MPIRICAL

Innovating Telecoms Training

Exploring 5G

Student Guide

www.mpirical.com

Exploring 5G

Student Guide

This course has been independently assessed and accredited by both the ITP (Institute of Telecommunications Professionals) and the CPD Standards Office (Continuing Professional Development). Their marks of excellence awarded to Mpirical provides you with the recognition that the learning you have completed conforms to best practice and is appropriate for inclusion in a formal training record.

Through successful completion of this class it will provide 3 CPD Learning Points.

Our certified courses mean you can be certain you're receiving the highest quality training. Both the ITP and the CPD use independent specialists to regularly assess all our products and services. Including our Learning.

The ITP is the UK's leading independent institution for people who work in telecommunications. It is dedicated to promoting the professional development of members through Professional Registration, training, mentoring and qualifications. The ITP collaborates with regulators, government associations and other leading bodies on projects that are important to the future of the industry. It has worked with businesses including; BT, Cable & Wireless Worldwide, Alcatel-Lucent, Huawei and Vodafone.

By providing independent accreditation, the CPD improve the quality of continuing professional development. Their marks of excellence for training providers demonstrate to individual professionals that the learning activity conforms to best practice and is appropriate for inclusion in a formal record.

Find out more about your certification at:
www.mpirical.com/certified-telecoms-training

Mpirical classes have been developed in accordance with the technical specifications published by the 3GPP. As such the 3GPP have granted Mpirical Limited the right to use the 3GPP logo to identify specifications, compliant products and services.

First published by Mpirical Limited in 2020

Contents

Figures

1 5G Commercial Considerations

1.1 Driving Factors for 5G

1.1.1 The Anatomy of a 2020 Mobile Network

Prior to looking specifically at the commercial considerations of a 5G network, we should first of all take this opportunity to review what a typical mobile service provider network actually looks like today in 2020. Surprisingly, as we begin the mass rollout of 5G technology, we still find the legacy technologies of GSM (Global System for Mobile communications), GPRS (General Packet Radio Service) and UMTS (Universal Mobile Telecommunications System) playing a part in the overall technology mix.

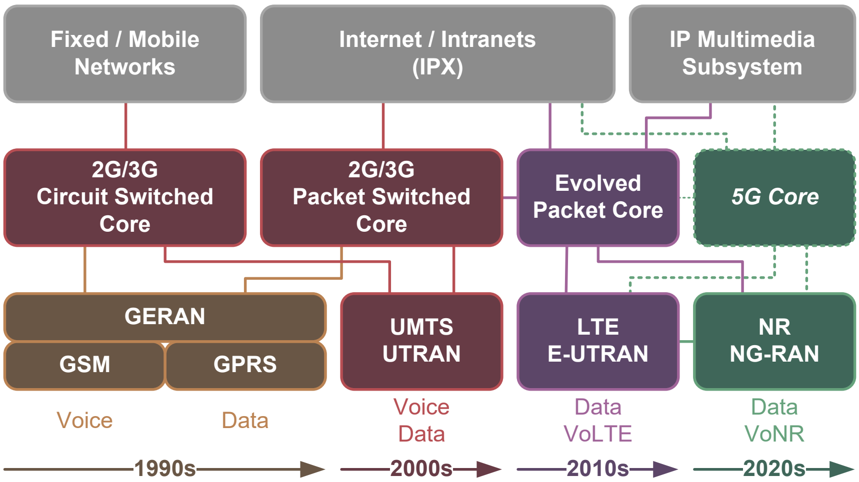


Figure 1-1 Anatomy of a 2020 Mobile Network

What Figure 1-1 doesn't show is the importance of each of the technology generations with regards to network operation or in other words, which technologies are carrying the traffic and generating the revenues? One way of visually representing this is to show how much of the limited and extremely valuable RF (Radio Frequency) spectrum is currently being used. Please note that Figure 1-2 only includes spectrum allocated in the UK although this is indicative of the current state of play in many countries. Furthermore, Figure 1-2 only includes half of the FDD (Frequency Division Duplex) allocated spectrum to take into consideration that this is comprised of paired frequencies.

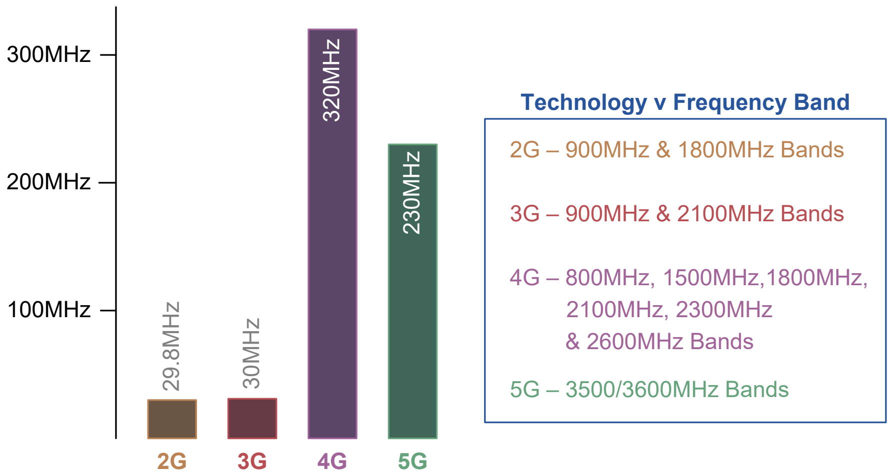


Figure 1-2 UK Radio Frequency Spectrum Usage 2020

So even though 2G and 3G are still in operation today, it would be fair to say that 4G is providing the real “heavy lifting” in terms of network capacity. Likewise, even though there is a significant amount of RF spectrum allocated to 5G, the reality is that the number of actual cell sites supporting the technology is still relatively low. Furthermore, spectrum allocation is still based upon relatively high frequencies which in general terms equates to lower coverage.

1.1.2 Why is 2G and 3G Still Operating?

Clearly as we move into the 2020s and 5G becomes mainstream, mobile service providers are going to want to “sunset” these legacy 2G and 3G networks to enable them to re-farm the spectrum, cut the cost of maintaining end of life equipment and rationalize their networks towards a virtualized service based architecture. However, doing so may not be that simple.

Figure 1-3 Sunsetting Legacy Technology

Coverage

There are still locations, particularly in rural areas, where 2G is the only signal available and thus in these regions, significant investment would need to be made to upgrade the infrastructure. Until this becomes economically viable, or government grants become available, it looks like 2G will still need to support voice services and limited data.

Telemetry

There are still a significant number of telemetry or machine-to-machine subscriptions that are in operation today using mobile radio equipment based on the 2G and 3G technologies. Therefore, depending upon the SLA (Service Level Agreement) that may be in place, mobile service providers may be liable when switching out devices for the more expense 4G and 5G variants. Examples include much of the current smartmeter network and eCall, the pan European accident / breakdown network for vehicles.

Age Profile

There are a significant number of subscribers, often of an older generation, who do not require a smartphone and would rather continue to use the “simpler” legacy feature phones to make calls and send texts. Therefore, simply switching off 2G would effectively “drop” these customers which, even though they may not be a net high revenue customer, it could leave the mobile service provider in an embarrassing position.

1.2 5G Standardization

1.2.1 From IMT Advanced to IMT 2020

Figure 1-4 From IMT Advanced to IMT 2020

In 2008, the ITU-R (International Telecommunications Union – Radiocommunications) published its technical performance requirements for IMT (International Mobile Telecommunications) Advanced which was later marketed as 4G or LTE (Long Term Evolution). In 2012, they embarked on a programme to develop IMT 2020, setting the stage for the development and deployment of 5G networks in the 2020s. In 2017, they published their final document outlining the minimum technical performance requirements and in so doing, identified three broad use cases and eight key capabilities.

Figure 1-5 IMT 2020 – Use Cases

1.2.1.1 Enhanced Mobile Broadband

The term Mobile Broadband was first introduced with the advent of the 3G technologies of HSPA (High Speed Packet Access) and later HSPA+ to describe the human centric use case of accessing multimedia content, services and data over a mobile network. Current forecasts indicate that demand will continue to increase, leading to Enhanced Mobile Broadband which incorporates new application areas and requirements such as super-fast access in dense areas, broadband access everywhere and support for higher user mobility.

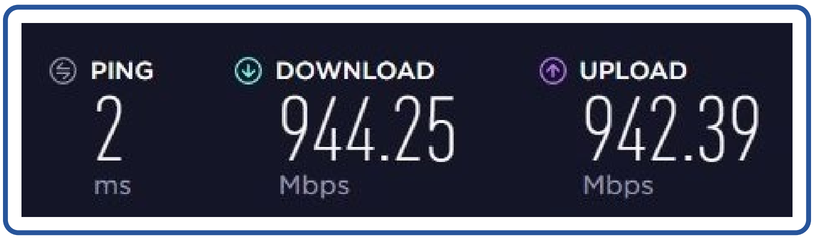


Figure 1-6 Enhanced Mobile Broadband

A key trend with each iteration of mobile technology is an increase in data rate. 5G is no exception, with user experienced speeds of over 1Gbps headlined in the telecoms press.

1.2.1.2 Massive Machine Type Communications

The second use case defined by the ITU-R characterizes MTC (Machine Type Communications) or more commonly termed, IoT (Internet of Things). In this use case, billions of low powered devices and sensors will wirelessly transmit relatively small volumes of delay tolerant data.

Although for many people, the increased data rates associated with 5G will be the biggest attraction, the reality is that 5G was designed to serve a huge population of end devices. The reason for this is attributed to the predicted growth of CIoT (Cellular Internet of Things) based applications and devices which will proliferate the 5G network.

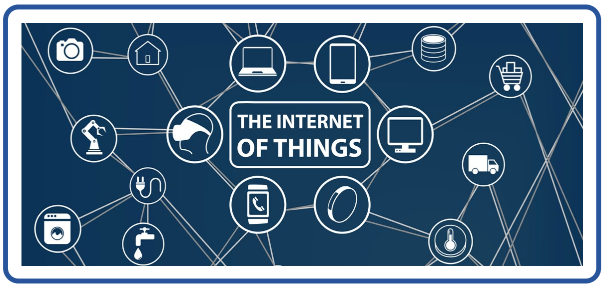


Figure 1-7 Massive Machine Type Communications

1.2.1.3 Ultra Reliable Low Latency Communications

This use case has several stringent capability requirements in terms of latency and mobility and therefore covers applications such as wireless control of industrial manufacturing, remote medical surgery and transportation safety systems. It is also used to describe the applications associated with Critical IoT and V2X (Vehicle to Everything).

Figure 1-8 Ultra Reliable Low Latency Communications

1.2.2 Performance Capabilities

For each of the three defined use cases, the ITU-R also defined eight key capabilities which they scored either Low, Medium or High in importance. These are illustrated in Figure 1-9.

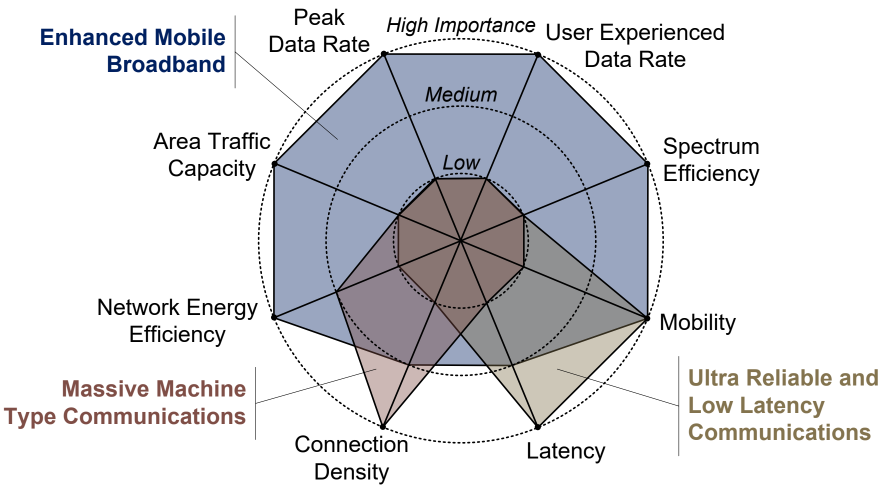


Figure 1-9 IMT 2020 Key Performance Capabilities

Therefore, by applying discrete values to these capabilities, the ITU-R defined the actual requirements for not only 5G technologies (IMT 2020) but also for the previous 4G mobile technology (IMT Advanced). This is illustrated in Figure 1-10.

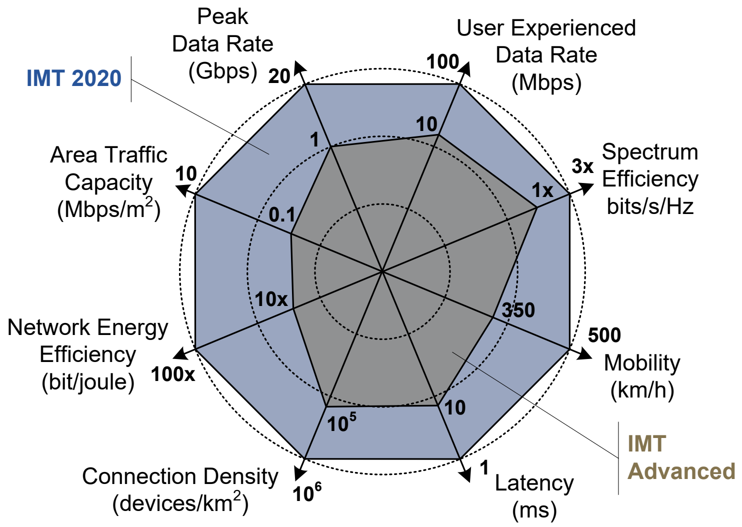


Figure 1-10 IMT Advanced v IMT 2020

1.2.3 The Role of the 3GPP

Whereas the ITU-R defined the capabilities of a 5G system in the broadest of terms, it is the 3GPP (Third Generation Partnership Project) which is responsible for writing the technical specifications which in essence define how the technology will actually work. The 3GPP, as the name would suggest, has been at the heart of standardizing mobile telecommunications for the past twenty years or so, since the introduction of 3G technology or UMTS. In so doing, they publish numerous technical specifications covering all aspects of the technology. These are broken down into Releases with 5G appearing at Release 15. However, some of the performance capabilities set out by the ITU-R only became achievable as a result of the Release 16 / Release 17 specifications.

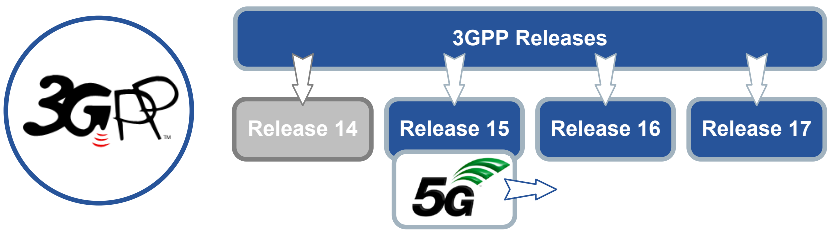


Figure 1-11 The Role of the 3GPP

1.3 5G Use Cases

The deployment of 5G technologies throughout the 2020s is set to deliver services and capabilities far beyond the traditional voice, text messaging and mobile broadband offered through the earlier technologies. As users of the technology, 5G will be more than just the new smartphone we hold in our hand. Instead, it will encompass all aspects of our life as the use cases of MMTC (Massive Machine Type Communications) and URLLC (Ultra Reliable Low Latency Communications) become more and more prevalent.

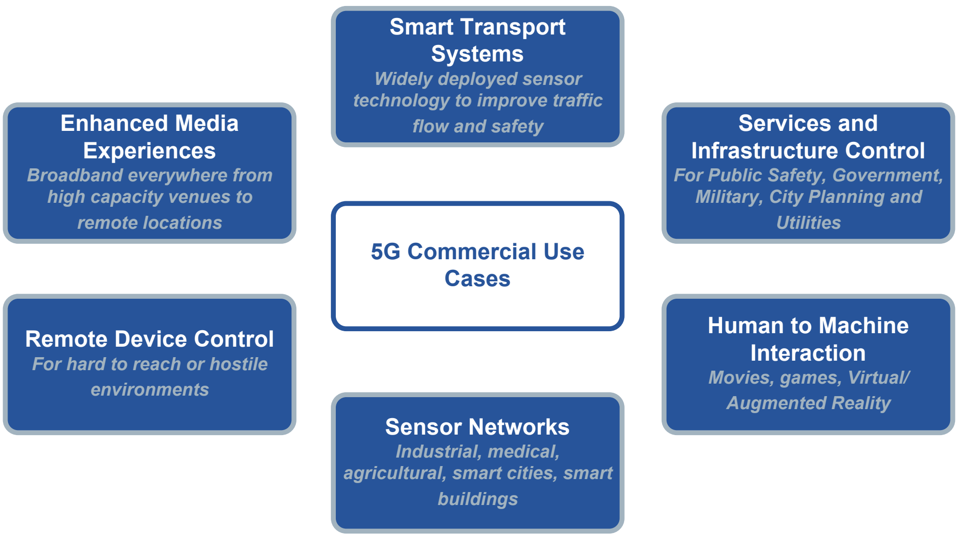


Figure 1-12 Commercial Use Cases in 5G

Turning our attention to five specific scenarios, we can illustrate why 5G is likely to make the most significant changes in the evolution of mobile telecommunications.

1.3.1 5G Small Cells

When planning a cellular network, two terms best describe the overall objective; Coverage and Capacity.

Coverage

During the early 1990s as second generation networks became more widespread and the use of mobile phones was not limited to business users, the service providers were focused on coverage. In other words, their primary objective was to make sure the radio signal emanating from the base stations reached the users; wherever they may be. This, in simplistic terms, was achieved by placing the base stations or cell towers on the tops of buildings or hills and turning transmit powers as high as they were permitted. These were later termed Macro Cells.

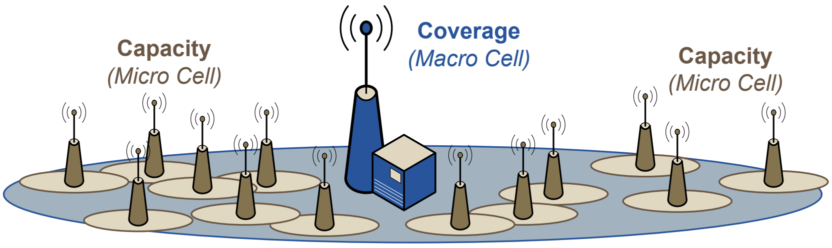


Figure 1-13 Coverage v Capacity

Capacity

As the networks became busier with more and more subscribers, the coverage based approach to planning had to change as there was too high a demand for the limited resources provided by one base station i.e. there were too many people under the coverage area of a cell. As such, the service providers had two options, either allocate more radio resources to the cell or add more cells. The first option however does not scale as the service providers only had a limited amount of RF spectrum to use across their entire network. Instead, they had to carefully reuse this limited resource by making the cells smaller (less coverage area) and reusing the RF spectrum. Thus, with fewer subscribers under the coverage area of a cell, demand could be better catered for.

In reality, both Coverage and Capacity needs to be taken into consideration, with service providers today using a combination of “Large / Macro” cells that provide coverage and “Small / Micro” cells that support capacity.

In the case of 5G, this principle remains the same. If we need to increase capacity within the network, then we need more and more small cells despite the fact that 5G technology is more efficient than its predecessors.

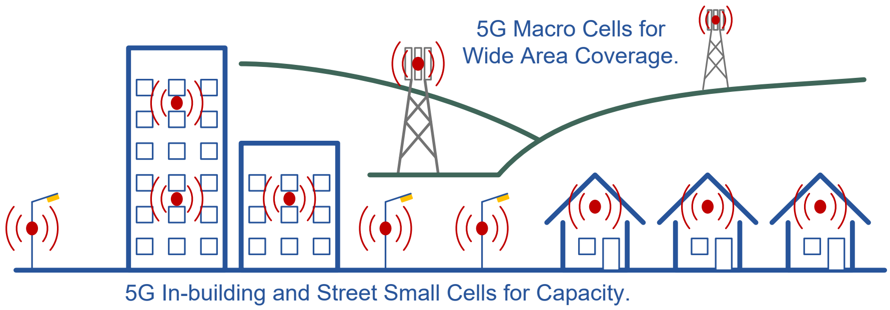


Figure 1-14 5G Small Cells

As mentioned previously, capacity can also be improved by allocating more RF spectrum and with the rollout of 5G, we are going to see new frequency bands coming into play. These however will be at much higher radio frequencies (GHz) than the ones in general operation today (MHz). At these higher frequencies, the radio wave will not travel (propagate) as far as lower frequencies and become more easily attenuated as it passes through walls etc. As such, the coverage area of cells operating at these higher frequencies (e.g. 6GHz) is significantly less than those operating at say 900MHz. This may first appear to be a negative facet however, that is not necessarily the case as these higher radio frequencies are better suited for reuse within a small geographic area and thus improve network capacity overall. Furthermore, there is more available spectrum at these higher frequencies which aids higher bandwidth.

