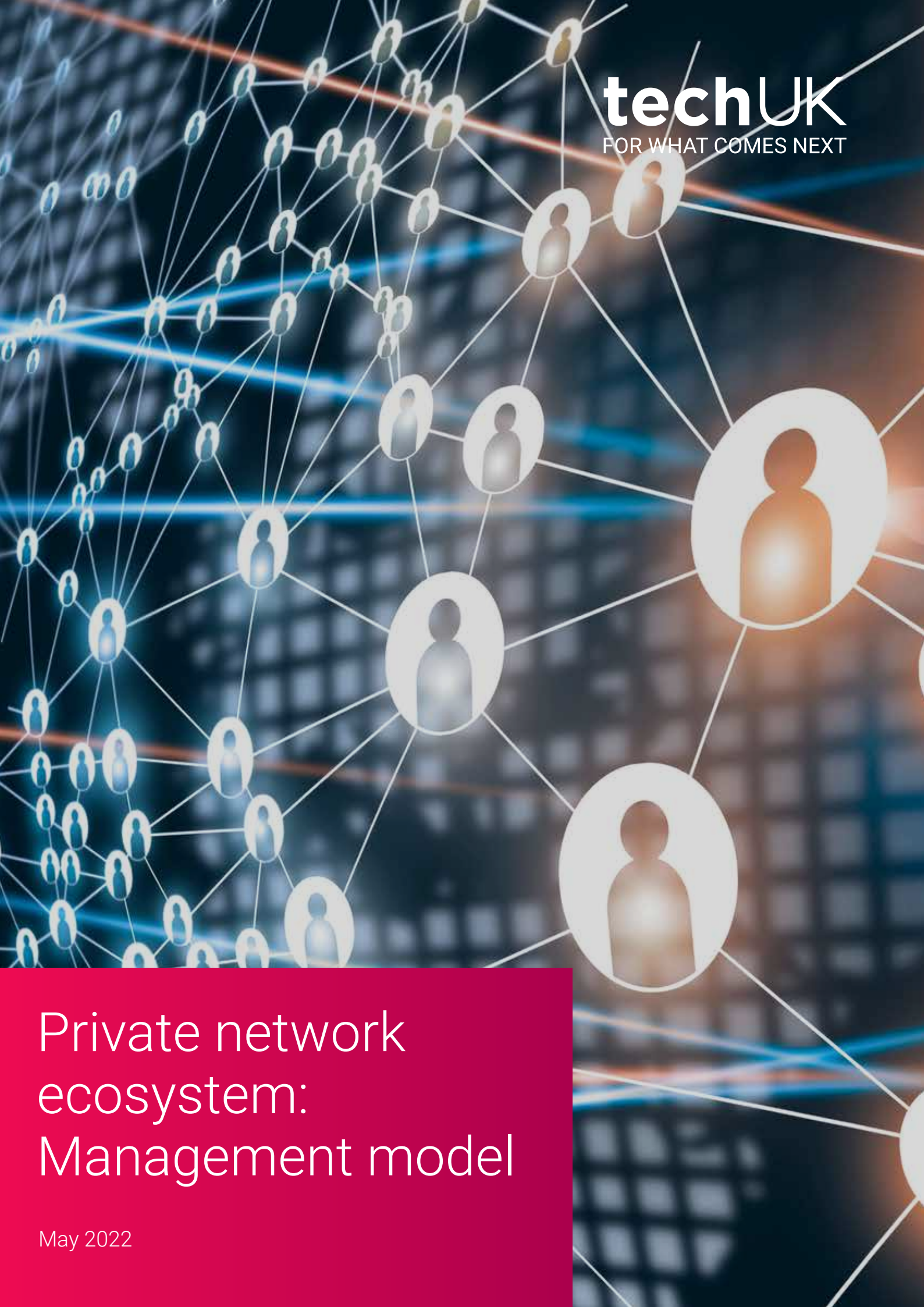
techUK
FOR WHAT COMES NEXT



Private network ecosystem: Management model

May 2022

Contents

Introduction	04
Overall architecture of edge native applications – components of ecosystem	06
Value chain and operating model	08
Resilience at the edge	11
Ecosystem operations	13
Neutral Host Networks	14
Market drivers	15
Neutral host challenges	16
Access to spectrum	17
5G Delivery over radio networks	18
Useful links and further reading	21
Annex 1	22
Acknowledgements	23

Introduction

This paper is targeted at organisations considering building out services using 5G networking technology. It introduces and describes the principles of neutral hosts, and then goes on to describe the architecture and ecosystem which supports the provision of shared services, particularly in the context of high capacity/low latency applications, which will drive 5G deployment. While this paper is focussed on 5G, many of the principles of neutral hosts, and the discussion of edge versus core provision will apply to other technologies such as Wi-Fi, including Wi-Fi 6.

