100s across the country
4. Site availability: some operators' edge compute deployment plans assume the nodes reside in the same facilities hosting their NFV infrastructure, however NFV deployments are work-in-progress and many telcos are still in the process of renovating these locations to turn them into (mini) data centres
5. Site ownership: operators don't have an advantage at sites they don't own – in the US, the cell towers are owned by the likes of Crown Castle, American Tower and SBA Communications, which limits how and where the US network operators choose to deploy edge compute

Where is the edge for telcos?

Below we have mapped out the potential "level of edge" and where different operators have so far announced they are testing or planned to deploy edge nodes.

© STL Partners



Source: Disruptive Analysis, Dave Burstein, STL Partners analysis

Level 1 edge clouds at the cell towers don't exist yet and may be too costly to deploy – 1ms is in the lab and years away from commercial deployment. Level 2 is exemplified by Vapor's approach, whose facilities are not at the tower itself, but connected close by which adds about 4ms. BT has also discussed that it estimates 1000 sites to serve 15-20ms. Level 3 is further back in the aggregation or core network, depending on the number of hops, and at level 4 we are the extreme edge of the network connecting to the internet, which becomes almost equivalent to peering points. As operators like Telefonica and Verizon simplify and accelerate their networks, the gap between levels 3 and 4 closes.

Level	Network	Latency	Notes / operators
Level 5	LTE	30-75+ms	Today
Level 4	5G or LTE	25-50+ ms	Deutsche Telekom (phase 1), Korea Telecom
Level 3	5G with servers in telco core	20-25 ms	Deutsche Telekom (phase 2), China (potentially), Vodafone
Level 2	5G with few hops	10-20 ms	Vapor (with Crown Castle), BT (potentially)
Level 1	5G with 1-10 ms air latency	5-12 ms	Not expected out of the labs for years Telefonica (long term)

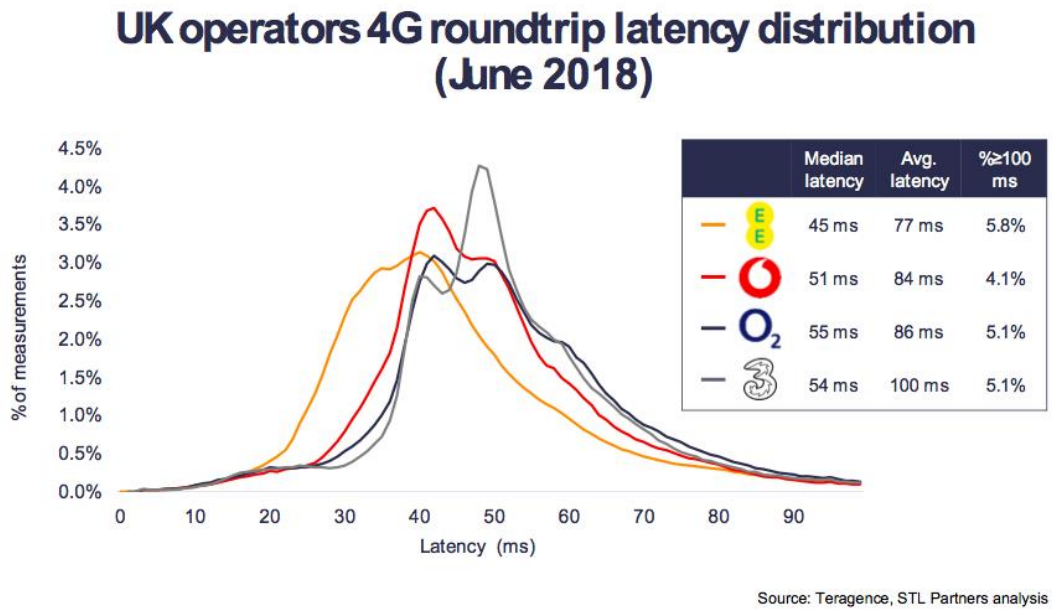
© STL Partners

But are we getting distracted by latency?

From our discussions with some of the application developers in the augmented reality, CDN and drone space, much of the debate has been around the ability to maintain a certain level of latency and reduce jitter, rather than meet an absolute latency limit.

Unfortunately, telcos do not have a great record of being able to ensure consistent latency. Analysis from June 2018 (below) shows the difference between the mean (average) latency and the median for the UK's top mobile operators. Although half of the population* actually benefit from latency of below 54ms across all 4 operators, almost 5% suffer from latency of higher than 100 milliseconds.

The challenge for operators in providing edge cloud will be to guarantee a level of service, particularly for enterprise customers who will need to depend on the network and the infrastructure if they are going to use edge for their (mission critical) applications.



*measured in this sample

This article has been adapted from Dave Burstein's articles at The Edge Reporter

© STL Partners

Dalia Adib is Edge computing practice lead at STL Partners. She has led major consulting projects with Tier-1 operators in Europe and Asia Pacific on edge computing strategies, use cases and commercial models, and is an active speaker at events including Edge Europe and Data Cloud Congress. Outside of edge computing, she supports clients in areas such as 5G, blockchain, digital transformation and IoT.

Get in touch with the author to learn more

dalia.adib@stlpartners.com

Or visit STL Partners' Edge Hub

www.stlpartners.com/edge-computing

© STL Partners