1.3.2 5G Wireless Backhaul

Cell densification is one of a number of solutions for fulfilling the coverage and speed requirements of IMT 2020. However, providing suitable backhaul connectivity (the means by which information is transferred between the bases stations and the core network) to a potentially huge number of small cells can be problematic for a variety of reasons. For instance, simply providing physical connectivity to the device through fibre optic transmission can be difficult and expensive. Wireless backhaul is a logical and cost effective solution, which does not necessarily have to be based on traditional microwave technology. Thus, 5G as a backhaul technology is considered to be a viable solution. Figure 1-15 illustrates the concept of using 5G as a wireless backhaul technology; in this case it has been assumed that a different radio frequency band has been used to mitigate against interference.

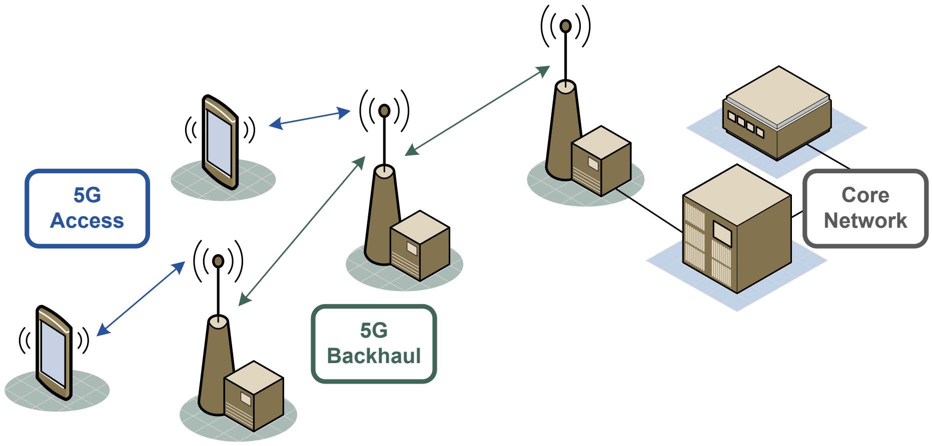


Figure 1-15 5G Wireless Backhaul

1.3.3 5G for Fixed Wireless Access

In traditional fixed telecommunication networks, the physical connection between the residential property and the local exchange, often termed the “final mile” is one of the more expensive parts of the network. In many developed countries, this infrastructure is many years old and commonly comprised of copper cables. Although this medium has been great at supporting voice calls, it struggles to provide the high data throughput rates we now demand from our Internet service providers. As such, networks have been ungraded and where financially viable, fibre optic cables have been brought as close as possible to the properties. Although this has significantly improved connectivity, the Gbps data rates we want are still unlikely to be available until a fibre optic cable can be terminated in the building itself.

With the introduction of 5G, this situation has changed as there is now a commercially viable alternative to running fibre optic cables. The data rates achievable with 5G mean the same technology can be used to span the “final mile”, a technique termed FWA (Fixed Wireless Access).

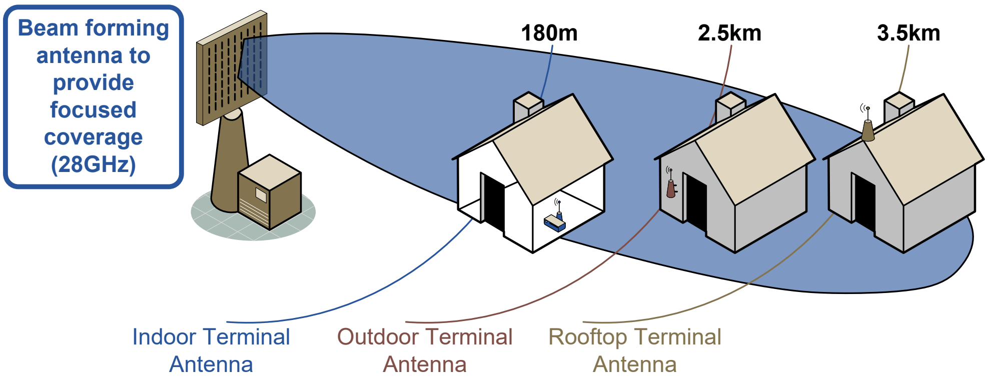


Figure 1-16 5G for Fixed Wireless Access

One advantage of 5G FWA over traditional mobile networks is that the subscriber (building) will not be mobile and as such, it is easier to plan the radio connection between the base station and the residential unit at the property. Furthermore, equipment only needs to be installed at properties taking the service rather than in the case of fibre optic, having to pass numerous properties on the way to connect one at the end of the street!

Figure 1-16 illustrates some typical distances which can be achieved using a 5G radio operating at circa 28GHz and beam forming technology.

5G Fixed Mobile Convergence

The convergence of fixed and 5G mobile services is not limited to FWA alone which itself falls within the overall 5G Fixed Mobile Convergence ecosystem. This enables closer working between the fixed access network infrastructure and 5G core networks enabling seamless mobility in terms of providing common services to both mobile and fixed devices. As such, a 5G Residential Gateway will be able to connect back to the core network via either a 5G wireless network (FWA) or via a fixed or wireline connection based on say fibre optic. This is illustrated in Figure 1-17.

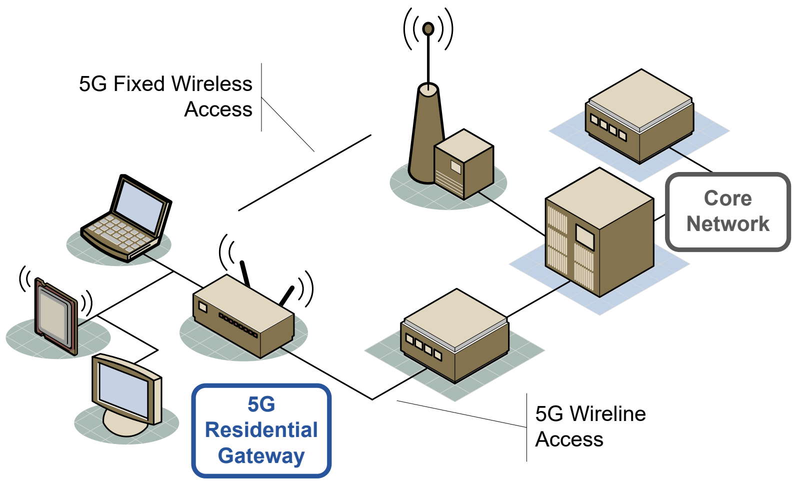


Figure 1-17 5G Fixed Mobile Convergence

1.3.4 Private 5G Networks

5G as a private networking technology is a particularly interesting proposition; with data rates in the order of Gbps, latency of sub 1ms and reliability exceeding 99.999%, 5G technology outperforms 4G and Wi-Fi but also provides advantages beyond simple performance.

For example, connection density of devices has improved tenfold over 4G, making 5G attractive for Massive IoT use cases and therefore, ideal for private deployments in a variety of scenarios with high numbers of end devices – enterprise, industry and healthcare etc. Moreover, 5G has inherent support for Network Slicing, making network management and control easier for the private network operator. Support for MEC (Multi access Edge Computing) is also beneficial, particularly in an industrial setting where AR (Augmented Reality) and VR (Virtual Reality) use cases are gaining traction.

Finally, the QoS (Quality of Service) mechanisms that 5G supports, albeit similar to 4G, now extend to delay critical services which require guaranteed bitrates (a prime requirement for industrial applications requiring deterministic message transfer).

Beyond the media hype surrounding 5G however, deploying the technology in a private network setting is not without its challenges. The technology is new, which obviously has implications on cost. Where early Private 5G trials are enabled with the financial support of equipment vendors, mobile service providers and academia, the price point for 5G will need to come down if Private 5G is to become a commercial success. That said, as 5G technology as a whole sees more widespread deployment, the cost should reduce.

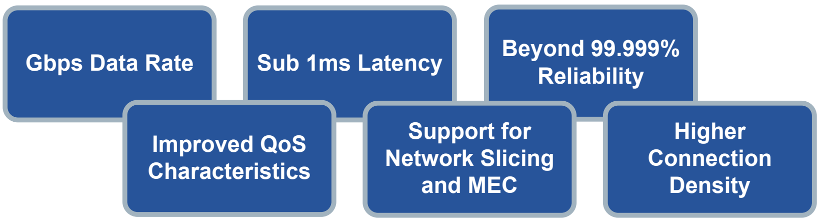


Figure 1-18 Advantages for 5G Private Networks

Perhaps one of the most significant challenges is spectrum or more specifically, in which band should Private 5G be deployed. At present, there's three options available: Licensed, Unlicensed and Shared.

Licensed

Although operating the Private 5G network within a licensed spectrum band will mitigate the risk of interference, the cost associated with this option is potentially too great. This however is scenario dependent; for an enterprise deployment of Private 5G, it's unrealistic to think that that enterprise would be in a position to purchase exclusive access to radio spectrum. However, for a national infrastructure company it may be an option.

Interestingly, spectrum regulators tend to understand these issues and consequently, some common-sense approaches to licensed spectrum usage are beginning to emerge. For example, the UK regulator OFCOM is permitting access to licensed spectrum owned by mobile service providers which may be either a) unused or b) not deployed in a specific geographical area. For a very small fee and under strict regulation (transmission range, geography, etc), a Private 5G network could potentially access this spectrum. Furthermore, regulatory authorities are also investigating making additional frequency allocations available, although these will be limited to small geographic areas and will need to align to international frequency bands otherwise compliant equipment will not be commercially available.

Note that the mining industry in particular could benefit from this approach since mines deep underground or geographically isolated are not going to interfere with public mobile networks.

Figure 1-19 Private 5G Spectrum Options

Unlicensed

With this approach, the Private 5G network will operate in the traditional unlicensed band which is shared by other technologies such as Wi-Fi (2.4GHz & 5GHz). Interference is obviously an issue, but the cost of the spectrum is free. In addition, in the USA, Europe and other parts of the world, freeing up a further 1.2GHz band of spectrum around 6GHz is currently being explored.

Shared

Around the globe today, spectrum regulators are assigning spectrum bands allocated to shared spectrum operation and reserved for a particular technology such as 4G or 5G. For a typically low fee, a Private 5G network can be deployed within this shared spectrum band, benefitting from a reduced risk of interference and potentially large channel sizes. Interference may still however be an issue and as such, coordination between private networks may be required.

1.3.4.1 Deployment Models

From the perspective of the 3GPP, Private 5G networks are termed “Non Public Networks”. Two high level deployment models exist:

Standalone Non Public Networks

This is a Private 5G network which is self-contained; the network owner is responsible for the RAN (Radio Access Network) and Core elements, with no interaction with a mobile service provider.

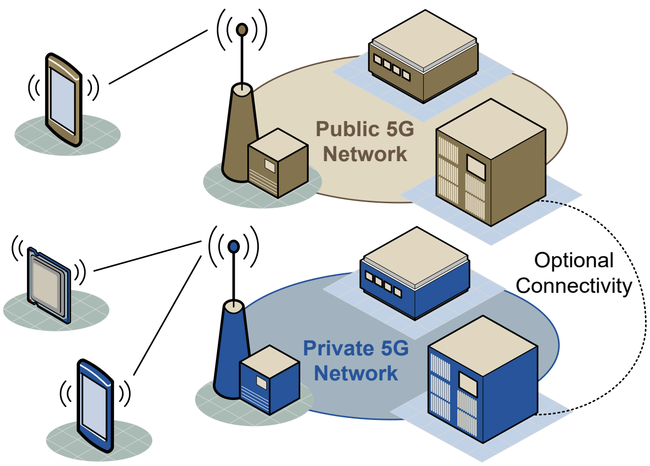


Figure 1-20 Standalone Non Public Network

Public Network Integrated Non Public Networks

With this approach, a mobile service provider will assist in the deployment of the Private 5G network. This could include RAN sharing, RAN and Core Sharing or full outsourcing, possibly instantiated as a network slice within the Public 5G network.

On a global basis, even Private 4G networks are a relatively new phenomenon, with deployment lessons still being learned. Although real-life deployments of Private 5G are out there, challenges related to cost and spectrum usage, coupled with a lack of knowledge within the private network industry are potentially going to be limiting factors. However, specific use cases such as Industry 4.0 could unlock a huge market for Private 5G, particularly since 5G can provide communication services which outperform the private networking technologies that are currently available today. Providing a combined approach between standardization bodies, spectrum regulators and equipment manufacturers can be achieved, Private 5G could mean the reach of 5G is extended way beyond the traditional mobile space.

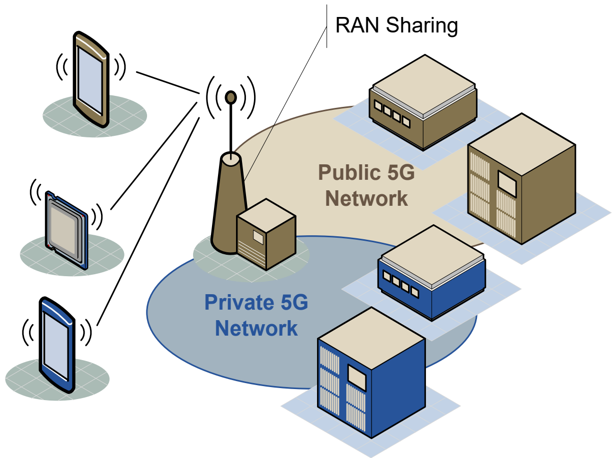


Figure 1-21 Public Network Integrated Non Public Network (RAN Sharing Example)

1.3.5 5G and Autonomous Vehicles

Future vehicles will not only be connected, they will also be autonomous. These connected and autonomous vehicles will require a new ecosystem with specific requirements capable of supporting the new workloads and satisfying real-time requirements. It will include not just the vehicles themselves and the required road infrastructure, but also a 5G network infrastructure and the supporting data network located in the cloud.

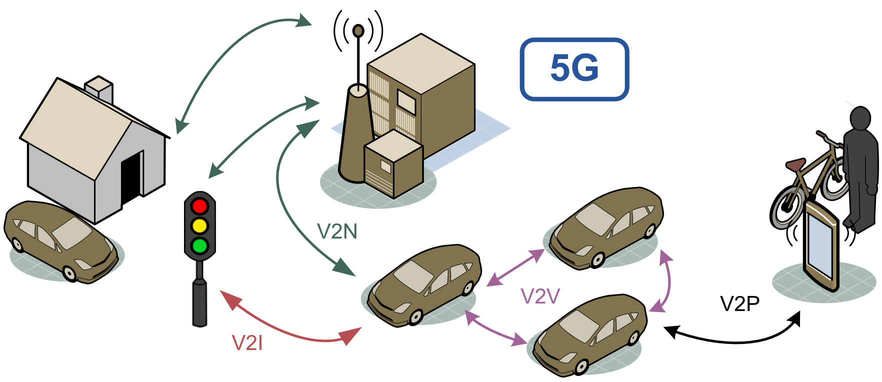


Figure 1-22 5G and Autonomous Vehicles

1.3.5.1 Driving Factors

The move towards fully autonomous vehicles is thought to be evolutionary, driven mainly by the automotive industry stakeholders. Automated driving will shape the future of mobility and also contribute to improving the quality of modern life.

Facilitating this automotive evolution and providing the specific requirements, particularly the need for very high reliability and low latency, an advanced communication technology is required to provide the network infrastructure. This will then allow connected autonomous vehicles to communicate with each other, share and access data, and provide the different services required for the driving process.

A 5G system deployed with V2X (Vehicle to Everything) capability will provide the required performance characteristics, a flexible air interface, the ability to coexist with multiple radio access technologies and network virtualization.

The connected autonomous vehicle evolution and the introduction of 5G has provided a number of opportunities for drivers, passengers, car manufacturers and service providers:

- Automated Driver Assistant Services - including driving assistant, parking assistant, self-parking and automated / unmanned vehicles.
- Telematics - such as traffic alerts, weather conditions and road conditions.
- Infotainment - music and video streaming.
- Location Based Services - supported by AV (Augmented Reality) and VR (Virtual Reality).
- Vehicle Management - remote and automated software / firmware updates will be a critical process for autonomous vehicles. A failure in software will not be acceptable.

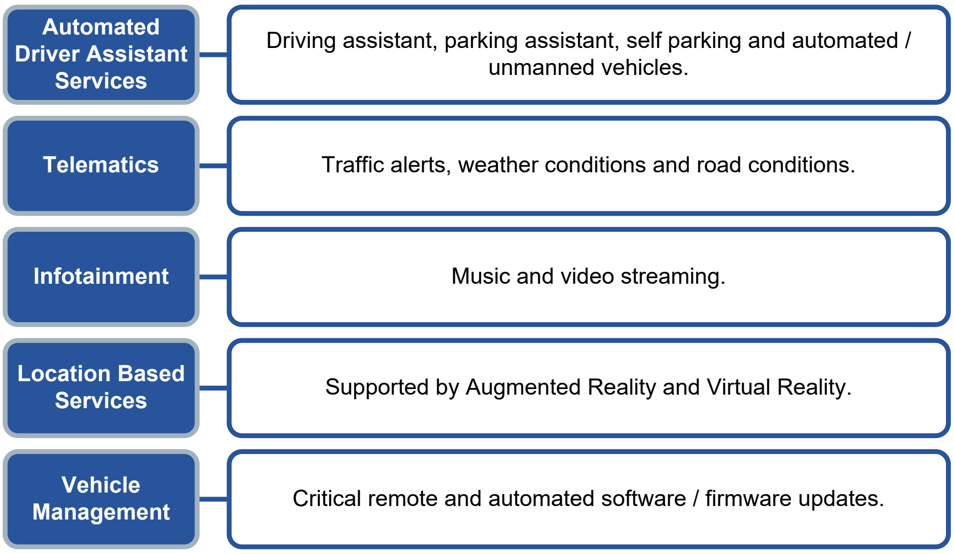


Figure 1-23 5G and Autonomous Vehicles Opportunities

1.3.5.2 Autonomous Vehicles

There are several key aspects to consider when studying autonomous vehicles; automated driving, road safety and traffic efficiency, digitization of transport and logistics information, intelligent navigation and the information society.

Figure 1-24 Autonomous Vehicles

Automated Driving

There are six defined levels of vehicle automation, as illustrated in Figure 1-25. Automated driving can occur outside of 5G coverage and without V2X communication, however this is restricted to the lower levels of automation, and not full automation. This is due to the need to access data servers for high-resolution digital mapping. In addition, local V2X communication capability will assist with inter vehicle communication and an improved understanding of the traffic situation.

Automation Level	Description	Driver in the Loop	Secondary Tasks	Final Fallback	Example
0	No Automation	Yes	None	Driver	Lane Departure Warning, Full Collision Warning
1	Assisted	Yes	None	Driver	Lane Keeping Assist, Adaptive Cruise Control
2	Partial Automation	Yes	None	Driver	Parking Assistance
3	Conditional Automation	No (optional)	Specific	Driver	Traffic Jam Chauffeur
4	High Automation	No (optional)	All	Automation	Parking Garage Pilot
5	Full Automation	No (optional)	All	Automation	Robot Taxi

Figure 1-25 Levels of Automation

Road Safety and Traffic Efficiency

Information collected through V2X for vehicle automation will enable warnings and increased situational awareness to be provided to the driver. This will greatly improve road safety, reducing collisions, injuries and fatalities.

Digitisation of Transport and Logistics Information

The digitisation of the transport and logistics infrastructure includes the collection and use of large amounts of data. The collection and subsequent analysis and use of this big data will assist with increasing the efficiency of the transport logistics chain. 5G and V2X will be a key enabler in assisting with this.

Intelligent Navigation

Digital mapping and geo-positioning are key components to providing navigation guidance to drivers via navigation systems. These will be further enhanced by the additional information provided by other vehicles, road authorities and traffic management centres. More value-added services beyond the current online servers will be provided, complementing navigation. The navigation systems themselves will continue to evolve. They will integrate information from additional sensors such as cameras and radars on the vehicle itself as well as from other vehicles and infrastructure in the vicinity.

Information Society

The connectivity performance that vehicle passengers have come to expect and demand in the home will continue with the vehicle. The use of partial or high automation will allow this to be extended to the driver, who will have the ability to use this connectivity whilst the vehicle is driving autonomously.

1.4 5G and Artificial Intelligence

5G and AI (Artificial Intelligence) has a “symbiotic” relationship in that each enhances the capabilities of the other. For example, the integration of MEC (Multi access Edge Computing) into the 5G SBA (Service Based Architecture) alongside the significant reduction in latency brought about by the new 5G air interface will enable Augmented Reality and Virtual Reality to become mainstream across both enterprise and consumer markets.

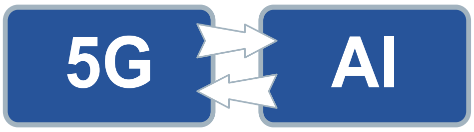


Figure 1-26 5G and AI

For example, video data from the smart devices will be rapidly transferred across the network, enabling machine learning algorithms operating on the edge to interpret the information and return “intelligent” responses in near real-time.