The full range of benefits that can be delivered over 5G requires other services to be provided in parallel. Devices specific to the application will need to be procured and deployed. Data will need to be gathered and processed, either in the device itself, at the edge of the network or centrally, depending on the specific use case.

The provision of edge compute resources will enable a new generation of applications to be developed. These *edge-native applications* depend on both the high-speed and high-capacity features of 5G and the availability of low-latency compute without impacting backhaul networks. The value chain for edge-native applications requires an ecosystem of providers and services, of which 5G is an important component.

Examples of such applications include:

- Services which require rapid location tracking, e.g., semi-autonomous robotics in a factory environment or vehicle platooning
- Services which require very high bandwidths and low latency, e.g., Virtual Reality/Augmented Reality
- Services that require processing very large amounts of data in time-limited applications, e.g., high-definition video analytics for quality control, security, transaction generations, etc.
- Services where highly confidential data must be processed within a secure zone to ensure total security of information

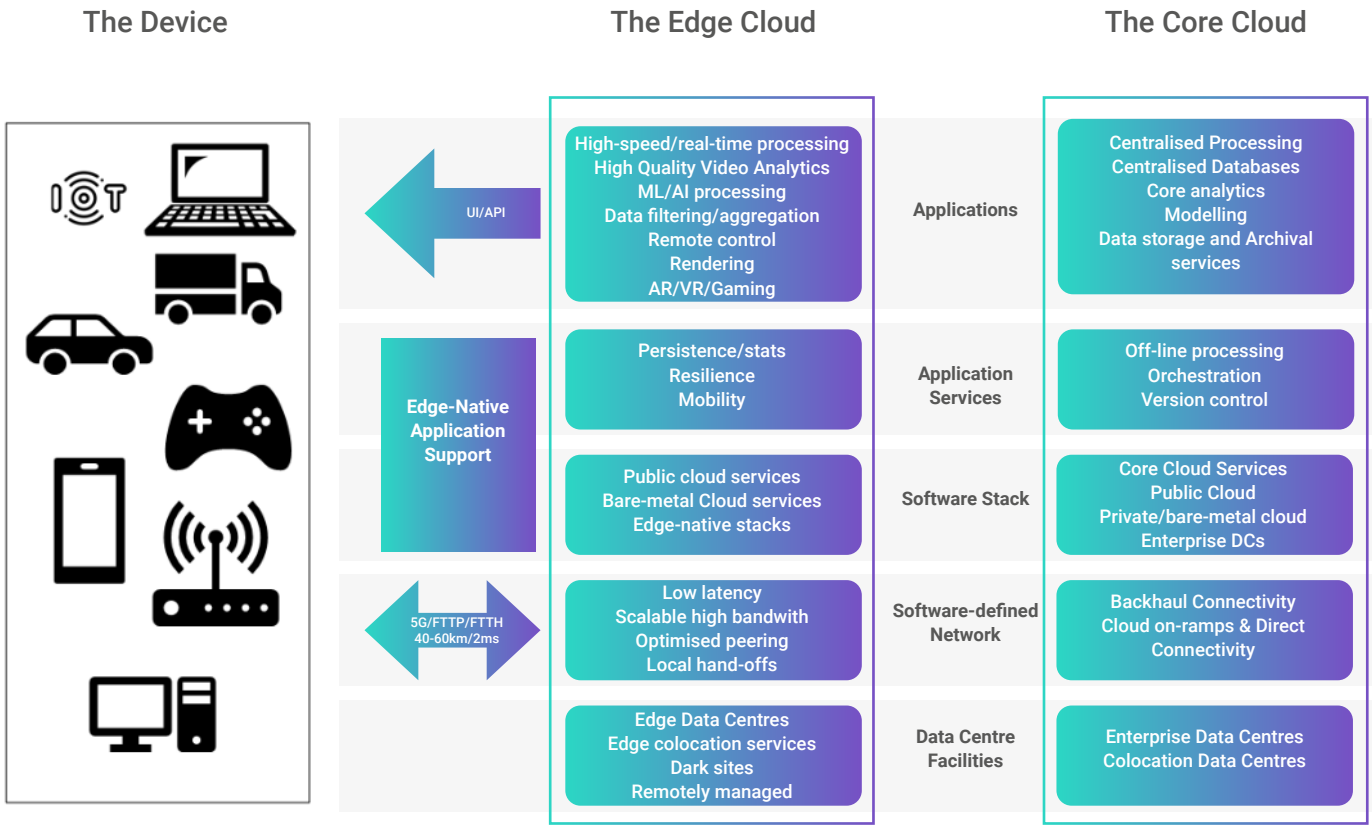
This paper focuses on 5G in the context of both private and public networks, recognising that some use cases may cross both environments. For example, tracking a high value manufactured good (for example an Electric Vehicle) through its creation in a factory (Private 5G), and then delivering services to it in-life (Public 5G). Many of the principles outlined in this paper can be equally applied to other communications technologies, in particular Wi-Fi, which is already widely used in private networks.