Conversely, the complexities of the 5G radio have necessitated the need for Machine Learning and AI in order to dynamically optimize the millions of 5G small cells, not to mention the issues of introducing Massive MIMO (Multiple Input Multiple Output) and beam forming / beam steering. Network Slicing, a key enabler introduced in 5G, can also benefit from AI as the algorithms will enhance automation and adaptability by processing real-time information on network performance, quality of service provision etc, thereby supporting a more efficient, dynamic provisioning of a network slice.

1.4.1 What is Artificial Intelligence?

There is no universal definition of Artificial Intelligence however it can be expressed as the ability for a machine to perform cognitive functions which we associate with the human mind, such as perceiving, reasoning, learning and problem solving.

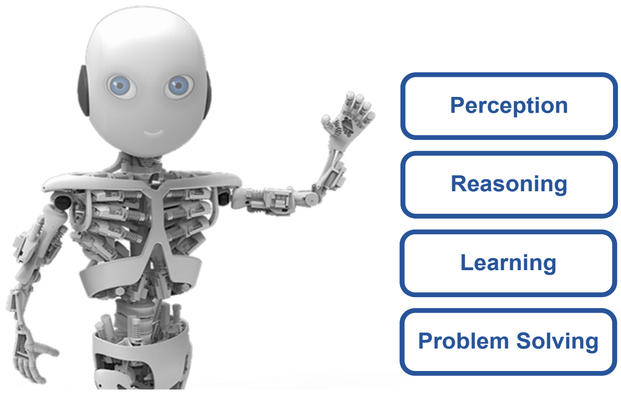


Figure 1-27 Artificial Intelligence – A Simple Definition

A conceptual view of AI can be described as an "intelligent agent" which in turn consists of three main elements; sensors, operational logic and actuators. The sensors collect raw data from the environment whilst the actuators act to change the state of the said environment. Clearly, the power of the intelligent agent is the operational logic as it must, for a set of defined objectives, take the input data from the sensors and provide output to the actuators. These could take the form of recommendations, predictions and decisions that in turn can influence the state of the environment. This concept is illustrated in Figure 1-28.

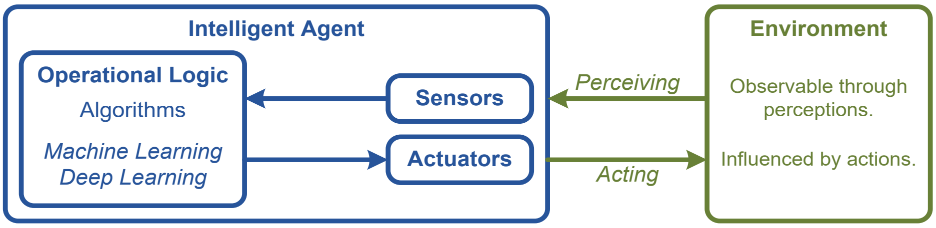


Figure 1-28 Artificial Intelligence – A Conceptual View

The operational logic element is comprised of various algorithms which are used to build a model of the environment and then to interpret this model to make predictions, recommendations and decisions etc. These algorithms form the basis of what we now describe as Machine Learning and Deep Learning.

Machine Learning

Most recent advances in AI have been achieved by applying Machine Learning to very large data sets. Machine Learning algorithms detect patterns and learn to make appropriate predictions and recommendations by processing data and experiences, rather than receiving explicit programming instructions. Moreover, the algorithms also adapt in response to new data and experiences to improve their efficiency over time. In other words, Machine Learning provides both predictions and prescriptions.

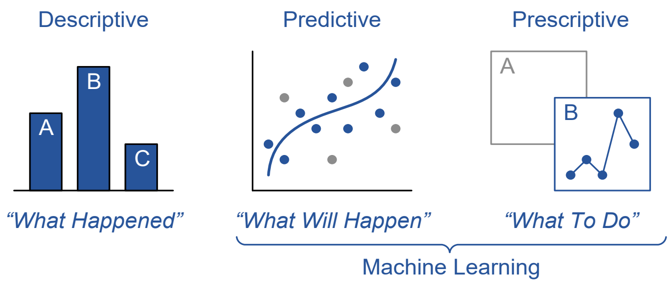
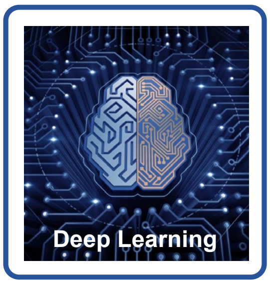


Figure 1-29 Machine Learning – Predictions and Prescriptions

Deep Learning



Deep Learning can be defined as a type of Machine Learning that is able to process a wider range of data resources, requires less data pre-processing by humans and can often produce more accurate results than the more traditional Machine Learning techniques discussed earlier. In Deep Learning, interconnected layers of software based calculations known as “neurons” form a neural network which can ingest vast amounts of input data and process it through multiple layers that learn increasingly complex features of the data. That is, the network will make a decision about the data, learn if its decision is correct and then use what it has learned to make better decisions in the future with regards to new input data

1.4.2 5G Driving Artificial Intelligence

There are a number of reasons why the widespread rollout of 5G technology in the 2020s is being seen as a significant factor in the deployment of AI in more and more everyday scenarios. 5G technology centres around three key use cases in terms of the services it aims to promote. Two of these are IoT (Internet of Things) centric and as such, they will aid in the generation of input data for the AI applications.

Therefore, with these capabilities in mind, the 5G network is able to provide a highly efficient communication network to support a wide range of AI applications.

Figure 1-30 5G Driving Artificial Intelligence

5G / AI Convergence on a Chip

With the anticipated mass adoption of devices operating on not only 5G but also other wireless technologies associated with IoT, it is anticipated that these radio modules will be converged with neural network processing or AI. In so doing, low power / low cost devices will become available for many mass market AI applications.

Billions of Devices / Lakes of Data

Furthermore, each of the billions of IoT centric devices will be capable of adding to the wealth of data available to AI applications to use for training, analytics and predictive analysis. Moreover, the ubiquitous connectivity offered through 5G should ensure a constant stream of data wherever the devices may operate.

High Speed Low latency Communications

As we have already discussed, 5G will bring about significant improvements in data throughput rates and reductions in latency. These performance advantages enable 5G to support AI DevOps pipeline workloads from data ingest and preparation to model building, training and serving low latency real-time streaming environments.

1.4.2.1 Case Study – Augmented Reality

A good example of how 5G and AI can operate symbiotically is the area of AR (Augmented Reality). Consider the example of visitors to a museum or gallery using their smartphone or perhaps smart glasses to be provided with near real-time information about the artefact or piece of art as they move around. This could include biographies of the artist, history of the item, short video of the sculpture being created etc.

For this to happen, from the 5G technology perspective, we need to be able to move information or data bidirectionally with minimal latency. This will be the images (video) of what the user is seeing which needs to be sent to the AI platform and the corresponding supporting information (text, audio, graphics, video etc) being returned; all at such speeds that the response arrives in synchronization with what the device is pointing at.

Figure 1-31 5G, AI and Augmented Reality

To achieve this, 5G technology has incorporated a flexible radio interface which is able to dynamically and incredibly quickly schedule resources. Furthermore, 5G network deployment will utilize millions of small cells (small coverage area) especially for in building solutions which will aid in reducing transmission times but also improve available bandwidth. Clearly transmission time is a key performance criteria for AR so as well as the time for the data to span the radio link, we also need to take into consideration the transmission time between the base station inside the museum and the AI application.

If it is necessary to backhaul this information to say another city just to leave the 5G network, not to mention the onward connection across the Internet to a Data Centre hosting the AI application, then we are going to have problems. Instead, we need to bring the AI application as close as possible to the user. In other words, we need to deploy MEC (Multi access Edge Computing), a 5G enhancement which sees applications typically housed in centralized data centres essentially migrated to the network edge and thus closer to the consumer.

1.4.3 Artificial Intelligence Driving 5G

Although AI will significantly benefit from the rollout of 5G, conversely, 5G will need to call upon AI technology if it is going to achieve the performance capabilities frequently touted.

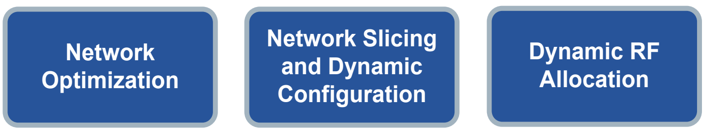


Figure 1-32 Artificial Intelligence Driving 5G

Network Optimization

In order to support the broad range of capabilities associated with 5G, mobile service providers will need to ensure that their networks become self-healing, self-managing, self-securing and self-optimizing. To achieve this state of network "nirvana", AI based applications will need to be deeply imbedded within the network ecosystem. In so doing, they will not only be able to detect patterns of behaviour but also predict outcomes which can be used to allocate resources, switch to standby capability etc as required.

Network Slicing and Dynamic Configuration

Network slicing and dynamic configuration enables mobile service providers to run several "virtual" networks utilizing the same physical infrastructure. With AI support, this can be achieved through the ability to predict and then dynamically configure network capability across the entire network; radio resources, MEC, transport etc.

Dynamic RF Allocation

The rollout of 5G connectivity will often require the installation of millions of small cells and as such the efficient reuse of the radio frequencies allocated to this technology. This, coupled with the added complexity of Massive MIMO and the associated techniques of beam forming and beam steering will require significant intelligence and processing – a purpose built role of AI.

2 Understanding the 5G Network

2.1 The 5G Network Architecture

In order to explain what a 5G network actually looks like, it is appropriate to do so through its evolution from 4G or LTE (Long Term Evolution). As such, we will first need to introduce the terms NSO (Non Standalone Operation) and CUPS (Control and User Plane Separation) before arriving at the full 5G network architecture.

2.1.1 Non Standalone Operation

The 5G networks which were launched during 2019 were based upon a specific approach termed Non Standalone Operation. This was primarily driven by the mobile service providers as a means of introducing 5G technology, and the associated increases in data throughput, without having to roll out an entire 5G architecture. As such, this approach introduced a new 5G RAN (Radio Access Network) however this was connected to the 4G RAN or E-UTRAN (Evolved – Universal Terrestrial Radio Access Network) and 4G Core or EPC (Evolved Packet Core). This is illustrated in Figure 2-1.

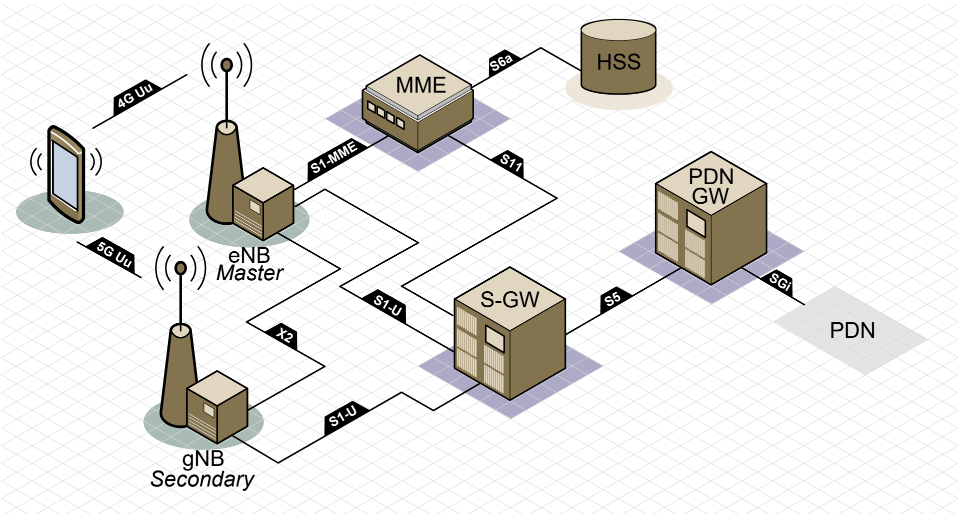


Figure 2-1 5G Non Standalone Operation

Therefore, 5G cells or gNB (New Radio Node B) were added to the connection as and when available to provide a “data boost”. Thus, the 4G cell or eNB (Evolved Node B) acted as the “Master” and as such was in control of the gNB which simply acted as the “Secondary” radio. Note, when a device is communicating with two distinct cells at the same time e.g. 4G and 5G, it is said to be operating in Dual Connectivity Mode.

Option 3

A number of different transition options were proposed to aid mobile service providers in their roll out of 5G. The methodology described above is referred to as Option 3 or EN-DC (E-UTRA – New Radio Dual Connectivity) and in this case, the 5G cell or gNB is more accurately termed an en-gNB.

Returning to Figure 2-1, it can be seen that User Plane traffic (emails, web pages, video etc.) can be transferred between the device and the PDN (Packet Data Network) via a host of different routes. For example, the traffic can be split at the S-GW (Serving Gateway) and sent to both the 4G and 5G cells for onward delivery across their respective radio interfaces. Alternatively, the traffic can be sent to one of the base stations and split here thus making use of the X2 interface. One favoured approach is referred to as Option 3x which supports two data flows; one via 4G entirely and the second splitting the bearer at the en-gNB and sending a proportion of the traffic up the eNB. This is illustrated in Figure 2-2.

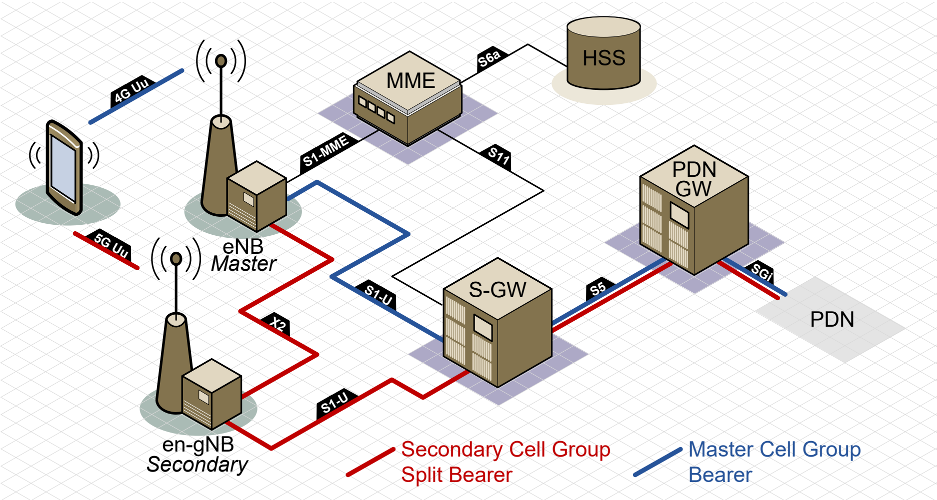


Figure 2-2 Option 3x

2.1.2 Control and User Plane Separation

The next stage in the evolution of the 5G architecture is to add the new core network or 5GC (5G Core). This is based upon an SBA (Service Based Architecture) utilizing the benefits of NFV (Network Functions Virtualization) and SDN (Software Defined Network) to improve network utilization and system elasticity, not to mention reduce both OPEX and CAPEX.

The benefits of virtualizing the core network are not restricted to 5G alone and as such, many mobile service providers are also adopting this approach for the 4G core or EPC, along with separating the signalling and traffic functionality of key network elements; a process termed CUPS (Control Plane and User Plane Separation). Figure 2-3 illustrates the CUPS architecture which in essence splits both the S-GW and PDN-GW (Packet Data Network Gateway) into two separate functions; one dealing with the Control Plane or signalling and the other the User Plane or traffic.

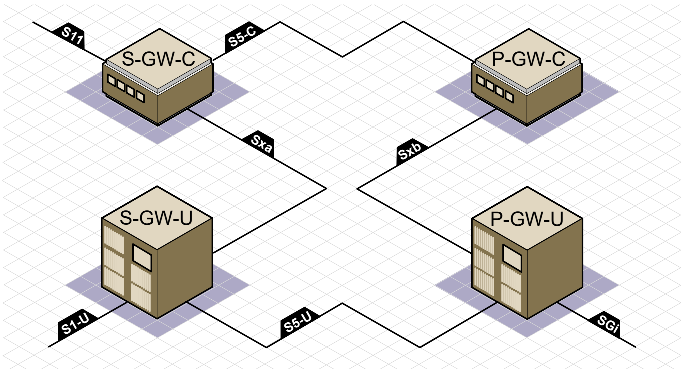


Figure 2-3 Control and User Plane Separation

This approach brings about several advantages not least:

- Reducing Latency – by selecting User Plane nodes closer to the RAN (Radio Access Network).
- Supporting Increased Data Traffic – by adding additional User Plane nodes without increasing the number of Control Plane nodes.

- Independently Scalable – with the advent of IoT, the relationship between User Plane and Control Plane will differ significantly.
- SDN (Software Defined Networking) – deliver User Plane data more efficiently by utilizing SDN functionality.
- 5G Convergence – improved interworking between the 4G and 5G core networks and the ability to deploy, the EPC, 5GC and IMS within the same Data Centre.

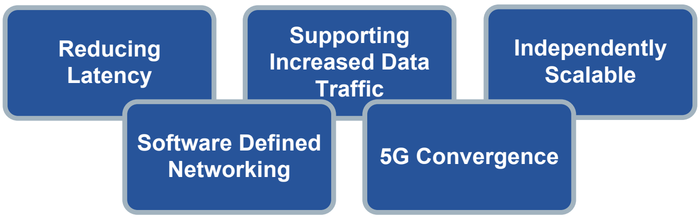


Figure 2-4 CUPS Benefits

2.1.3 5G Network Architecture

Figure 2-5 illustrates the simplified architecture of a 5G network. This is comprised of a number of key network functions rather than physical elements due to the virtualized nature of the 5GC.

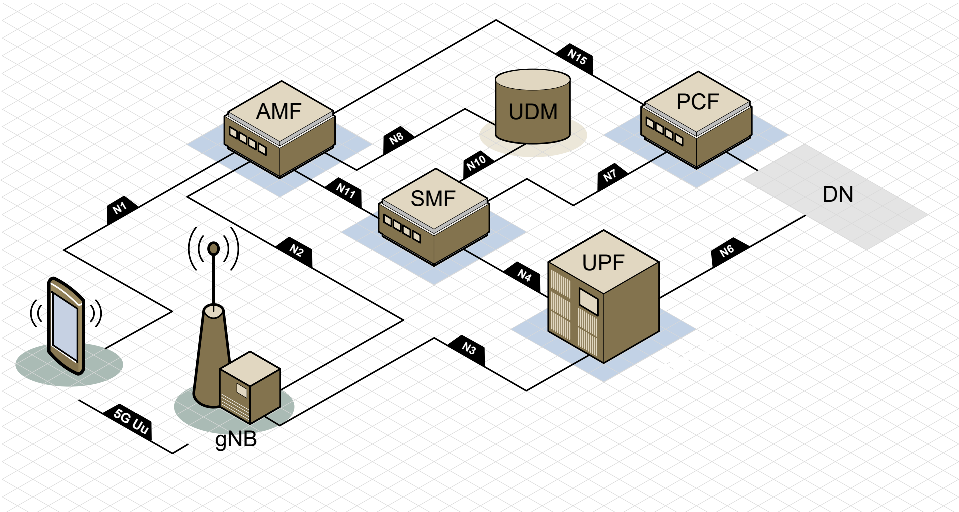


Figure 2-5 5G Network Architecture

Access and Mobility Management Function

The primary task of the AMF (Access and Mobility Management Function) is to manage the mobility of the subscriber. In particular, the AMF will play a key role in the device registration process (including security), as well as track the device's mobility for reachability and paging purposes.

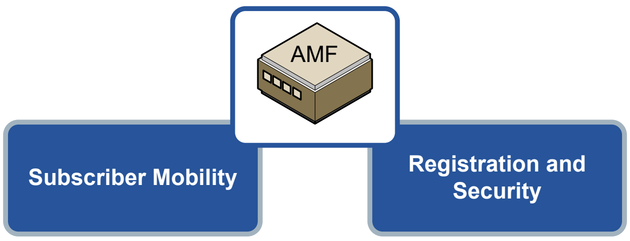


Figure 2-6 AMF Functionality

Session Management Function

The SMF (Session Management Function) manages the PDU (Protocol Data Unit) Sessions associated with an individual subscriber. As such, the SMF will routinely interact with the PCF (Policy Control Function) to determine exactly which Data Networks the device is allowed to connect to, as well as the QoS (Quality of Service) profile it can expect to be allocated. The SMF will also liaise with the UPF (User Plane Function) to establish PDU Session connectivity. Finally, the SMF will be responsible for allocating a suitable IP address to the device (IPv4 or IPv6), assuming an IP PDU Session is active.

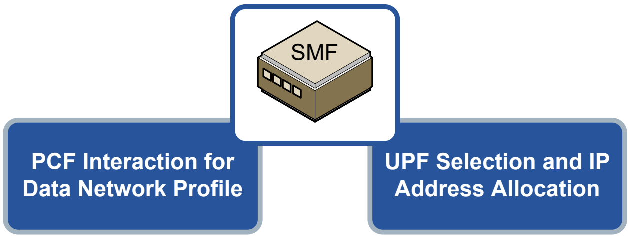


Figure 2-7 SMF Functionality

User Plane Function

The UPF is the only network element within the core that handles User Plane traffic. As can be seen in Figure 2-5, the UPF will have User Plane connectivity to both the gNB and also the DN (Data Network). As such, the UPF is responsible for ensuring data is placed on the correct downlink QoS Flow, as well as ensuring that any dynamic policy rules are suitably enforced. Moreover, as handovers take place in the RAN, the UPF will remain the core network anchor point for the User Plane traffic.

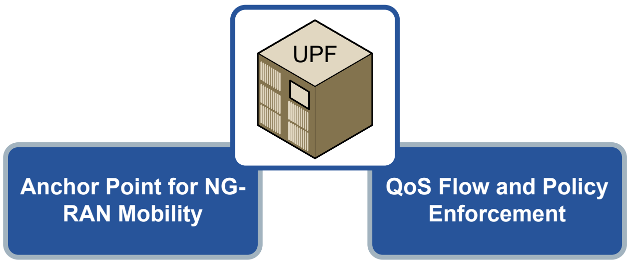


Figure 2-8 UPF Functionality

Unified Data Management

The UDM (Unified Data Management) is essentially a central repository of subscriber information which can be used by several different network elements. Information stored in the UDM includes access restrictions, mobility restrictions, QoS profiles and roaming permissions.

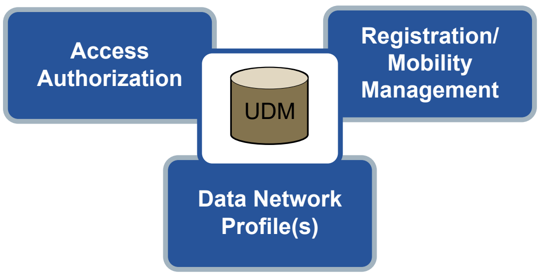


Figure 2-9 UDM Functionality

Policy Control Function

The PCF can provide policy decisions to the AMF and SMF on a dynamic basis. These policy decisions or often based on conditions being active within the network, such as congestion, subscriber geolocation or billing. In addition, the DN can also potentially provide session level information, such as the subscriber wishing to make a call or view a video. In all of these scenarios, the PCF can dynamically change the way in which the subscriber receives their service, from establishing the correct QoS Flows in the network to completely terminating a PDU Session.

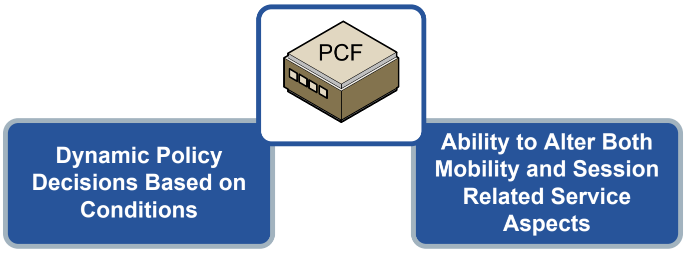


Figure 2-10 PCF Functionality

2.1.4 Virtualization in the RAN and Core

2.1.4.1 Virtualization

Network virtualization is a key technology which has impacted the architecture of today's core networks and to a degree, access networks also. In essence, service providers are seeking to deploy the various elements of their core network as VNFs (Virtualized Network Functions) running over a shared NFVI (NFV Infrastructure). As such, what was once potentially a piece of dedicated hardware performing the role of a specific network element now becomes a software based implementation. The NFVI essentially provides all of the compute, storage and network resources that the VNFs require. This concept is shown in Figure 2-11.

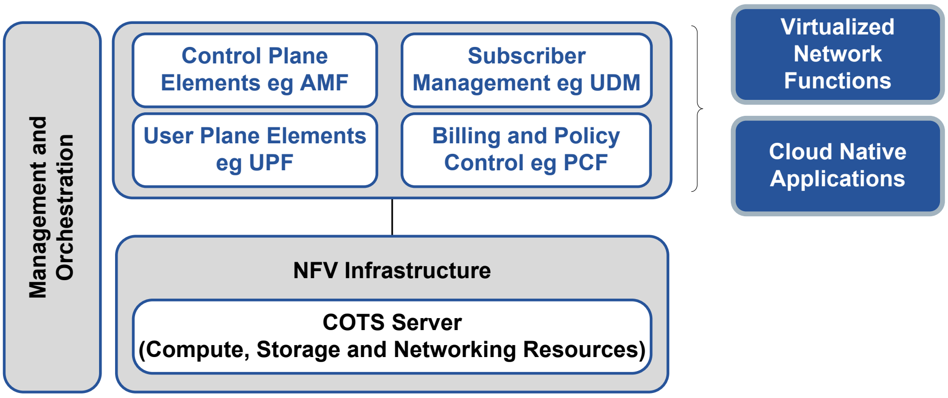


Figure 2-11 Virtualization

Note that MANO (Management and Orchestration) performs a critical role in the virtualized infrastructure, ensuring that VNFs can be managed (instantiated, modified and removed) accordingly. This provides huge benefits to the service provider because it allows the network to be highly flexible and highly scalable. In essence, if more or less capacity is required in the network, rather than deploying or removing hardware (which can take weeks or even months), capacity can be added or removed in a matter of minutes (since the network functions are software based and as long as the NFVI resources are available, it is just a matter of allocating resources accordingly).

Note that service providers who do not adopt the VNF approach to virtualization tend to opt for the newer approach based on “Cloud Native Applications”. Although the concept is essentially the same, the implementation is different (and arguably better) than the VNF approach.

2.1.4.2 Centralized RAN

C-RAN (Centralized RAN), like that of NFV, is not restricted to 5G networks and in fact was introduced several years ago. However, it too will play a significant part in the deployment of the 5G RAN.

The 5G RAN can support a number of different deployment configurations. A centralized deployment would be based on C-RAN, which sees a single CU (Central Unit) in control of one or more DU (Distributed Units). A non-centralized deployment would be based on a traditional, monolithic architecture. Key terminology is explained in Figure 2-12, although it should be noted that different standards bodies use different terminology.

CU **Centralized Unit**	Centralized processing, housing higher layer elements of the gNB protocol stack (one CU to multiple DUs)
DU **Distributed Unit**	Houses lower layers of the gNB protocol stack, with reduced processing requirements
RRU **Remote Radio Unit**	Deals with physical layer transmission and reception, supporting technologies such as Carrier Aggregation and Massive MIMO (Multiple Input Multiple Output).

Figure 2-12 C-RAN Terminology

Ultimately, the CU will take processing requirements away from the DU and house them in in a centralized location. As such, DU hardware is cheaper to deploy (additional advantages are also involved).

Figure 2-13 outlines the terminology used to define the various legs of the transport network that provides connectivity between the C-RAN elements.

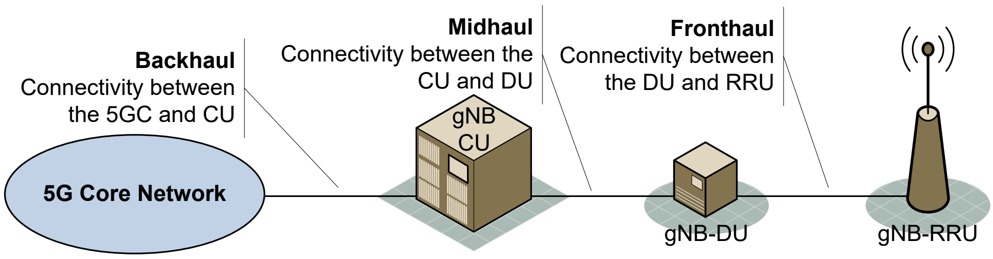


Figure 2-13 Basic C-RAN Architecture

Beam Forming, Beam Steering and Massive MIMO

Massive MIMO (Multiple Input Multiple Output) is considered to be a key technology used within the 5G RAN. In principle, Massive MIMO is a radio antenna technology, with individual 5G gNBs featuring large arrays of antennas working together to improve both coverage and also increase data rate.

Figure 2-14 illustrates how, by using massive antenna arrays, narrower beams of radio energy can be directed towards a subscriber or group of subscribers (contrary to regular antennae which radiate over greater coverage areas). This has the effect of providing much improved coverage for those subscribers within the beam, hence increasing data rates.

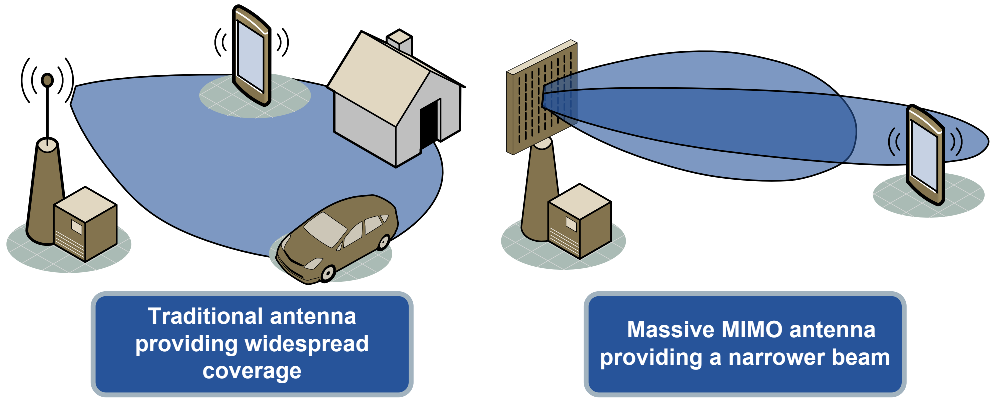


Figure 2-14 Beam Forming and Massive MIMO

In Figure 2-15, the beam forming concept is taken one step further, in which the beam of coverage is steered in order to first of all focus on a "target", then move in step with the movement of the "target". This requires the device to continuously provide feedback to the antenna array with information on the received signal, since factors such as the subscriber turning their head, road traffic passing by or street furniture may require the antenna to adjust its beam steering so that optimal coverage is provided. Figure 2-15 shows that beam steering can potentially accommodate a variety of subscriber mobility scenarios, although fast moving devices will pose significant challenges. Furthermore, deploying a RAN based upon beam forming and beam steering is still very expensive when compared to the more traditional techniques.

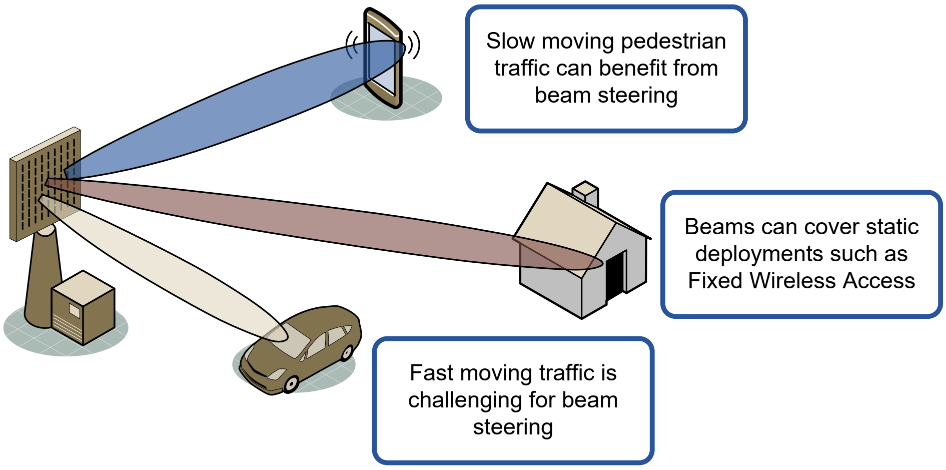


Figure 2-15 Beam Steering

Increased Spectrum

One of the key factors in delivering the data rates and capacity requirements of IMT 2020 is the release of more licensed spectrum for mobile service providers to use. Certain parts of the RF (Radio Frequency) spectrum are already severely congested however, higher up the frequency range, there are more resources available. Consequently, 5G will utilize parts of the RF spectrum that are new to mobile networks. Unfortunately, these new bands pose their own challenges, as outlined in Figure 2-16.

When one considers factors such as geography, topography, regulation and population density, it is clear that no single frequency range will be able to satisfy all of the criteria of IMT 2020. As such, 5G mobile service providers will operate their 5G services across a range of frequency bands.

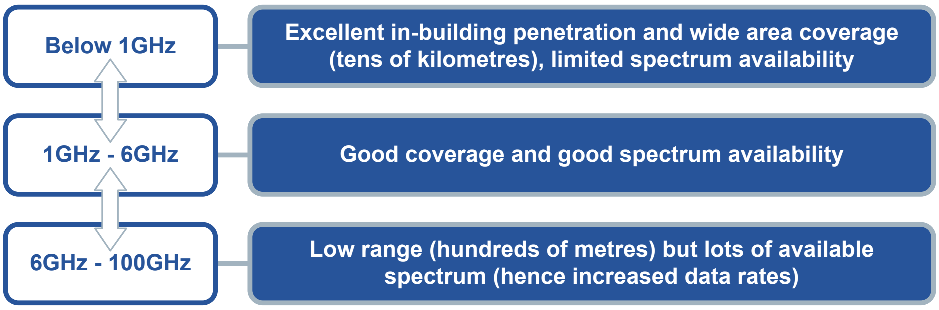


Figure 2-16 RF Spectrum Usage

Gaining additional spectrum is not that straight forward and in the majority of countries, it is acquired through expensive auctions which generates hundreds of millions of pounds / dollars for the respective governments. Mobile service providers will re-farm spectrum from legacy technologies however during the roll out of 5G, this spectrum is still required to support 2G, 3G and 4G. Furthermore, the high bandwidth capabilities of 5G are amongst other things achieved by allocating wide bands of spectrum still further driving demand for this limited resource.

2.1.4.3 5G Service Based Architecture

Due to the fact that the 5G specifications included support for NFV, the 3GPP also defined a service based representation of the 5G System (Figure 2-17), relative to Control Plane entities. This SBA (Service Based Architecture) is a conceptual view of how virtualized network elements communicate with one another via an SBI (Service Based Interface) (as opposed to traditional reference points). Note that all Control Plane network functions in 5G will use the Service Based Architecture approach for communication.

With the Service Based Architecture, a NF (Network Function) can offer its NF services to other network functions via an open API (Application Programming Interface). For example, the virtualized AMF network function can use an API to access specific subscriber information from the virtualized UDM network function.

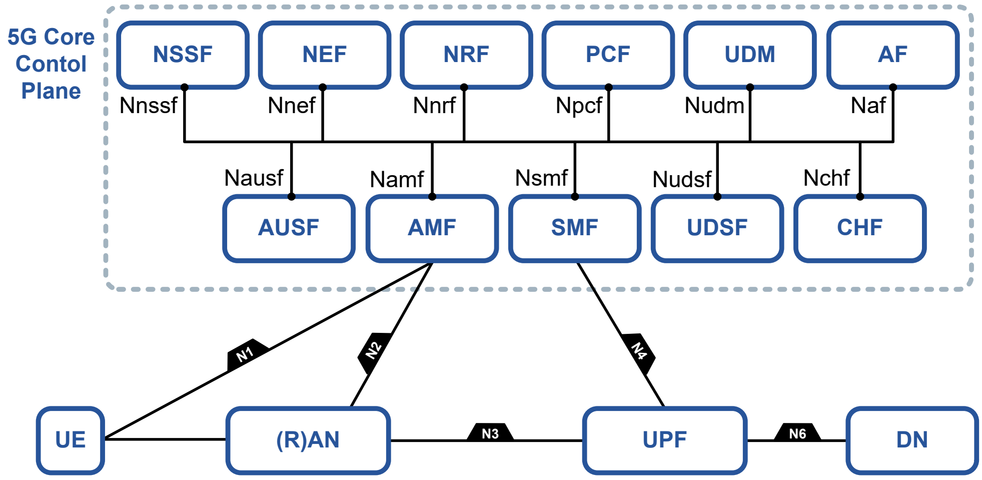


Figure 2-17 5G Service Based Architecture

Note, Figure 2-17 includes several other network functions not previously discussed. Some of these will be discussed later in this section however more information can be found in NetX.

2.1.5 2020 Network Architecture

Figure 2-18 illustrates the combined architecture of a 4G and 5G network after transitioning through Non Standalone Operation and Control and User Plane Separation. Note, that the SMF and the P-GW-C functions have been co-located as to the UPF and P-GW-U. As such, devices will be able to switch between 4G and 5G yet remained anchored through the same core network function thereby maintaining their IP address and remaining connected to whichever service they were using.

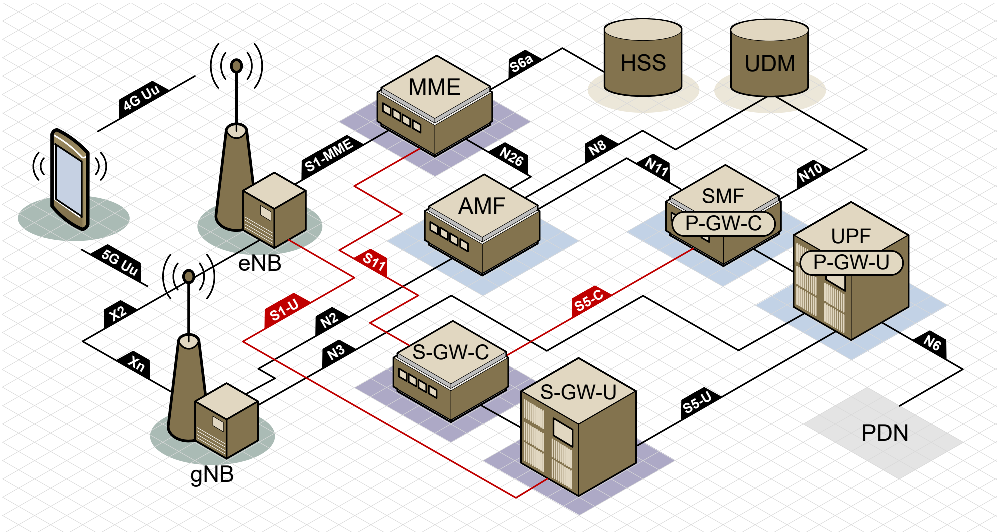


Figure 2-18 5G / 4G Network Architecture

There are a number of interfaces / reference points not included on the diagram. For example, the technical specifications support the connecting of the E-UTRAN (4G) to the 5GC. In this case, the eNB will need to be upgraded and as such is referred to as an ng-eNB. For a complete representation of the network, see NetX.

2.2 Transferring Data in 5G

2.2.1 PDU Sessions

The 5G network has been designed to provide PDU (Protocol Data Unit) connectivity between the devices and one or more DN (Data Network) such as the Internet or Intranets. To facilitate this, PDU Sessions are created which span the 5G network.

Within a PDU Session, several QoS Flows, each with different QoS characteristics, can be established to accommodate the varying transport requirements of the User Plane data (no two QoS Flows will have the same QoS characteristics). For Internet connectivity, this may be a simplistic approach where only one QoS Flow is required to provide a Best Effort service. However, for voice services, separate QoS Flows may be established to handle the differing transport requirements of signalling and voice.

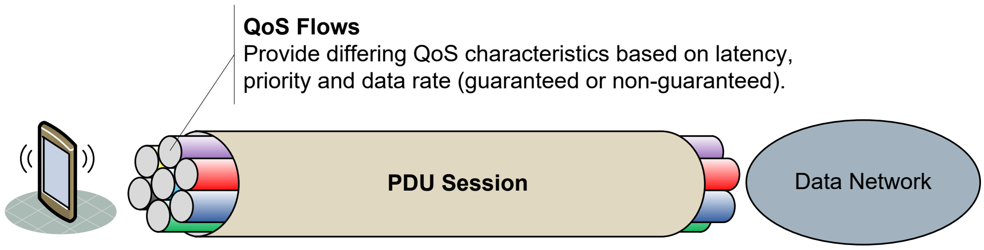


Figure 2-19 PDU Sessions and QoS Flows

PDU Sessions are able to support different traffic types e.g. IP (IPv4 and IPv6), Ethernet and Unstructured Data or NIDD (Non IP Data Delivery). The latter being primarily focused towards IoT (Internet of Things) type applications where the size of the data packets that needs transferring is typically very small with respect to the additional overhead required to “wrap” it inside an IP datagram.