The ecosystem consists of:

- Public network providers (MNOs)
- Fixed network providers
- Data centre providers (including edge data centres)
- Private network providers (May be an MNO, a System Integrator, Neutral Host, equipment vendor or other provider of wireless telecommunications infrastructures)
- Wireless infrastructure operators
- Device providers
- Edge cloud providers
- Specialist application providers (specific to the use case)
- Provider(s) of ongoing support for the deployed solution

This is not an exhaustive list, and specific providers may deliver in multiple aspects. For example, an MNO might offer a full solution. A System Integrator or Mobile Network deployment specialist might provide a solution bringing components together from various parties, bringing specific domain expertise relevant to the customer.

Overall architecture of edge native applications – components of ecosystem



The overall ecosystem that 5G is a vital part of provides a technical architecture and infrastructure that can support edge native applications on high-speed and high-capacity networks. The smartphones, vehicles, cameras, and devices used by consumers and businesses will require access to a wide variety of cloud and edge services in order to fully exploit the possibilities offered by edge native applications.

In a distributed environment, system developers need to make decisions on where to execute functionality, and how to balance this between the edge and the core elements.

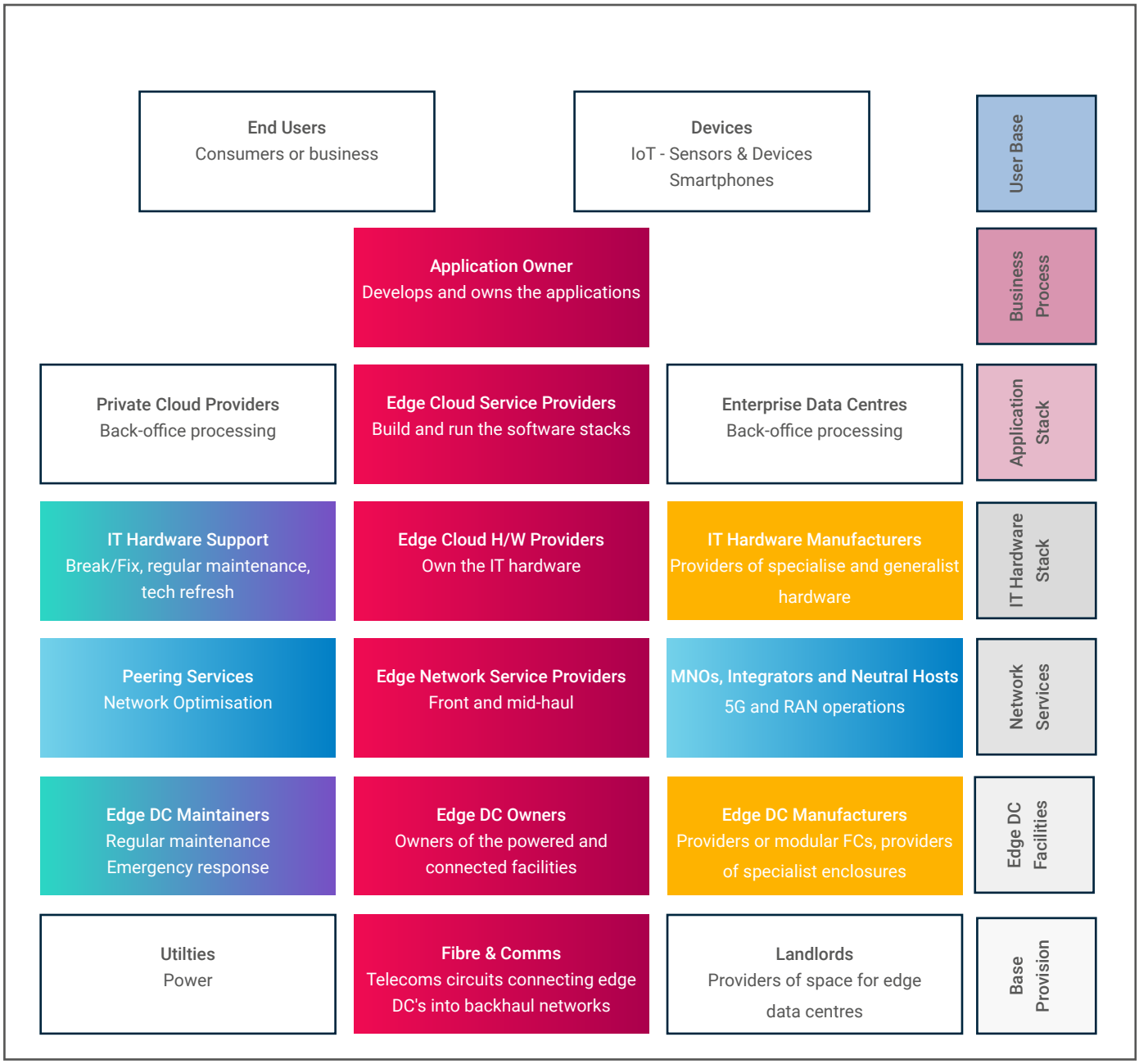
The edge cloud is likely to be most appropriate where latency is critical, or where very high volumes of data needs to be analysed, but not fully stored.