2.2.2 Transferring Data via the UPF

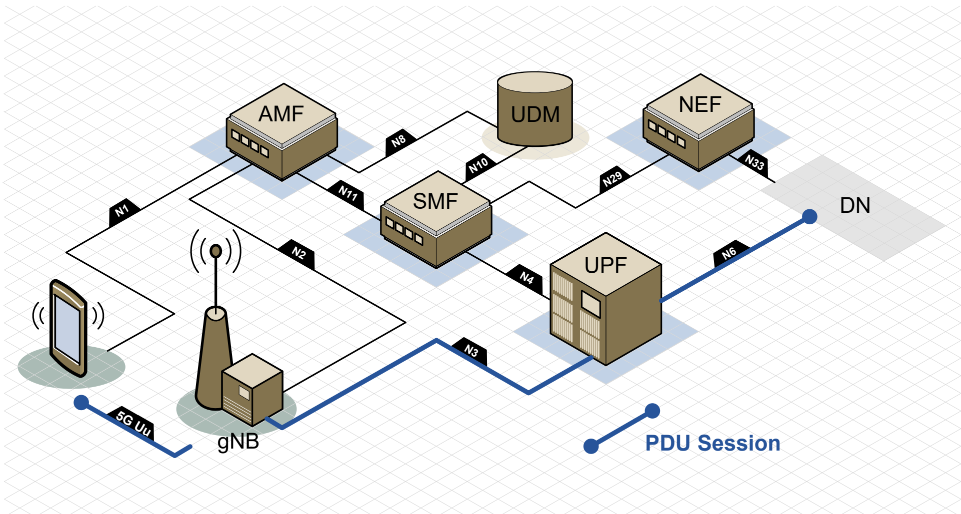


Figure 2-20 Transferring Data via the UPF

Once a PDU Session has been established, it can be described as a logical connection running between the device and a UPF / DN. As previously described, the PDU Session will be defined by the way the traffic is treated as it spans the 5G network (QoS) as well as the type of traffic it is carrying (IP, Ethernet etc.) Figure 2-20 illustrates a PDU Session spanning the 5G network and thus the route traffic will take as it is transferred between the device and the DN.

2.2.3 Transferring Data via the NEF

Alternatively, NIDD can be delivered by the NEF (Network Exposure Function) which in turn exposes a series of Northbound APIs across N33. As such, an IoT AS (Application Server) can pass Non IP Data to the NEF using one of the available APIs from where it is delivered as control plane information to the device. Note, the data will pass through the SMF and AMF before being passed to the gNB for scheduling and onward transmission across the radio / 5G Uu interface. This methodology is considered to be more efficient when transferring very small amounts of data to potentially millions of devices as in the case of MMTC (Massive Machine Type Communications) deployments.

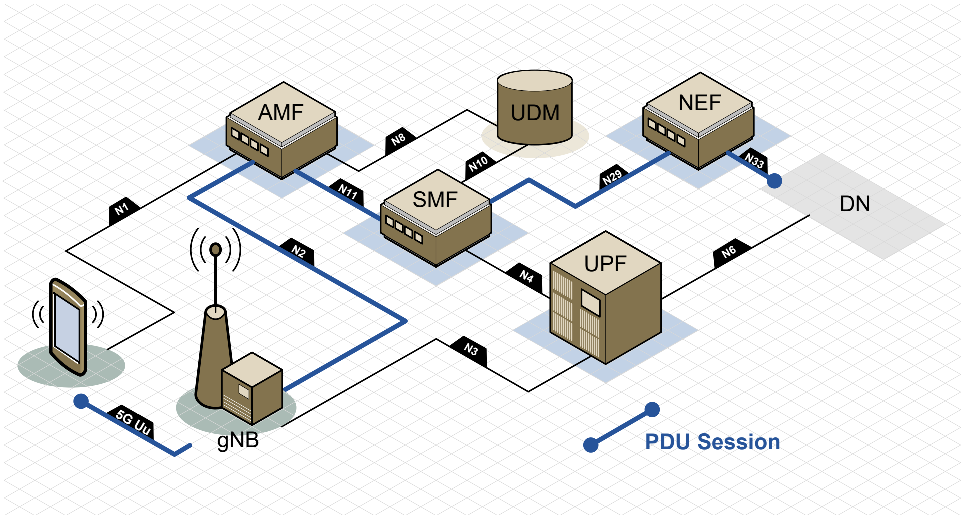


Figure 2-21 Transferring Data via a NEF

The NEF is not only restricted to NIDD but in fact can also support network monitoring, provisioning and policy / charging configuration.

2.3 Network Operation

2.3.1 Finding the Network

Prior to attaching to the 5G network, the device must first of all find its home network and then find a suitable cell on which to "camp". The high level process that supports this is outlined in Figure 2-22 and is much the same as earlier 2G, 3G and 4G technologies.

Figure 2-22 Finding the Network

- PLMN (Public Land Mobile Network) Selection – before choosing a suitable cell, the device must first find a suitable 5G network. This may involve a frequency scan to find available gNBs, after which point broadcast system information can be read by the device to facilitate network selection.
- Cell Selection – once a suitable network has been found, the device will choose a suitable cell within that network (taking into account cells which have access limitations). This will include taking measurements for all neighbouring cells and ultimately choosing the cell that will provide the best coverage.
- Radio Connectivity – once a cell is chosen, the device will initiate a radio connection, which grants signalling resources to the device to allow it and the gNB to exchange signalling messages related to the 5G RAN. Crucially, the connection also allows the device and the AMF to communicate with one another via the gNB.

Following these initial procedures, the device can conduct the Registration procedure.

2.3.2 Network Registration

The Registration procedure is conducted by the device in order to “Attach” to the 5GC. During this process, the device will be allocated an AMF and the network will undertake a series of security procedures which are designed to authenticate the device and establish secure communication.

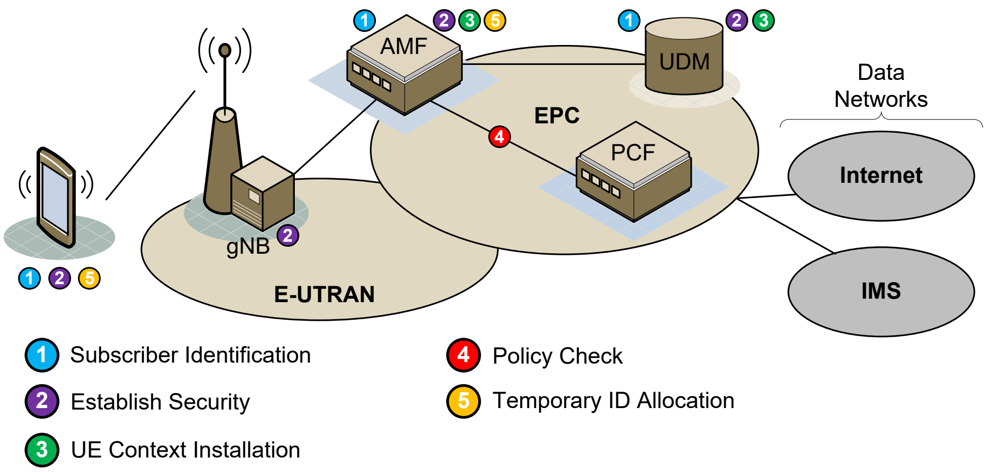


Figure 2-23 Key Activity within the Network Attach Procedure

2.3.2.1 Subscriber Identification

The device will typically provide to the network its last known temporary ID, assuming it has one. As such, in order to support security and subscriber profile acquisition, the network will need to ascertain the subscriber’s real identity. If the AMF cannot resolve the temporary ID to the permanent ID of the subscriber, the AMF will query the subscriber device directly.

Note that in 5G, a “Concealed Identity” can be used, which is designed to protect the subscriber’s permanent identity (typically their IMSI (International Mobile Subscriber Identity)) from being snooped. Essentially, any time the device wishes to send its permanent ID to the network, it will first encrypt it using the public key of the mobile service provider.

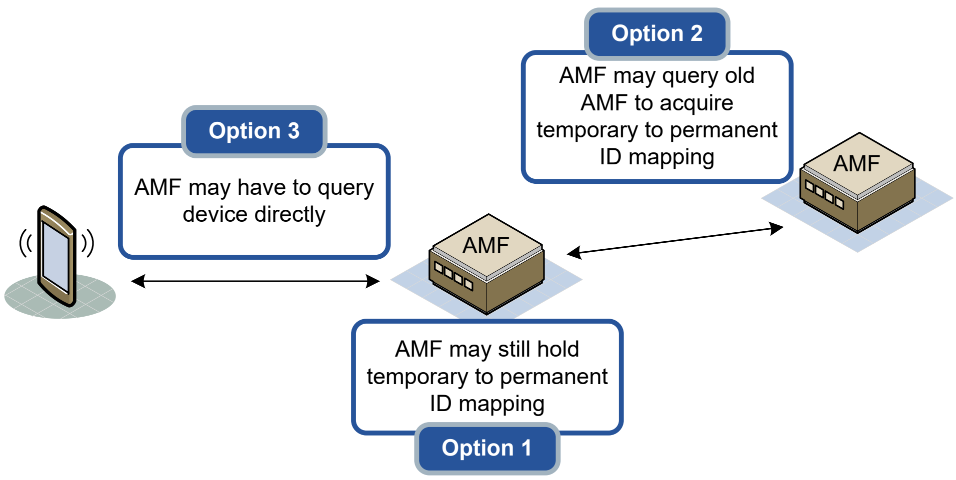


Figure 2-24 Subscriber Identification

2.3.2.2 Establish Security

Several layers of security are used in 5G to protect the user’s data and signalling in transit. This security is established as part of the overall Registration procedure and is based on a process termed 5G AKA

(Authentication and Key Agreement). In essence, the technique relies on the notion of a shared secret key; the SIM (Subscriber Identity Module) stores one copy and the UDM stores the other. During the Registration procedure, the secret key is used as the basis for mutual authentication (both the device and the network prove to one another that they are legitimate) and also encryption. All signalling traffic between the device and the network can be encrypted to prevent snooping, in addition to all user traffic.

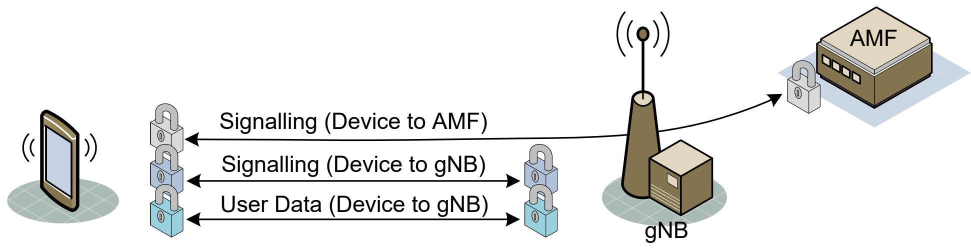


Figure 2-25 Establish Security

2.3.2.3 UE Context Installation

Every valid subscriber has a subscriber profile stored in the UDM. This profile will define the DNs that a subscriber is permitted to connect to, including the QoS profile they are allowed to have for that network. Additional examples of information in the subscriber profile includes an overall bandwidth cap, roaming permissions and billing information.

During Registration, the AMF must acquire the subscriber profile from the UDM. The AMF then uses this information to check if barring or roaming restrictions are in place for the subscriber. The information provided by the UDM is stored in the AMF as a UE Context.

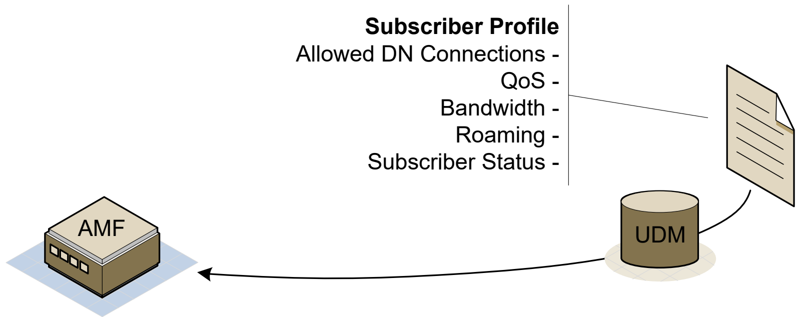


Figure 2-26 UE Context Installation

2.3.2.4 Policy Check

During the Registration procedure, the AMF may typically conduct a policy check with the PCF. Policy control is often influenced by conditions, such as the time of day, location of the user, network congestion, handset type etc. Therefore, before the AMF allows the subscriber to register with the network, a check can be made with the PCF, essentially requesting that based on the current network conditions, is the subscriber permitted access to the network?

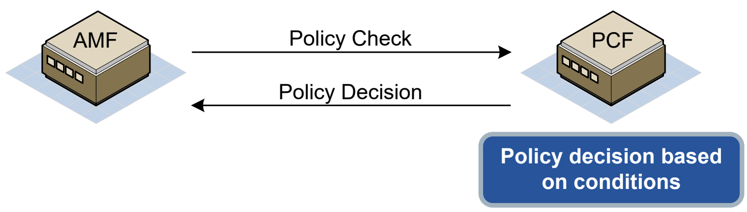


Figure 2-27 Policy Check

2.3.2.5 Temporary ID Allocation

The final part of the Registration procedure is the allocation of a temporary ID to the subscriber. This temporary ID is created by the AMF and is used for the remainder of the registration period or until the device is allocated a new temporary ID, potentially due to an AMF change.

2.3.3 Establishing a PDU Session

When the device is required to exchange user plane data with the network, a PDU Session will need to be established. The main elements of the PDU Session Establishment procedure are outlined in Figure 2-28.

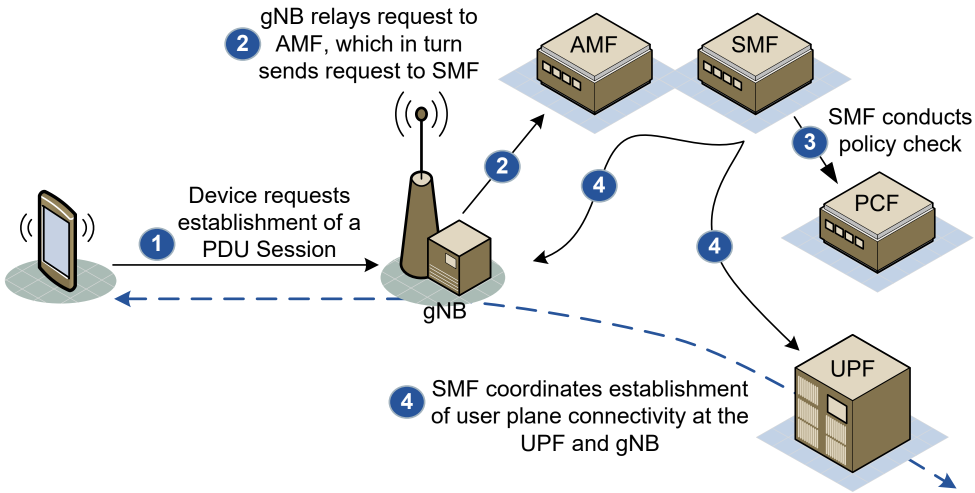


Figure 2-28 Establishing a PDU Session

The data connections established include a radio bearer between the device and the gNB and a connection between the gNB and the UPF. These connections are logical in nature, with the network supporting thousands of logically separate PDU Sessions.

A subscriber's PDU Session is unique to them; throughout its existence, a PDU Session will only carry user data specific to one subscriber.

If a subscriber has multiple PDU Sessions, each PDU Session needs its own set of connections across the network. For example, if a subscriber has three PDU Sessions in place, this means there will be three radio bearers between the device and the gNB, and three connections between the gNB and the UPF.

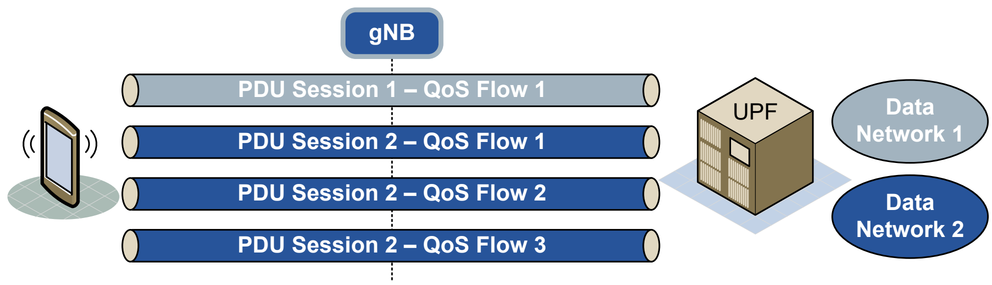


Figure 2-29 User Data Connections

Note that within a PDU Session, several separate QoS Flows may be in operation. Each individual QoS Flow will have a QoS profile different to any other QoS Flows within the same PDU Session. To allocate a particular piece of user data to a specific QoS Flow, the device or UPF must add an additional header to the data packet which contains a QoS Flow ID.

2.3.4 Sending and Receiving Data over 5G

2.3.4.1 Idle and Connected Mode

Once registered with the network, the device will constantly pass between being “Idle” or being “Connected”. The key differences between the two modes are shown in Figure 2-30, but the high-level distinction is that Idle mode is used to save battery life; if the device does not have any data to exchange with the network, it is a waste of battery power to keep the device’s 5G radio 100% active.

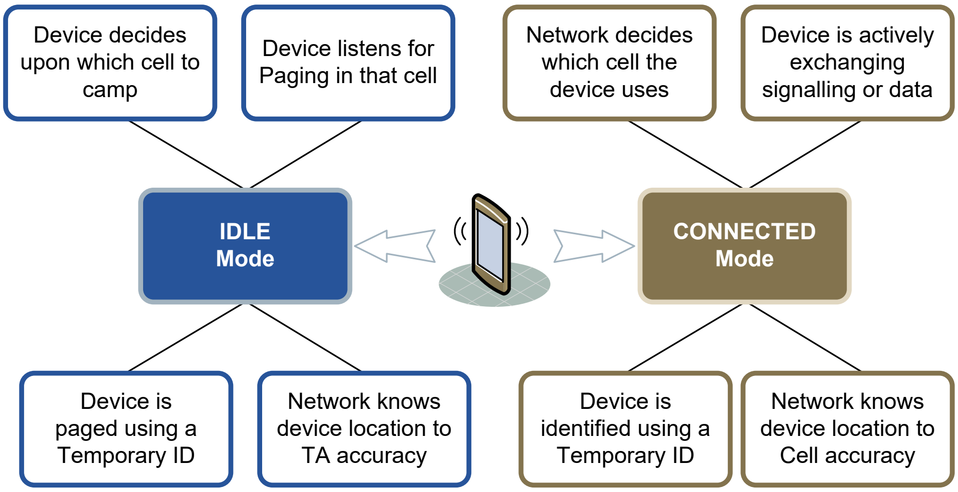


Figure 2-30 Idle and Connected Mode

The transition between Idle and Connected modes is a constant occurrence, which is largely attributed to the way in which apps on a typical smartphone are “chatty” with the network (assuming a smartphone is the device in question). Even if the subscriber is not actively checking Facebook, sending messages, checking emails etc, the apps on the device may still be actively exchanging data with the network. The smartphone can potentially drop to an Idle mode and literally within milliseconds be transitioning back into a Connected mode. Moments later, the phone may transition down to Idle again, with the process repeating as and when data needs to be sent.

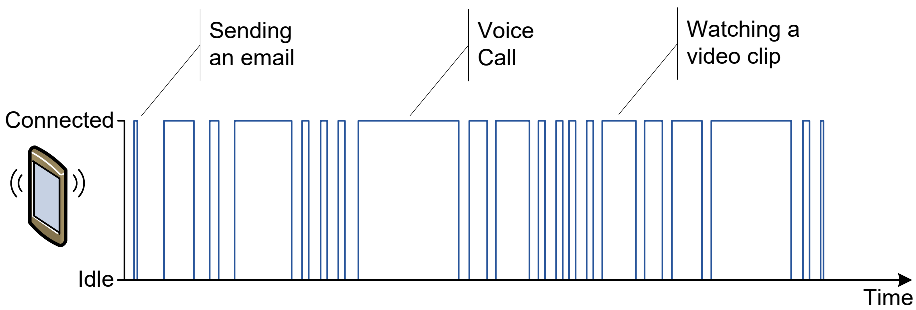


Figure 2-31 Transitioning between Idle Mode and Connected Mode

Note that some IoT devices will behave differently in the network, since they may have much stricter requirements on optimizing the life of the battery. As such, devices may access the network on a much more infrequent basis, and may even operate in MICO (Mobile Initiated Connection Only) mode, which prevents the network from paging the device.

2.3.4.2 Moving to Idle Mode

At some point, a device in a Connected mode will not have any data to send or receive in the network. It is the gNB which monitors this situation, essentially looking out for a fixed period of user inactivity. If the gNB deems that the device has been inactive for, say 15 seconds, it will send a request to the AMF to be permitted to send the device to Idle. At this point, connectivity resources will be torn down in the 5G RAN but the PDU Session itself will remain.