An example might be visual analytics, where a real time video feed can be analysed for specific events (e.g., an intruder in a largely static scene, an unattended toddler in a shopping mall) which need to be flagged to the core for further action, but the video content itself is mostly discarded, except for a small window of time before and after the event.

This is the core technique used in “cashier-less retail”, taking multiple video streams (as data) and producing retail transactions (as information). Another use case is for generation (rendering) of images for AR/VR or gaming systems, where low latency is required to generate imagery that avoids the “travel sickness” issues of existing remote rendering systems and provides a sufficiently reactive gaming experience.

The core is likely to be most appropriate where data from multiple sources is aggregated together over a period of time, or where static data is held in a system of record. The core will also act as the central repository of data for centralised processing of data from multiple sources, the majority of which may be at the edge. The core is also likely to be the venue for transaction processing (e.g., payment systems, ledgers, etc.) for events originated in edge locations – for example cashier-less retail. In a mobility scenario, the core may also be used to deliver a subset of the services available at the edge, e.g., in a city, where edge compute may be available, it might be possible to deliver a wider range of services to an urban setting, where edge might not have been deployed, in which case some fallback to the core network might be appropriate to deliver some (but not all) of the services.

Value chain and operating model



The diagram above illustrates the range of components needed to deliver a complete service. These are unlikely to all be provided by one provider, so defining each in turn helps to provide a framework enabling each element to be described in a way which enables the group of components to be designed and procured to work together effectively.

Base Provision

The fundamental resources needed to build an edge node. Typically, edge nodes will be small by data centre standards. Current deployments go from 50kW to 500kW of power, requiring only 500-2,500ft². Fibre connectivity is an important pre-requisite for most edge nodes although examples have been noted using 5G for connections into backhaul networks – only suitable over short distances.

Edge DC Facilities

There are numerous designs for edge node facilities. Many are based on “container”-like solutions – metal boxes that resemble (or are actually built using) ISO containers. They have the advantage of being very easy and fast to deploy – built off-site in fabrication yards or factories, transported to site and lifted or rolled into place.

These systems are usually highly modular and can start as small units and then expanded as demand requires. However, edge nodes can be situated inside existing buildings just as easily and will often benefit from enhanced physical security.

Edge DCs will almost always have to operate “dark” – with no permanent on-site staff – with remote support for planned and unplanned maintenance services. Likewise, monitoring of critical systems (mechanical, electrical, fire detection and suppression, leak detection, physical security and CCTV, etc.) are expected to be monitored centrally, ideally through a “single pane of glass” model. Automation of controls within the edge data centre is likely to be used at a much higher level than in traditional data centres and several Artificial Intelligence/ Machine Learning (AI/ML) sub-systems have been developed to assist with automation.

Edge data centres are usually designed to carry high-density cabinets, with 15kW/cab being quoted by many operators as a standard density. This is to cater for the hyperscale and other cloud workloads that incorporate high-density server technology and specialised kit for AI/ML workloads.

Network Services

The Network Services layer is a vital element in the edge data centre ecosystem. Most of the technological benefits delivered by edge compute derive from the network layer. Intelligent routing forms the basis of the distributed model that allows local data processing to happen without the default behaviour of backhauling all traffic to Internet Exchanges (IXs). In effect, many edge nodes will act as mini Internet Exchanges for the applications that they carry. This is done in a transparent way that does not require application developers to "understand" or be aware of the network topology.

IT Hardware Stack

Edge compute is typically based on a mixture of standard server technology and more specialised hardware utilising GPUs for AI/ML workloads. Estimates for power density are constantly increasing and edge sites are typically dense compared to standard data centres – 15kW/rack is not uncommon.

Monitoring of IT hardware is a remote, centralised service and in most cases is carried out by the third party operators providing facilities monitoring (as above). Break/fix and other unplanned maintenance services must be considered as long lead-time services, with long service-level agreements (SLAs).

Application Stack

The application stack provides useable IT cloud components and capacity enabling applications to be created and operate, either at edge or in the core. The range of applications available is constantly evolving and depends on the specific cloud provider. It typically includes compute elements, storage and virtual networking, together with management and deployment tools, enabling a complete application to be constructed and operated at scale.

Edge data centres can mostly be seen as an extension of cloud services into edge locations, so the application stacks act as provisioning platforms in a multi-cloud environment. As with the network layer, the application stack needs to provide a transparent platform that application developers can rely upon to provide the essential services they require. This includes persistence, which forms an important part of a resilient service and also one that works for applications where the end-user is moving (e.g., in cars).

The operating model for the whole ecosystem contains a lot of components but some are of particular importance for 5G services. In particular, the network service layer must be set up to allow applications that run across 5G services to be correctly routed to avoid unnecessary use of backhaul networks. In effect, the edge services that 5G networks connect to need to act as IXs so that data that can be processed locally in edge compute data centres remains "local".

The Open Grid Alliance is a not-for-profit industry body that is helping define the characteristics and standards for an open architecture for supporting edge native applications.[1]

Resilience at the edge

What does “resilience at the edge” mean? Usually resilience means fault tolerance – the ability to carry on or recover when something goes wrong. For edge native applications, which rely on edge infrastructure, this means that the application should be able to recover from any type of failure or fault in its supporting ecosystem. As the edge is, by definition, a distributed environment, resilience for edge-native applications needs to be delivered by the infrastructure – in particular the edge cloud stack and the network.

Historically, one of the most common approaches for providing resilience has been to build resilient data centres with fault tolerance built into the electrical and mechanical systems that support the IT. A top-end data centre should be able to continue to operate without interruption except in the most catastrophic scenarios. Providing this level of fault tolerance is, however, very expensive and is not an approach that scales well at the edge. Not only would a highly resilient edge data centre be prohibitively expensive for all but the most specialised uses, but space is often at a premium and resilience uses a lot of space for redundant systems. This can also be seen as ranking low in sustainability. Lots of redundant kit is required, it is inherently less energy efficient and produces a bigger carbon footprint.

Whilst 5G and edge systems should be mostly reliable, real-world instances will suffer from unpredictable failures and it is reasonable to expect some maintenance and repair services to be relatively slow to deliver. Thus, 5G and edge infrastructure ecosystems must be designed with resilience that can account for catastrophic loss of communications or processing sites.

The best channel for resilience is via software. A combination of the software-defined network layer and the edge cloud stack that applications run on should provide resilience. Ideally, if an edge compute node was unable to continue to run an already running application, another node on the network should be able to take over seamlessly, without any user or application intervention. Even more ideally, without any noticeable delay. It is worth considering that the problem being solved here is extremely similar to the “mobile edge” – for example, a vehicle moving along a road, connected to an edge data centre that is running applications that the vehicle is using will need to connect to different edge nodes as it progresses in its journey. In this example, edge data centres need a way to “hand off” the running apps between each other, in much the same way as mobile networks do in exactly the same scenario. It is the same in the failover processes that provide resilience. If a node fails, the nearest can take over.

The main problem is ensuring persistence. This requires the state of a system or application to be storable and transferable. This should be provided by the edge cloud stack and enabled by the software defined network interconnecting the edge nodes.

Application architecture considerations

The architecture and value chain described above provides a framework to help an application owner design and operate a solution.

During the design phase, the primary considerations are likely to be:

- **Functionality** – what needs to be delivered to support the defined business problem being solved?
- **Devices** – what devices will be used (sensors, displays, VR headsets, actuators etc) and how do they impact the performance requirements?
- **Mobility** – how far and how fast does the user move while using the service? What does this mean for the radio access solution? Is wide area or local coverage needed?
- **Data privacy** – how sensitive is the data, and what does that mean about where it is stored and managed?
- **Performance** – what are the performance criteria (latency, throughput, reliability, volume of data, number of devices) needed to deliver the service?
- **Caching and replication** – are native solutions offered by the cloud stacks or are specific sub-systems required to ensure data consistency and availability?
- **Security and resilience** – are there specific requirements to secure the connection, maintain a private network connection? What happens when the system degrades (due to a failure elsewhere), and how should it recover from a failure?

Ecosystem operations

Once the solution has been deployed and proven to work through formal Acceptance Testing, it enters an operation. In practice, this may be phased over multiple releases as functionality is deployed, but there should be a clear handover between “Deployment” and “Operation”. The team providing this operational support may be from one of the deployment contractors, or a new contractor focussed on operation of complex technical estates.

During the operational phase, the following need to considered:

- What elements need to be actively managed?
- What orchestration and provisioning mechanisms are required?
- How are elements provisioned and de-provisioned (SIMs, devices, data, users etc.)?
- How will in-life upgrades be delivered (over the air) to devices?
- How will performance be monitored and maintained?
- How can the system be scaled cost effectively as the user base grows?
- How will security threats be identified and neutralised?

These issues arise as a result of transitioning applications into a fully distributed environment. This builds upon the problems first experienced by enterprises when using third party cloud services. The automation of processes to support provisioning, orchestration, monitoring, and management of distributed applications requires a change in thinking for those with operational responsibilities.