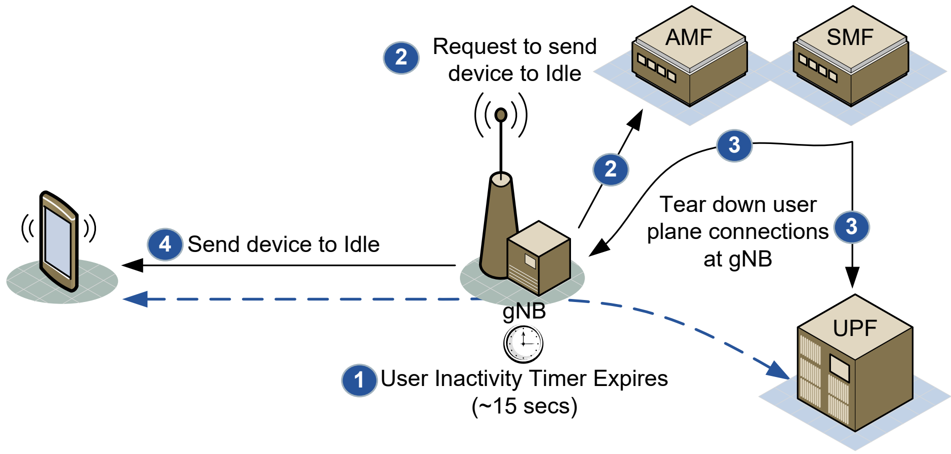


Figure 2-32 Moving to Idle Mode

When in Idle mode, the resources of the device's PDU Session(s) which span the RAN are not in place. This is due to the fact that the device, when Idle, will be determining the best cell to use and will switch to a new cell without informing the network (as long as the new cell is not in a new Tracking Area). Therefore, there is no point in maintaining data connectivity at a particular gNB if the device has already potentially decided to camp on a different cell.

2.3.4.3 Getting Connected – Service Request

When a device in an Idle state has user data to send, it will use a Service Request to move back into a Connected State. Bearing in mind that until this point the device was simply listening for paging, the first task it must complete is to establish a radio connection to the gNB looking after the cell in which the device is camping. Once this is achieved, a Service Request is sent from the device to the AMF via the gNB.

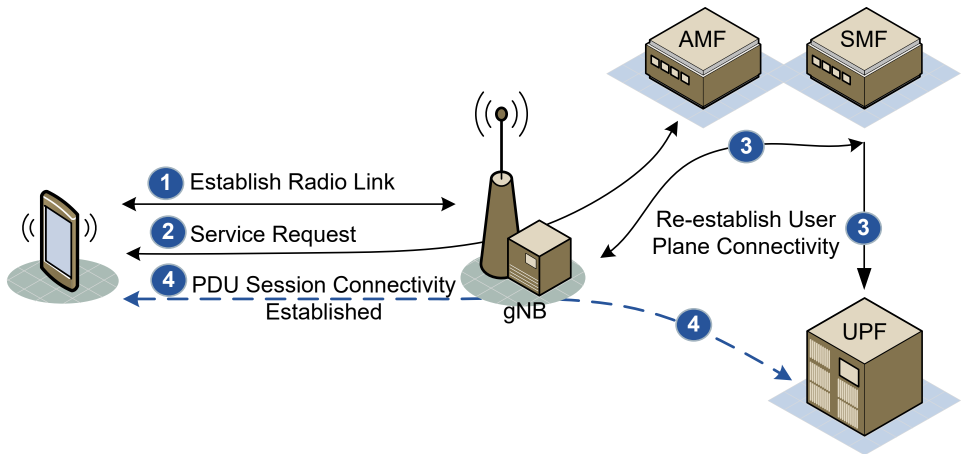


Figure 2-33 Getting Connected – Service Request

The aim of the Service Request is to establish NG-RAN connectivity for each of the PDU Sessions that the device currently has active (see Figure 2-33 – the connectivity was torn down when the device transitioned to Idle). Note that in 5G, the device can be selective as to exactly which PDU Sessions have their data plane connectivity re-established.

2.3.4.4 Getting Connected – Paging

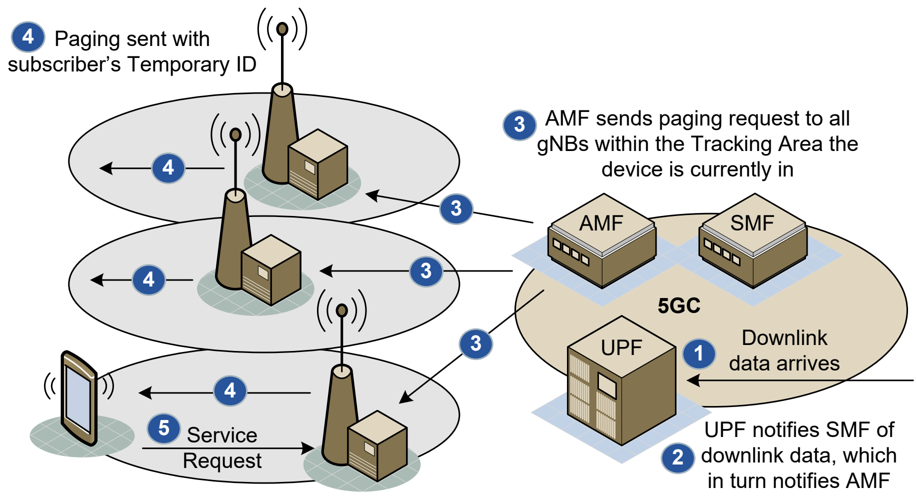


Figure 2-34 Getting Connected - Paging

When the device is in an Idle state, it is possible that downlink data traffic arrives at the UPF destined for the user. For example, the subscriber may have been sent a WhatsApp message or perhaps a call establishment request. The problem is, when the device is Idle, the network only knows the location of the user to the accuracy of a Tracking Area. As such, when downlink data arrives for the device, the AMF will conduct the Paging process in which a Paging message is sent to every gNB within the subscriber's known Tracking Area. In turn, each gNB will broadcast the Paging message within every cell that the gNB is responsible for (note that contrary to this, the mobile service provider will undoubtedly use an optimized Paging strategy.

From the device's perspective, in an Idle state it has camped on a particular cell (typically the cell that provides the best signal) and is periodically switching on the 5G radio and listening out for a Paging message. The Paging message carries the temporary identity of the subscriber and as soon as the device recognizes this, the phone is triggered into conducting a Service Request procedure (outlined in Figure 2-33).

2.4 Mobility and Interworking

2.4.1 Tracking Area Updates

Once a device is attached to the network, it is imperative to the success of the Paging procedure that the network is kept informed as to the current Tracking Area in which the device is residing. This is only relative to when the device is in the Idle mode; as soon as the device is Connected, location information is much more accurate (down to the granularity of an individual cell rather than a Tracking Area).

The Tracking Area Update process allows the device to keep the network informed of the current Tracking Area in which the device can be found. As an example, if a device moves from one cell to another and finds out that the new cell belongs to a different Tracking Area, this will trigger the device into conducting the Tracking Area Update procedure. This is just one example; there are several other reasons to conduct a Tracking Area Update.

The procedure itself is very straightforward, with a simple message exchange between the device and the AMF (this occurs after the device has established a radio link with the gNB of the current cell).

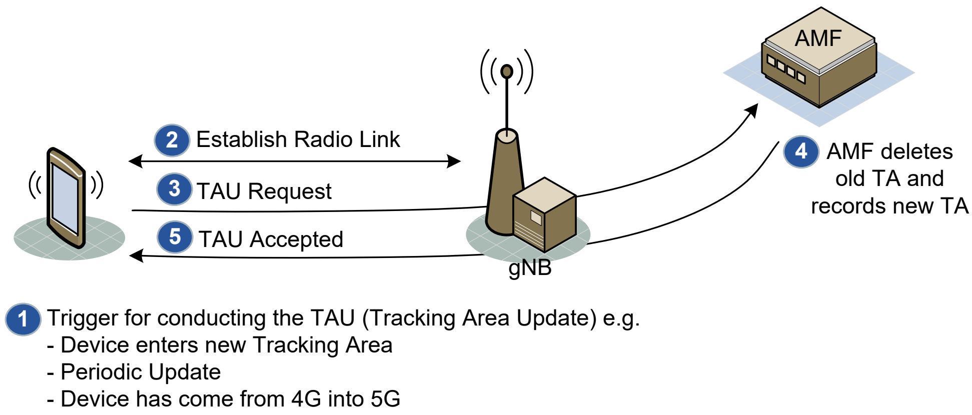


Figure 2-35 Tracking Area Update

2.4.2 Handovers

Handovers take place when the device is in the Connected mode, under the control of the network. That is, the network decides when the device should be handed over from one cell to the next.

Two mechanisms for conducting handovers are available, termed the Xn Handover and N2 Handover.

2.4.2.1 Xn Handover

Xn based handovers are coordinated largely by the two gNBs involved in the handover process – the Source gNB and Target gNB. Using the Xn interface that links them, the Source gNB will supply the Target gNB with all the information it needs to receive the subscriber. Such information includes security information, as well as the key characteristics of the PDU Session(s) that the device is currently utilizing.

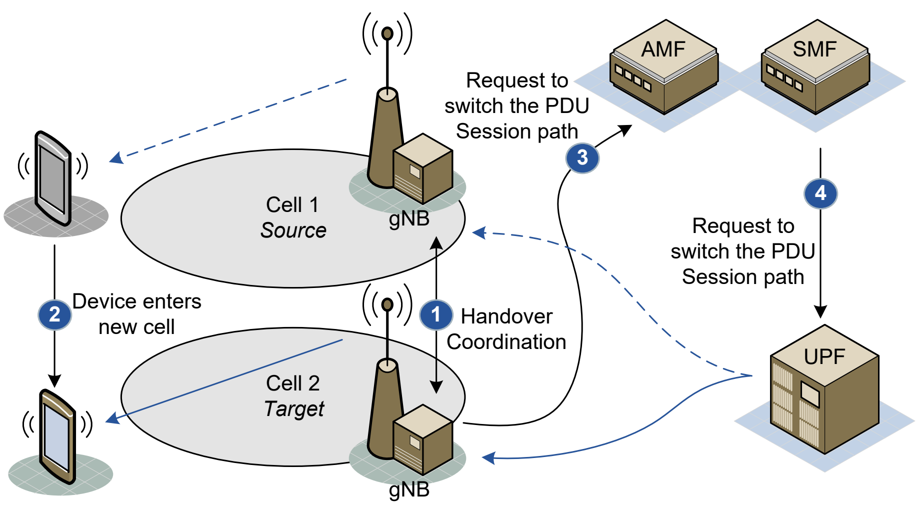


Figure 2-36 Xn Handover

Ultimately, the Target gNB will determine as to whether or not it can accommodate the handover. Assuming the answer is yes, all that remains is to redirect the device to the new cell and redirect the PDU Sessions(s) to the new gNB. Due to the fact that there is little involvement by the core network in the handover process, the Xn Handover is a very popular technique for facilitating handovers.

2.4.2.2 N2 Handover

N2 Handovers typically occur much less frequently in the network, largely because Xn Handovers are more prevalent. Typically, N2 Handovers are used because an Xn Handover is not possible (possibly due to network architecture or lack of provisioning).

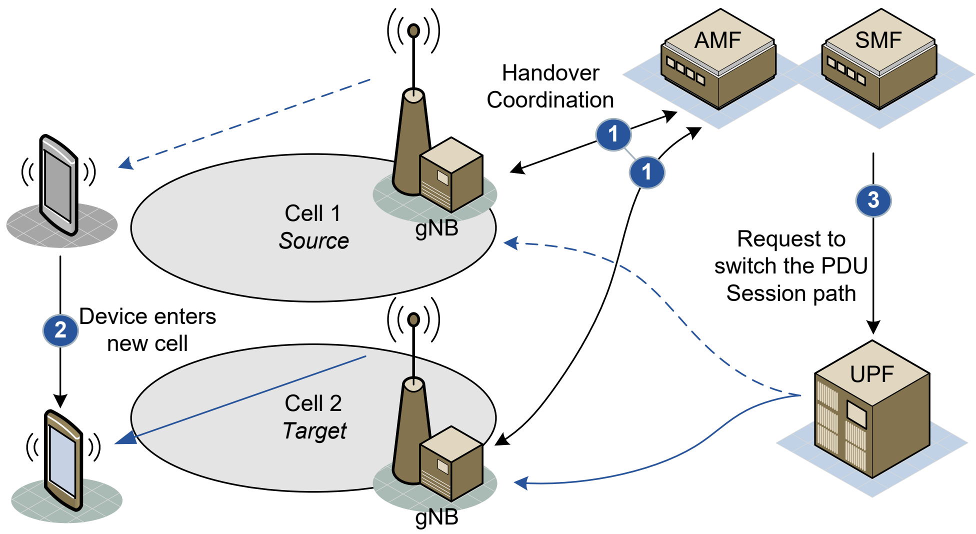


Figure 2-37 N2 Handover

Both handover techniques achieve the same goal; moving the device from one cell to another. However, with an N2 Handover the core network coordinates the overall process instead of the gNBs. In essence, the AMF acts as a go between for the Source gNB and Target gNB, relaying all the required messages. Fundamentally, the Target gNB still needs to know all the necessary information relative to the device, and it still needs to make a decision as to whether or not the device can be accepted. If the handover takes place, the device will be directed to the new cell and the AMF / SMF will ensure that the PDU Session(s) are redirected to the new gNB.

2.4.3 Interworking with 4G

Clearly nobody is going to buy a 5G device if it is only capable of working on a 5G network so as such, the ability to interwork with legacy networks is going to be critical from the outset.

A device registered on the 5G network and effectively under the control of the AMF and SMF will need to be transferred to 4G and the control of the MME. Clearly this will be initiated by the gNB determining that the 5G connection can no longer be maintained and that a handover to 4G is required. As such, signalling throughout both the 5GC and EPC (Evolved Packet Core) will be necessitated in order to transfer the PDU Session spanning the 5G network to a PDN Connection spanning 4G. This is illustrated in Figure 2-38.

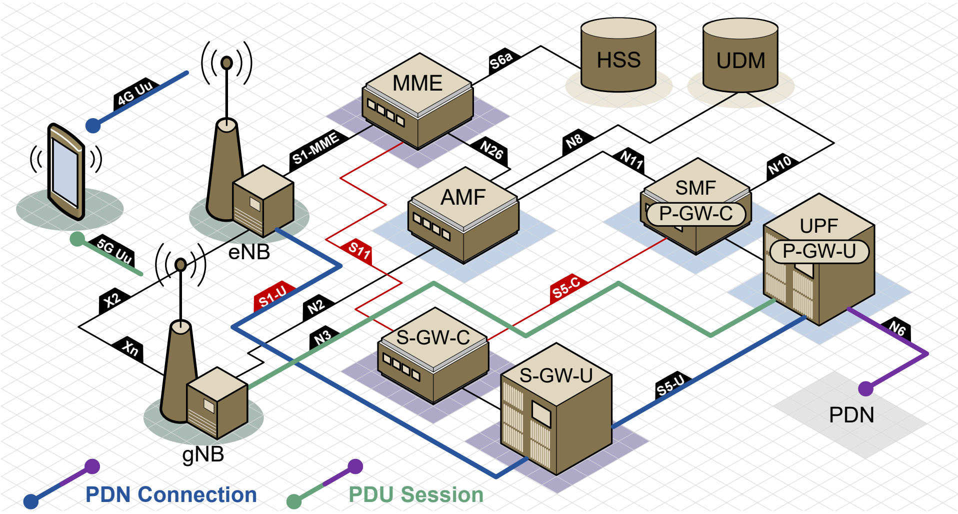


Figure 2-38 Interworking with 4G

As the UPF and the P-GW-U are one in the same, the data anchor can be maintained with regards to the PDU Session / PDN Connection and thus there will generally be no requirement to change the device's allocated IP address. However, there may need to be a change in QoS as in this example, the 4G network may not be able to offer the same level of service.

2.4.4 Interworking with Wi-Fi

Cellular networks interworking with Wi-Fi have been commonplace for some time, especially with the advent of VoLTE (Voice over LTE) and Wi-Fi Calling. As such, similar collaboration is taking place between 5G and Wi-Fi as illustrated in Figure 2-39 with the inclusion of the N3IWF (Non 3GPP Interworking Function). This terminates the IPSec (IP Security) tunnel spanning the connection to the device (Y1 and Y2) and as such, includes the insecure Wi-Fi element. From the N3IWF, the data will pass to the UPF via N3 which is consistent with normal 5G operation, whereas signalling between the device and the 5GC will be passed to the AMF across the N2 reference point.

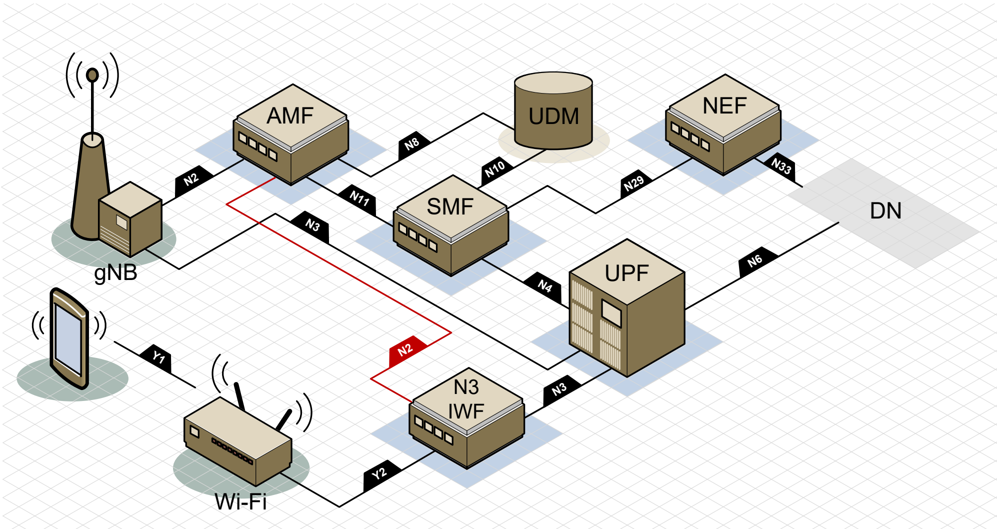


Figure 2-39 Interworking with Wi-Fi

2.4.5 Roaming

The 5G System supports two different roaming scenarios, namely Home Routed or Local Breakout.

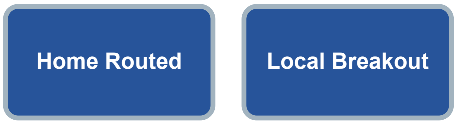


Figure 2-40 Roaming in 5G

Figure 2-41 depicts the architecture for the Home Routed scenario, whereby all User Plane traffic will be sent from the Visited Network to the DN (Data Network) via the Home Network (utilizing the N9 reference point between the Visited and Home Network). Control Plane traffic is also sent back to the Home Network where necessary, utilizing the N8, N12 and N16 reference points. For example, during registration, the Visited AMF would acquire subscriber information from the Home UDM via N8. Moreover, the V-SMF would liaise with the H-SMF via N16 in order to establish the PDU Session.

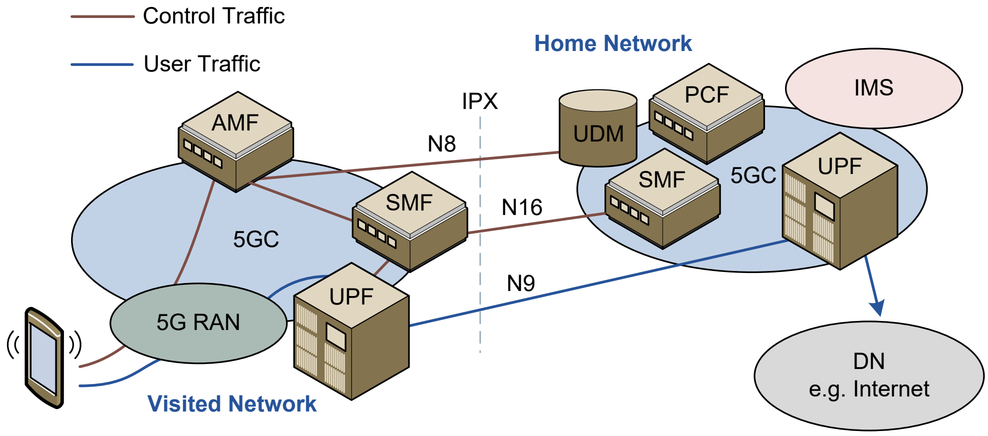


Figure 2-41 5G Roaming – Home Routed

Figure 2-42 shows the architecture associated with Local Breakout, in which the User Plane traffic is sent to the DN directly from the Visited Network. The Home Network's involvement is only required for Control Plane purposes.