The challenges require new approaches to developing the ideal of a “single pane of glass” for system management. There are new elements to distributed applications and ecosystems that require special consideration:

- A more complicated value chain – more layers of system delivery and operators
- A more complicated application architecture
- More components requiring monitoring than ever
- Several components of the ecosystem are leading or even bleeding edge technologies – there are limited resources available who can effectively diagnose and resolve problems
- There is limited collaboration between players in the ecosystem

Neutral Host Networks

One of the most interesting developments in the telecom industry recently has been the neutral host network, whereby a neutral third party builds and operates part of the network offering private and public connectivity, as opposed to the individual deployment and operation by each Mobile Network Operator (MNO) or other provider of wireless telecommunications infrastructures. This represents a significant benefit for operators, who can expand their reach more easily and with less investment by utilising neutral host networks, and that anyone can use a Neutral Host, not just an MNO.

Examples already exist of neutral host networks in multi-tenanted offices, hospitals and stadiums (e.g., Stanhope and the Freshwave Group in London's Angel Court[2]) along with pilot programmes for small cell deployments (e.g., Dense Air in Dublin[3]).

The four MNOs in the UK make use of shared masts via two consortia which operate them on their behalf. This is well-established form of shared facilities delivering value. Similarly in the fixed network, Openreach are required to provide access to ducts and poles, via the Passive Infrastructure Access (PIA) product. This reduces the costs for fixed operators entering the market.

There will be challenges in deploying neutral hosts at scale, such as standards for approach, the process of connecting and activating capability in a neutral host, and the use of shared resources such as radio spectrum. The technology is maturing rapidly, but the ecosystem needs to develop so that it functions at scale. Each solution is likely to be different initially, but as the industry matures, so standards and approaches will become more consistent.

Market drivers

Neutral host networks help improve mobile phone and data performance for consumers and businesses in a cost-effective way. In some specialist use cases, for example trackside provision of mobile connectivity through a sparsely populated area, it may be the only practical way in which coverage can be deployed cost-effectively.

The recent Ofcom changes in licencing means that local spectrum licences, limited by range and frequency, allows a new class of operator to enter the market. Local spectrum licences also enable private networks for enterprises to address IoT and voice services. Private networks can also offer capacity to MNOs to improve mobile coverage.

Neutral host challenges

Deploying neutral host networks entails a number of challenges, for example flexible authentication methods supporting different device types, client confidentiality as well as encryption will be required services that neutral host providers will have to solve.

Indoor mobile coverage and reliable indoor cellular coverage is important to Industry 4.0, but is becoming more difficult to guarantee due to the higher frequency bands being licensed and the use of materials in buildings that are not conducive to high-frequency radio signals, such as low-emissivity glass. Neutral host networks can help mitigate these issues and help provide indoor connectivity in a wide range of buildings, e.g., public amenities; hospitals and healthcare centres; educational establishments and campuses; museums, exhibition halls and convention centres; with a single infrastructure that can cater for all MNOs.

Likewise, in rural and remote areas, and on major transportation trunks, neutral host networks offer the most economically viable method to provide high service levels that all operators can access.

Neutral host providers and operators will have to collaborate on developing a common standardised approach, including security, integration, and performance. For neutral host networks that connect with public mobile networks, permission to connect will be dependent on stringent tests and obligations. Roaming and interconnectivity services will be of prime importance.

Users will require seamless and continuous coverage whilst in transit. Interconnectivity, hand-off and roaming processes are required to enable full mobility between networks so that network selection can be fully automated. Failure to cater for these requirements will prevent the correct operation of data and voice services.

Neutral host providers will charge for access to their network services, but the charging mechanisms are likely to be complex and will require new billing systems to deliver and manage performance sufficiently for MNOs and other users.

It should be noted that neutral host providers require a high level of internal governance and controls: asset ownership, revenue flows and operational responsibilities and accountability are crucial areas to cover when defining company structure. Neutral host networks are critical systems that require seamless, efficient, and reliable delivery. End user customers will demand very high levels of service, resilience, and performance and this will need to be reflected in the operators' internal processes and structure.

Access to spectrum

In the UK, spectrum is managed by Ofcom. Larger spectrum blocks for commercial use are auctioned to the MNOs, enabling them to deliver national coverage. More recently, access to shared spectrum has been made available, and routes to access this are described below.

For bodies that are not telecoms firms or mobile operators, there are routes available to being able to use 5G spectrum. This is frequently found in certain areas such as shopping centres and sports arenas.

JOTS – Joint Operators Technical Specifications

The JOTS forum publishes technical specifications to enable mobile operators and their partners to deploy high-quality shared wireless solutions for the benefit of their customers.

Joint Operators Technical Specification Forum

The aim of the JOTS forum is to specify the performance, coverage and reliability of wireless systems that are shared by mobile operators. The JOTS forum specifications should be referenced by equipment providers and installers when they deploy shared radio solutions for the benefit of mobile operators and their customers.

Over the past twenty years, this original JOTS document has evolved to include GSM, UMTS and LTE and stands as a reference and a framework covering the design and test requirements (non-commercial) of indoor Distributed Antenna Systems (DAS).

JOTS Distributed Antenna System specification

The JOTS Distributed Antenna System (DAS) specification sets out all the technical requirements for all types of third-party in-building solutions based on Distributed Antenna System solutions.

JOTS Neutral Host In-Building specification

The JOTS Neutral Host In-Building (NHIB) specification sets out all of the technical requirements for shared in-building solutions using small-cell base stations.[4]

5G Delivery over radio networks

Radio Access Network (RAN) - A radio access network (RAN) is the part of a mobile network that connects end-user devices, like smartphones, to other parts of a network through radio connections.

The RAN is the final link between the network and the phone. It is the visible piece and includes the antennae we see on towers, on top of buildings or in stadia, plus the base stations. It is a major component of modern telecommunications with different generations of mobile networking evolving from 1G through 5G.

	C-RAN *"Centralisation"*	**vRAN** *"Virtualisation"*	**O-RAN** *"Disaggregation"*
Baseband hardware	Proprietary BBU Centralised	COTS-based BBU May be centralised	
Baseband software	*Proprietary software*	**Virtualised functions** Proprietary software	**Virtualised functions** *Open interface software*
Radio hardware	Proprietary RRU		**COTS-based RRU**
Fronthaul (BBU-RRU) interface	Proprietary interface		**Open interface**
Interoperability	Baseband HW/SW and radios must come from the same vendor	Baseband SW and radios must come from the same vendor	**Baseband HW/SW and radios can come from multiple vendors**

Source: STL Partners

Open RAN (O-RAN) – O-RAN is an architecture for building virtualised RAN on open hardware and cloud services. It provides industry-wide standards for RAN (Radio Access Network) interfaces that support interoperation between vendors' equipment and offer network flexibility at a lower cost. The main purpose of Open RAN is to have an interoperability standard for RAN elements including non-proprietary white box hardware and software from different vendors. Network operators that opt for RAN elements with standard interfaces can avoid being stuck with one vendor's proprietary hardware and software.

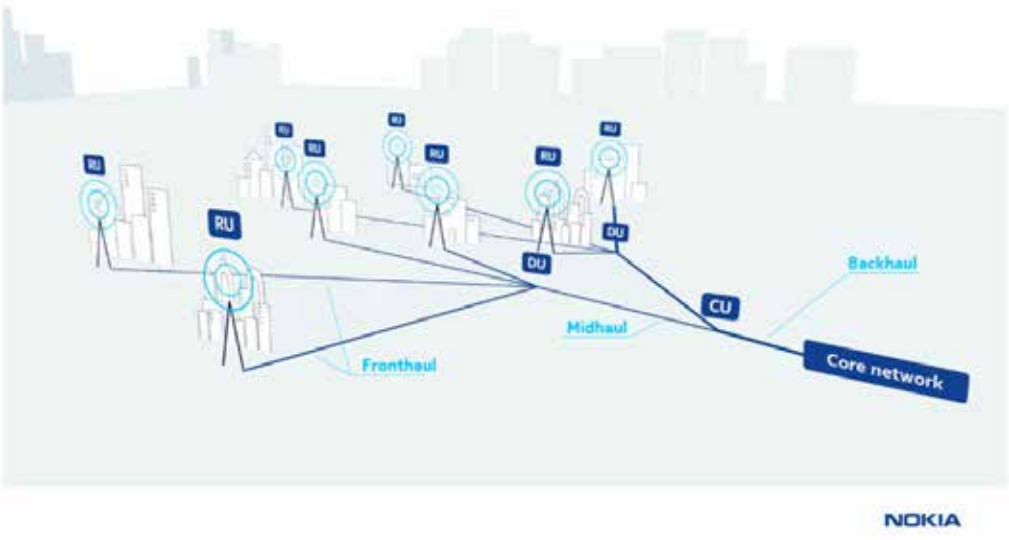
In an Open RAN environment, the RAN is disaggregated into three main building blocks:

- the Radio Unit (RU)
- the Distributed Unit (DU)
- the Centralised Unit (CU)

The RU is where the radio frequency signals are transmitted, received, amplified, and digitised. The RU is located near, or integrated into, the antenna. The DU and CU are the computation parts of the base station, sending the digitalised radio signal into the network. The DU is physically located at or near the RU whereas the CU can be located nearer the Core.

The key concept of Open RAN is "opening" the protocols and interfaces between these various building blocks (radios, hardware, and software) in the RAN. The O-RAN ALLIANCE has defined 11 different interfaces within the RAN including those for:

- Fronthaul between the Radio Unit and the Distributed Unit
- Midhaul between the Distributed Unit and the Centralised Unit
- Backhaul connecting the RAN to the Core



Centralised RAN (C-RAN) - C-RAN is an architectural shift in RAN design, where the bulk of baseband processing is centralised and aggregated for a large number of distributed radio nodes. In comparison to standalone clusters of base stations, C-RAN provides significant performance and economic benefits such as baseband pooling, enhanced coordination between cells, virtualisation, network extensibility, smaller deployment footprint and reduced power consumption.