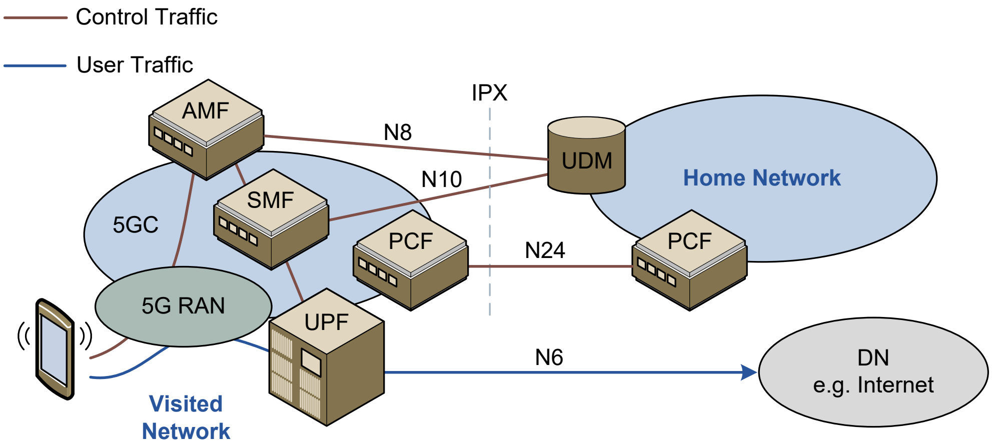


Figure 2-42 5G Roaming – Local Breakout

3 5G Services and Technology Enablers

3.1 Understanding Network Slicing

Over the past three decades, communication technology has played a significant role in the digitization of our society and with the advent of 5G, there is every reason to think this will continue through the decade ahead. Many organizations spanning numerous verticals have already embraced mobile communications to drive both effectiveness and efficiency however there are many others such as automotive, manufacturing, agriculture healthcare and energy which are only just beginning to embrace the opportunities. This sub-optimal use is probably down to the fact that different industries require different attributes from their communication networks; ultra reliability, ultra high bandwidth, low latency etc. In essence, they all need a different network.

3.1.1 What is Network Slicing?

From a functional standpoint, building dedicated networks to cater for the specific requirements for each customer would appear the only way forward, rather than forcing a "one size fits all" approach which to some extent has been the strategy adopted by previous technologies. However, a much more efficient approach is to operate each dedicated networks on a common platform or in other words, to create a "slice" for each customer or industry.

> ***Network Slicing is the embodiment of the concept of running multiple logical networks as virtually independent business operations on a common physical infrastructure in an efficient and economical way***
>
> *GSM Association*

Figure 3-1 What is Network Slicing?

Therefore, from the mobile service provider's perspective, a network slice is an independent end-to-end logical network that runs on a shared physical infrastructure which is capable of providing a negotiated quality of service. From the customer's perspective however, the network slice supporting this negotiated quality of service is completely transparent.

Figure 3-2 illustrates the network slicing concept in 5G and as such, it can be seen that the slice can span across multiple parts of the network; RAN (Radio Access Network) and 5GC (5G Core) network, not forgetting the underlying transport networks.

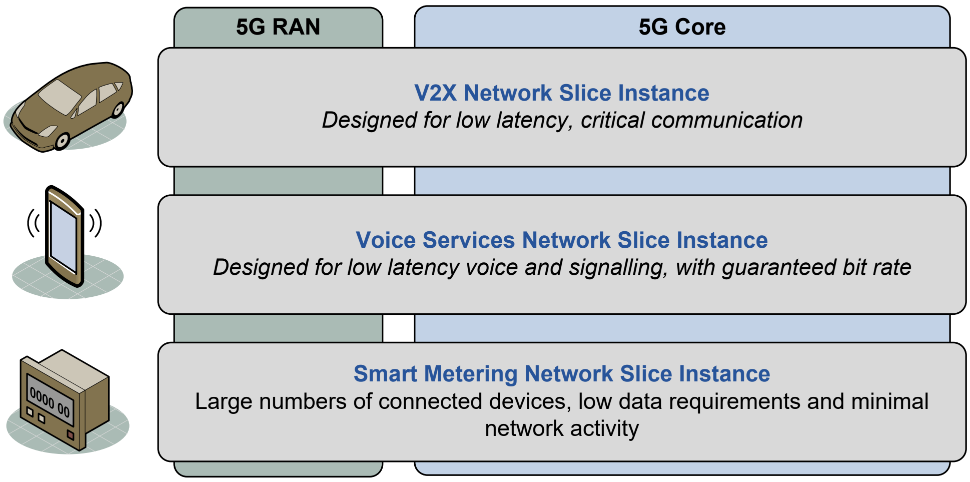


Figure 3-2 Network Slicing In 5G

Slice types can be defined from either a functional or behavioural perspective. Therefore, mobile service providers can deploy a single slice type that satisfies the needs of multiple verticals, through to dedicated slices that are packaged as a single product and targeted towards specific business functions which may have very diverse requirements e.g. V2X (Vehicle to Everything) and autonomous vehicles.

> ***5G networks, in combination with network slicing, permit business customers to enjoy connectivity and data processing tailored to the specific business requirements that adhere to a SLA (Service Level Agreement) agreed with the mobile service provider. The customisable network capabilities include data speed, quality, latency, reliability, security and services***
>
> GSM Association

Figure 3-3 Defining a Network Slice

Therefore, in order to accommodate the specific features required by the customer, the mobile service provider is able to provision different characteristics for each network slice in terms of Functionality, Performance and Users which in turn map down to the technical characteristic dealing with throughput, latency, reliability, security, availability, mobility etc.

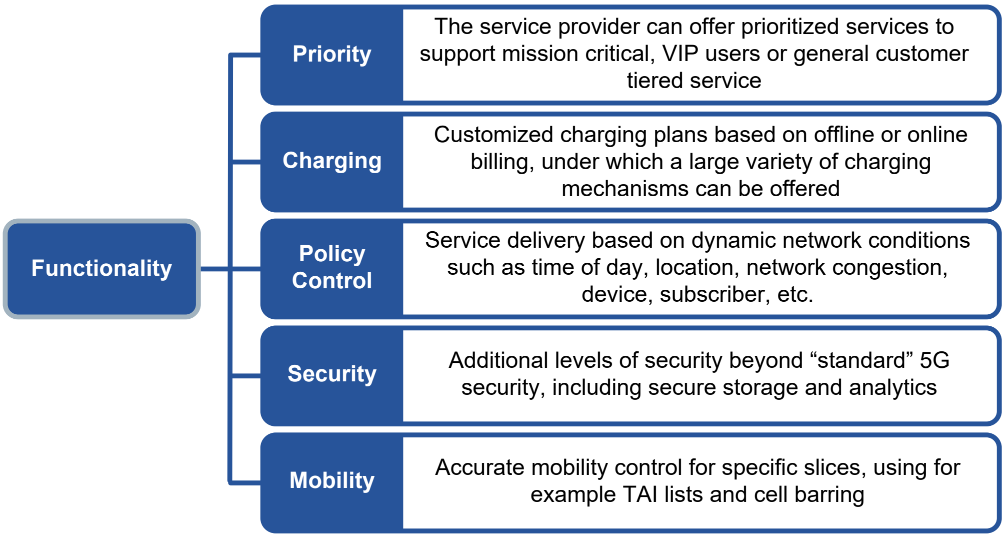


Figure 3-4 Network Slice Characteristics – Functionality

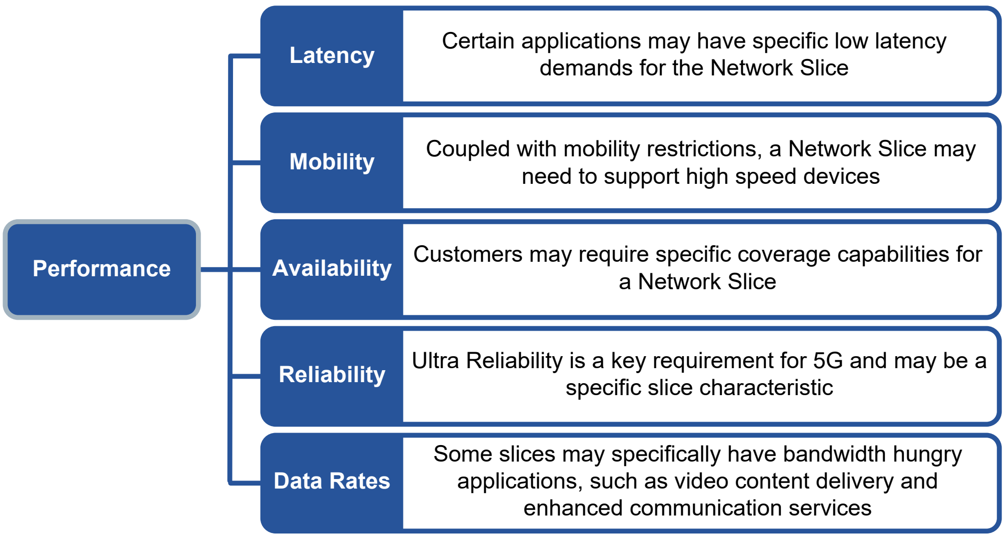


Figure 3-5 Network Slice Characteristics – Performance

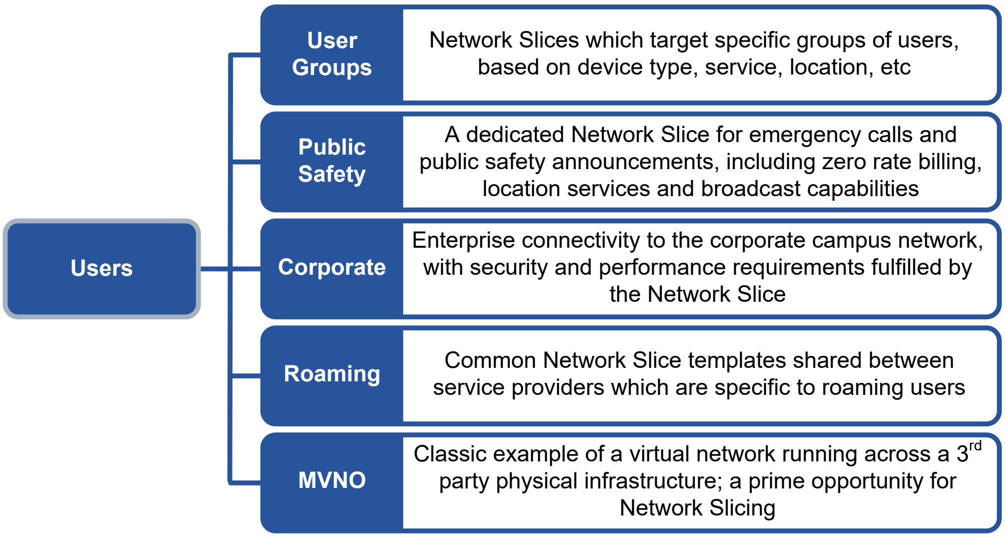


Figure 3-6 Network Slice Characteristics - Users

3.1.2 Deploying a Network Slice

As mobile service providers endeavour to transition away from simple “bit pipe” providers, network slicing enables them to provide a more integrated solution to drive the business proposition. This can be achieved in different ways depending upon the customers’ requirements; Hosting, Integration and Capability Exposure.

Figure 3-7 Deploying a Network Slice

Hosting

Mobile service providers now have the ability to host applications on behalf of their customers and collect the relevant data within the network slice. This information, along with data derived from other sources can be processed and fed into AI (Artificial Intelligence) algorithms. In other words, the mobile service provider will be able to expand their service offering and provide an end-to-end solution. This is particularly relevant for the MMTC (Massive Machine Type Communications) use case where many businesses have the requirement but not necessarily the technical expertise.

Integration

Some business customers may already have a communication or suitable network infrastructure in operation e.g. within an industrial setting. Therefore, this infrastructure can be integrated into a network slice. For example, an organization may implement their own 5G core network as part of their overall IT infrastructure but may integrate with the RAN from several mobile service providers, potentially on a global basis.

Capability Exposure

Alternatively, mobile service providers could offer their customer the capability to manage their own services across their slice through a series of API (Application Programming Interface) within the pre-defined SLA. As such, network specific information will be accessible to the customer including service quality, network conditions, device status and mobility to name but a few.

3.1.3 Network Slicing Use Cases

Clearly there are numerous verticals that could benefit from network slicing and the closer integration with the mobile service provider's network that would bring. As such, and to further aid in the explanation of the commercial implications of network slicing, we have identified seven here.

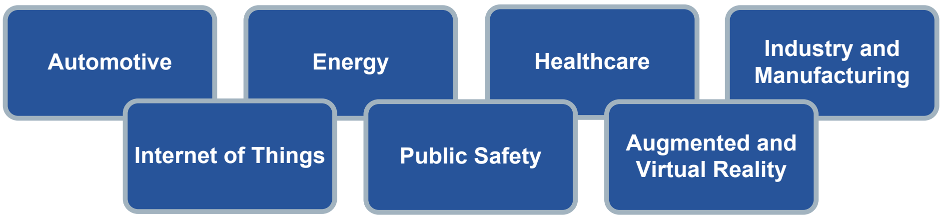


Figure 3-8 Network Slicing Use Cases

Automotive

Several use cases for network slicing exist for the automotive industry, as outlined in Figure 3-9.

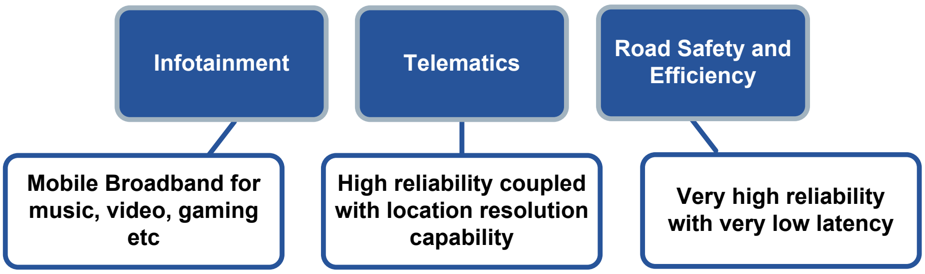


Figure 3-9 Automotive Use Cases and Requirements

- Infotainment – although not safety critical, infotainment services can be high on the agenda. Services such as streaming music and video, along with gaming, will typically require mobile broadband characteristics from the network slice.
- Telematics – typically designed to assist the driver by using a broadband connection to an application server(s), in support of services such as navigation, vehicle health and parking availability. This will require high reliability as well as location capability from the network slice.
- Road Safety and Efficiency – collision warnings, platooning control, autonomous vehicles and cooperative driving are all potential services facilitated through the network slice. Such services will all require very high reliability and very low latency.

Energy

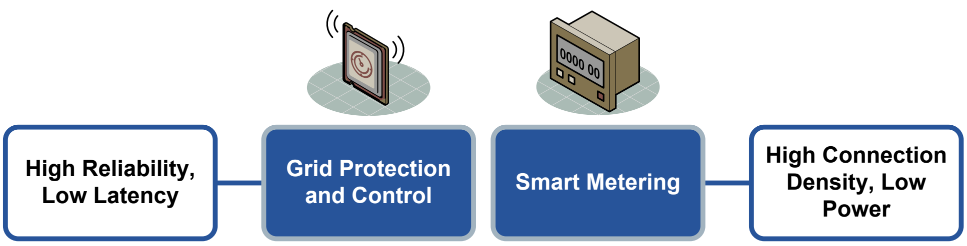


Figure 3-10 Energy Use Cases and Requirements

Both energy grid protection and control, along with control of smart meter networks are two key areas of network slice utilization for the energy sector. Note that the two areas are very different in their demands placed on the network; the former requires high reliability and very low latency, whereas the latter requires high connection density, coupled with low power requirements.

Healthcare

Although remote surgery seems to be a widely cited potential capability for 5G, the reality is that it will take a long time before 5G networks are considered reliable enough for surgical operations to become routine traffic on a 5G network slice.

For healthcare, the more mainstream use cases are centred around patient monitoring and tracking, including location tracking in some cases. In terms of network slice requirements, availability and reliability are clearly crucial, followed by latency and throughput.

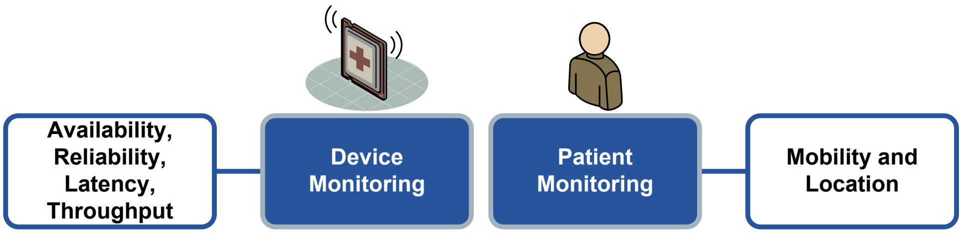


Figure 3-11 Healthcare Use Cases and Requirements

Industry and Manufacturing

In relation to the industry and manufacturing vertical, five main use cases are generally considered, as outlined in Figure 3-12. Due to the fact that this vertical is so diverse, the associated requirements for the vertical are also diverse.

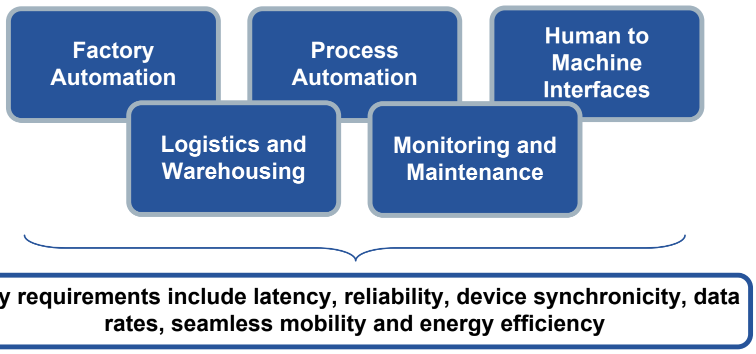


Figure 3-12 Industry Use Cases and Requirements

Internet of Things

Although the Internet of Things can apply to lots of potential use cases, the core focus with respect to Network Slicing is the support for low power devices, distributed across a wide area. A large number of areas can exploit IoT connectivity, including agriculture, environment monitoring, consumer, industrial, logistics, smart buildings, smart cities and utilities.

Figure 3-13 outlines a selection of common use cases for IoT applications.

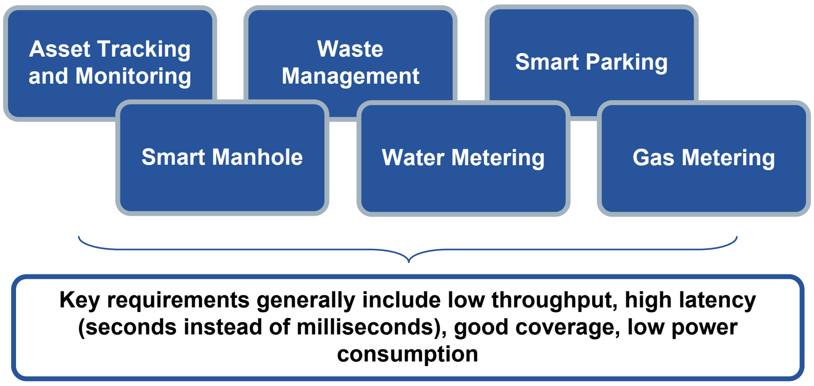


Figure 3-13 LPWA Use Cases and Requirements

Public Safety

Public Safety networks have long been a function of the cellular industry and although the trend is currently towards LTE based deployment, it is anticipated that network slices within 5G could well support Public Safety services in the future.

As can be seen in Figure 3-14, Public Safety encompasses a number of use cases, all of which have quite varied requirements for the network. As such, different network slices would be used for each use case. In terms of commonality between the use cases, security, priority and pre-emption are clearly high on the agenda.

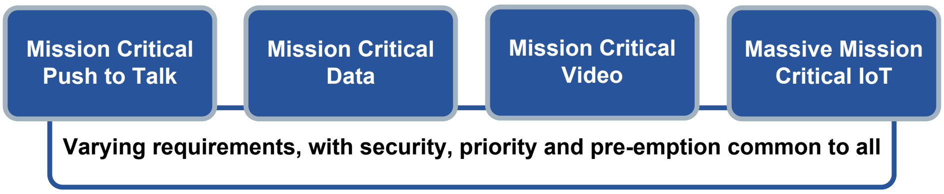


Figure 3-14 Public Safety Use Cases and Requirements

Augmented and Virtual Reality

AR (Augmented Reality) applications are designed to complement a subscriber's experience of the real world, typically via auditory, visual or haptic enhancements. Conversely, VR (Virtual Reality) provides a more immersive experience for the user, which can be based on a real environment with real time interaction between other VR users. In either case, potential applications for AR and VR impact many other market verticals, including entertainment, gaming, education, healthcare and training.

	Entry Level VR 8K Definition 2D/3D	Advanced VR 12K Definition 3D	Ultimate VR 24K Definition 3D
Data Rate	120Mbps (2D) 200Mbps (3D)	1.4Gbps	3.36Gbps
Round Trip Time	10ms	5ms	5ms
Packet Loss	10^{-6}	10^{-6}	10^{-6}

GSM Association

Figure 3-15 Augmented and Virtual Reality Use Cases

From the perspective of network slicing, AR and VR have the potential to place heavy demands on the mobile service provider's network. Figure 3-15 illustrates this fact by specifying the network requirements for a range of VR scenarios in terms of data rate, round trip delay (latency) and packet loss (reliability).