Virtualised RAN (V-RAN) – V-RAN uses software-based network functions (network functions virtualisation or NFV) instead of hardware-based network functions to virtualise the functionality of RAN systems. RAN virtualisation is required for 5G networks to provide more visibility, automation, and adaptability that hardware-based RANs cannot provide. The ability to scale and intelligently adjust the network to changing conditions is significant when the demands on 5G networks increase both from mobile phone users and, more significantly, IoT devices.

Useful links and further reading

- **“Business Aspects of the Neutral Host Model: The Immersive Video Services Case”**
  https://link.springer.com/chapter/10.1007%2F978-3-030-49190-1_3
- CGI: A paper on “Neutral host networks and how to support them” via Telecoms.com
  https://telecoms.com/opinion/neutral-host-networks-and-how-to-support-them
- GSMA: A paper on “Enabling Neutral Host – Network Economics”
  https://www.gsma.com/futurenetworks/wp-content/uploads/2018/09/180920-CCS_GSMA_Case_Study-FINAL_NE-Modelling-removed.pdf
- GSMA: Mobile communications industry body
  https://www.gsma.com/
- MobileUK: The Joint Operators Technical Specifications Forum
  https://www.mobileuk.org/jots
- Ofcom: resources for ECC, including how to register
  https://www.ofcom.org.uk/phones-telecoms-and-internet/information-for-industry/policy/electronic-comm-code
- Open Grid Alliance: Industry body promoting grid compute models
  https://www.opengridalliance.org/
- STL Partners: Specialist consultancy and research company with useful resources about 5G, edge and RAN topics
  https://stlpartners.com/
- techUK: Advanced Communications Services Working Group
  https://www.techuk.org/communications-infrastructure-programme/advanced-communications-services-working-group.html
- The 5G Infrastructure Public Private Partnership
  https://global5g.org/

Annex 1

Electronic Communications Code

The electronic communications code (the Code) is set out in Schedule 3A of the Communications Act 2003. It is a set of rights that are designed to facilitate the installation and maintenance of electronic communications networks.

The Code confers rights on providers of such networks and on providers of systems of infrastructure to install and maintain apparatus on, under and over land and results in considerably simplified planning procedures. It contains a mechanism for companies wishing to become "wireless infrastructure providers" to obtain the same rights as Telecoms firms and MNOs with respect to installations and planning.

The Code confers "code rights" on a person with Code powers. A code right is a right to:

- install electronic communications apparatus on, under or over the land;
- keep installed apparatus, which is on, under or over land;
- inspect, maintain, and operate apparatus;
- carry out any works on the land to enable apparatus to be installed and maintained;
- gain access to land to maintain or operate apparatus;
- connect to a power supply;
- interfere with or obstruct a means of access to or from the land (whether or not any electronic communications apparatus is on, under or over the land); and
- lop or cut back any tree or other vegetation that could interfere with apparatus.

In connection with these rights, the Code allows persons to whom the Code applies to:

- construct and maintain electronic communications networks and infrastructure (such as ducts, cabinets and poles) on public highways without the need to obtain a street works licence to undertake such works;
- construct communications infrastructure which is classified as 'permitted developments' under Town and Country Planning legislation (such as certain types of masts, poles and cabinets) without the need to apply for planning permission;
- in the event that agreement cannot be reached with the owner or occupier of private land, the Code allows an operator to apply to the Court to impose an agreement which confers the Code right being sought or for the Code right to bind the landowner or occupier; and
- claim compensation from a local authority in circumstances where that local authority has obstructed access to electronic communications apparatus in certain stipulated circumstances.

Acknowledgements

This paper was drafted by Duncan Clubb, Head of Digital Infrastructure Advisory at CBRE, and Alan Nunn, Consulting Director, Telecoms at CGI (and Chair of techUK's Advanced Communications Services Working Group).

Thanks to the following contributors and organisations:

- Jonathan Freeman, Strategic Growth and Regulatory Director, Cellnex UK and Vice-Chair of techUK's Communications Infrastructure and Services Council
- Members of techUK's Advanced Communications Services Working Group

techUK
FOR WHAT COMES NEXT

About techUK

techUK is a membership organisation that brings together people, companies and organisations to realise the positive outcomes of what digital technology can achieve. We collaborate across business, Government and stakeholders to fulfil the potential of technology to deliver a stronger society and more sustainable future. By providing expertise and insight, we support our members, partners and stakeholders as they prepare the UK for what comes next in a constantly changing world.

linkedin.com/company/techuk

@techUK

youtube.com/user/techUKViews

info@techuk.org