3.1.4 Network Slicing and Roaming

In addition to being able to operate within a network, devices utilizing a specific network slice also need to be able to roam and be afforded the same service. Clearly for some verticals, this is critical and as such, has significant commercial implications.

To aid this, the GSM Association have identified three examples of how network slicing can be implemented through the instantiation of a network slice as a device roams on a visited network.

Standardized Slice Type

In this example, a globally agreed slice type allocated to the roaming device is instantiated in the visited network.

1. The device roams onto a visited network.
2. The visited network instantiates the same standardized slice type.

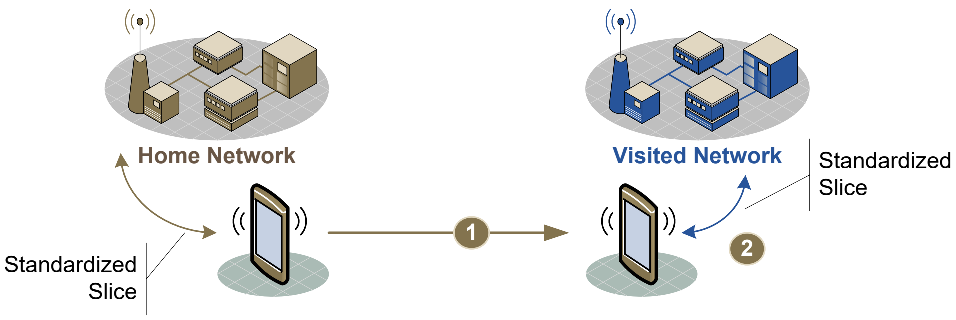


Figure 3-16 Roaming – Standardized Slice Type

This type of approach would work for universal slice types such as mobile broadband or standard voice services.

Export Slice Blueprint

For this example, the home network provides the slice blueprint to the visited network which in turn instantiates if for the roaming device.

1. The home network exports the custom slice template.
2. The device roams onto the visited network.
3. The visited network instantiates a slice identical to the one used on the home network.

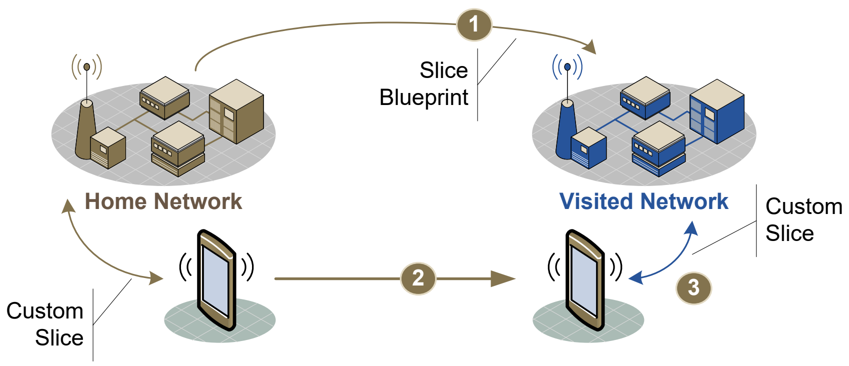


Figure 3-17 Roaming – Export Slice Blueprint

Virtual Home Slice

When granted by the visited network, the home network takes control of the network resources in the visited network. This may be tricky to implement due to matters relating to trust and security.

4. The device roams onto the visited network.
5. The home network requests permission to control the visited network.
6. The device uses the same custom slice which extends from the home network.

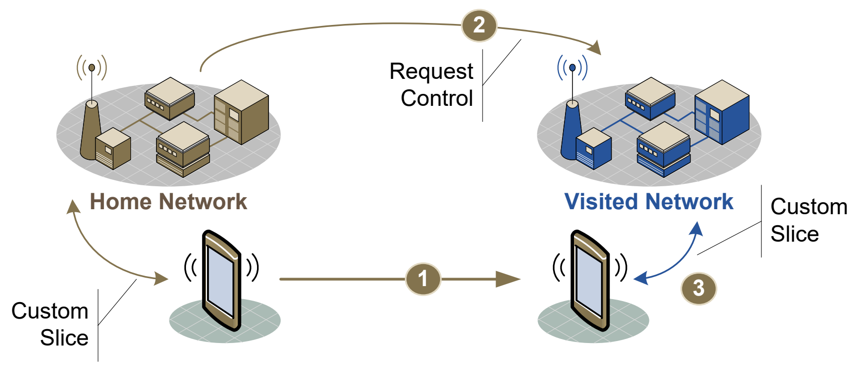


Figure 3-18 Roaming – Virtual Home Slice

3.2 The Role of Multi access Edge Computing

Following on from our review of network slicing, we shall now turn our attention to MEC (Multi access Edge Computing) and its ability to enable a range of new exciting applications to become mainstream in the years ahead.

3.2.1 What is MEC?

In essence, MEC is cloud computing for both consumer and enterprise type applications within both the fixed and mobile service providers' network although in this section, we will focus specifically on its implementation within the mobile environment. Furthermore, the advantages brought about by the introduction of MEC can equally apply to private 5G networks and as such, drive new services and applications within a range of different verticals.

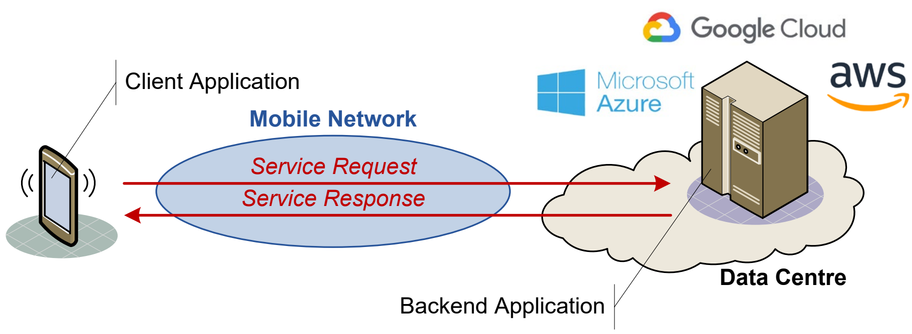


Figure 3-19 Legacy Application Transactions

Today, many of the applications that run on smartphones are comprised of two fundamental elements; the client application on the device itself e.g. news, sport, social media etc. along with a backend IT application typically hosted in a private Data Centre or public cloud such as AWS (Amazon Web Services), Microsoft Azure or Google Cloud. The client device is relatively "thin" and requests service form the more powerful data centres which carry out the processing and return the information back down to the client. As such, the mobile network simply acts as a "bit pipe" carrying information between the two endpoints.

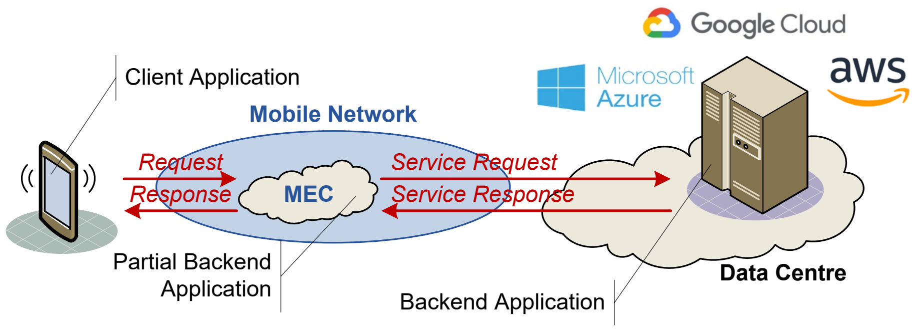


Figure 3-20 MEC Based Application Transactions

As we transition towards the next generation of rich applications such as VR and AR, the traditional approach is unlikely to be able to support this mode of operation. As such, the cloud needs to come to the devices with the creation of a mini cloud, cloudlet or fog within the mobile service providers network. This mini cloud is effectively MEC and is illustrated in Figure 3-20.

3.2.2 Benefits of Introducing MEC

This approach brings two distinct advantages. First of all, instead of running the whole backend application in the cloud, some of it can now be run on the cloudlets within the mobile service providers network, thereby improving the application response time or drastically reducing latency (critical for applications such as VR and AR).

Secondly, by adopting a MEC based approach, the full client software no longer needs to run on the client itself which therefore enables some of the computation to be offloaded to the network edge. This improves battery life, increases processing capabilities and reduces issues associated with heat dissipation, not to mention the form factor of the device; all important aspects for the growing IoT space, including new wearable devices such as glasses / AR/VR headsets.

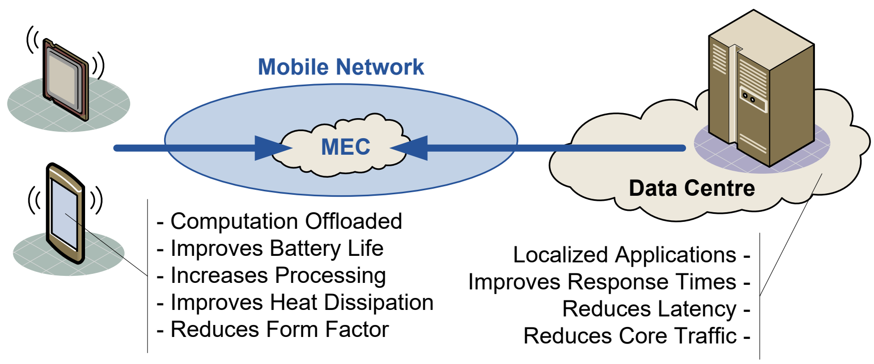


Figure 3-21 Benefits of Introducing MEC

Furthermore, MEC is able to filter / analyse data at the edge, yet still inside the mobile service provider's network, before said data is sent to private or public clouds or is made available through API exposure to other third parties. This is particularly useful for video data as it would enable various polices to be applied such as obscuring faces of any children or licence plates on vehicles. Moreover, MEC aids in the deployment of innovative cyber security mechanisms, utilizing AI to recognise and predict potential threats to the network from hacked devices. Equally, it will also protect the devices themselves from intrusion and cyber-attacks emanating from the network.

3.2.3 Deploying MEC

Figure 3-22 illustrates that MEC can be deployed throughout the access network, with the lowest latency applications potentially mandating that MEC functionality is housed either at or very close to the RAN site (gNB DU / gNB CU). Although the main focus of this course is 5G, MEC also applies to fixed line networks.

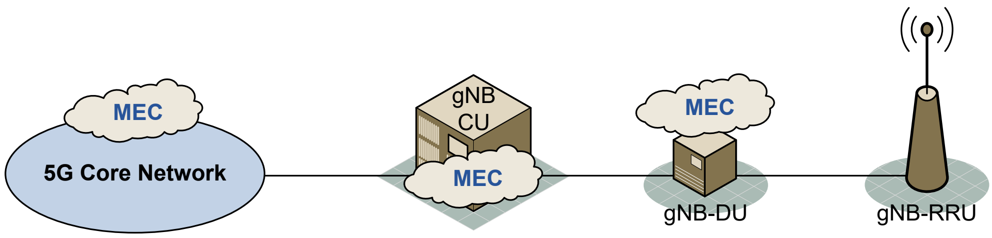


Figure 3-22 Deploying MEC

3.2.4 MEC Use Cases

Autonomous Vehicles

A localized MEC application using AI to build a protective shield for vulnerable road users would take video images from both vehicles in the vicinity, alongside images from additional cameras fitted to street furniture. The AI algorithms would fuse this information in order to identify potentially dangerous situations. As and when they are identified, the system can send back information in near real time, due to the proximity of the AI application, in order to alert the driver or perhaps more importantly, alert the onboard computer controlling the car.

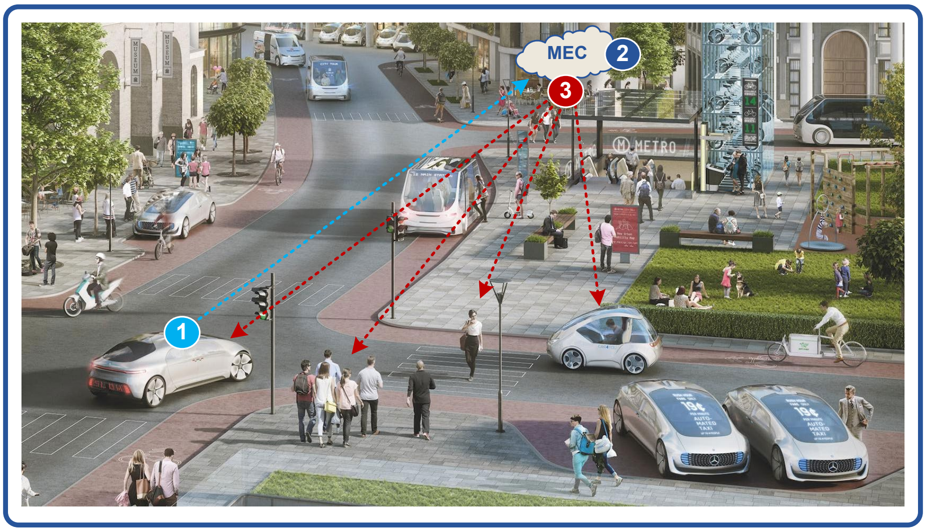


Figure 3-23 MEC Use Case – Autonomous Vehicles.

7. Car send video images to AI application.
8. AI application predicts possible danger.
9. AI application send warning to both vehicles and pedestrians in local area.

As such, this approach may aid in enabling autonomous vehicles to co-exist within an environment containing traditional vehicles, pedestrians and cyclists.

Augmented Reality

In this second use case, MEC is being used to provide additional capabilities to look and listen and then proactively give advice in near real-time. This capability would not be possible on a smartphone operating in a standalone fashion as it would not have either the computational power or storage of the necessary training data to be able to react to the broad range of situations it may encounter. Likewise, pushing this intelligence back beyond the mobile network as applications such as Alexa and Siri do, would bring about too long a delay and thus render much of the potential benefits null.

The actual use cases for this Augmented Reality approach are plentiful: instance guidance for those with hard of hearing or a visual impairment; a dynamic instruction manual reacting to your behaviour as you struggle to build flat pack furniture. The list goes on and covers most verticals including industry, retail, manufacturing, education and training, not to mention healthcare and support for our aging population.

Figure 3-24 MEC Use Case – Augmented Reality

3.3 Voice Services in 5G

Even though 5G is generally discussed in terms of providing enhanced mobile broadband and super-fast data throughputs, not to mention the introduction of billions of IoT devices, 5G still needs to be able to support voice services. This is not purely for altruistic reasons as these services still contribute significantly to the profitability of these mobile networks.

3.3.1 Delivering Voice

Ignoring 1G systems which generally only catered for traditional telephony, the 2G and 3G technologies of GSM (Global System for Mobile communications) and UMTS (Universal Mobile Telecommunication System) both supported voice services using their circuit switched network. This changed with the advent of 4G or LTE (Long Term Evolution) as this was purely a packet switched based technology and as such, voice needed to be carried as VoIP (Voice over IP) and was termed specifically VoLTE (Voice over LTE). Furthermore, to support VoLTE, mobile service providers needed to augment their infrastructure with additional call processing capability in the form of an IMS (IP Multimedia Subsystem).

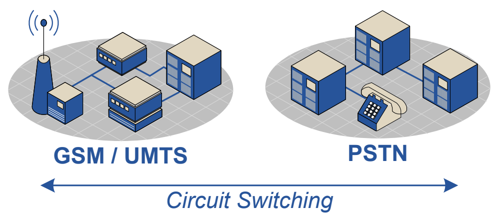


Figure 3-25 Delivering Voice – 2G and 3G

With regards to delivering voice services on 5G or NR (New Radio) networks, the industry has already decided that the IMS telephony service will continue to be used but this time interworking with 5G or NR, hence the term VoNR (Voice over New Radio). Note that depending upon the technology in operation within the PSTN (Public Switched Telephone Network) or other PLMN (Public Land Mobile Network), it may be necessary to “convert” the media flow (voice) back to a circuit switched connection. This conversion would take place on the interconnect between both networks.

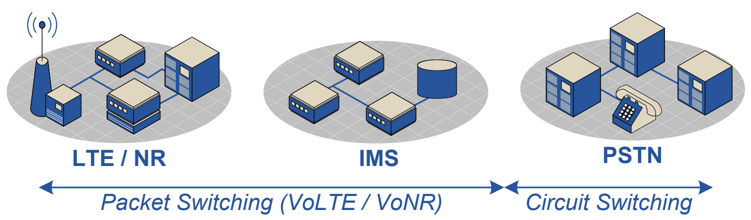


Figure 3-26 Delivering Voice – 4G and 5G

Therefore, the evolution of voice services on 5G begins with 4G VoLTE.

3.3.2 VoLTE / VoNR Principles

Without going into the technical details, the principles behind VoLTE and VoNR are fundamentally the same. Both networks will provide a packet switched connection capable of carrying both the signalling and the voice traffic itself. In fact, due to the specific QoS (Quality of Service) requirements of each, these are comprised of separate logical connections termed EPS (Evolved Packet System) Bearers in 4G and QoS Flows in 5G; one for the signalling dealing with call setup etc and the other carrying the voice.

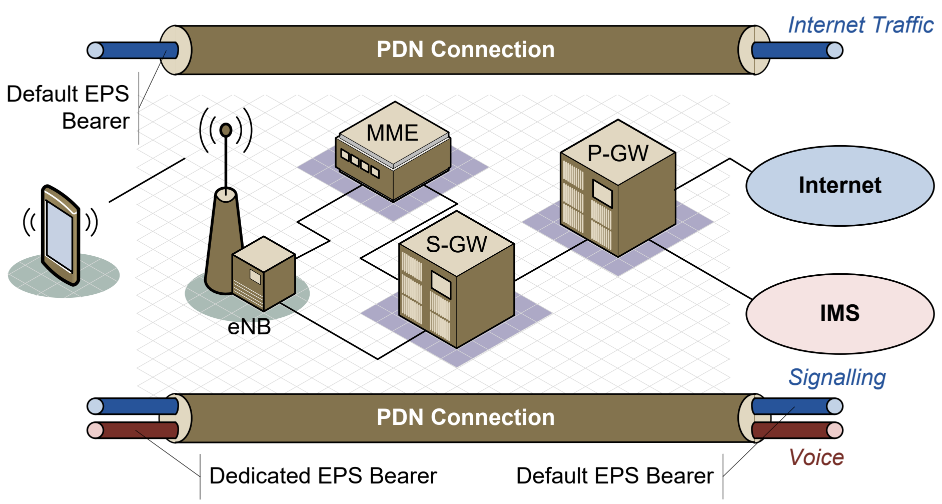


Figure 3-27 VoLTE – PDN Connections and EPS Bearers

When a device powers on and registers with the 4G network, it will typically set up two different types of connection; one destined to carry Internet type traffic such as emails, web pages, movies etc. and a second targeted at the

IMS ready to support telephony. As these terminate on different networks, they are considered to be different PDN (Packet Data Network) Connections. However, within the PDN Connection supporting IMS connectivity, there will be typically two EPS Bearers established, the first one termed the Default EPS Bearer will be used to carry the signalling and will typically remain in place for the entire time the device is connected whilst the second, a Dedicated EPS Bearer, will only exist during the time a call is in place.

This concept is illustrated at Figure 3-27. With regards to VoNR, the same concept holds, with a PDU Session being created towards the Internet and a second towards the IMS; the latter containing two QoS Flows for the signalling and voice traffic respectively. This is illustrated in Figure 3-28.

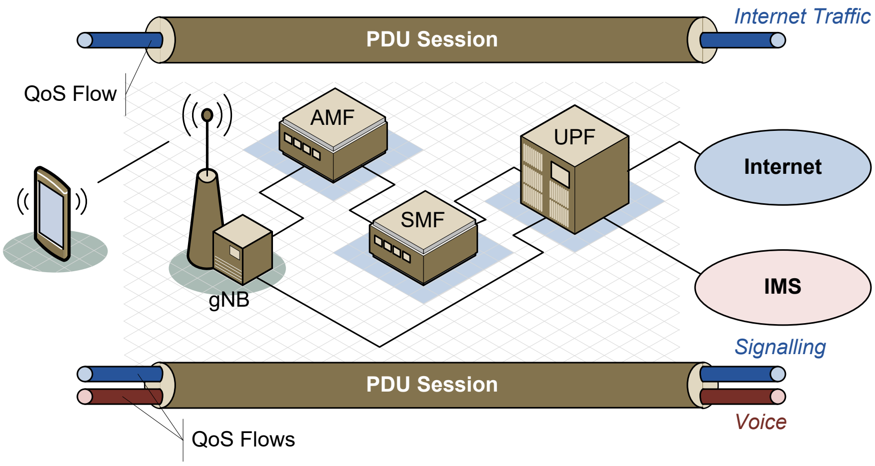


Figure 3-28 VoNR – PDU Sessions and QoS Flows

During their establishment, a QoS profile is associated with each of the EPS Bearers / QoS Flows to ensure the traffic spanning the 4G and 5G network respectively is offered the correct service in terms of Bit Rate (Guaranteed / Non-Guaranteed), Reliability, Precedence etc. These have typically been standardized by the 3GPP (Third Generation Partnership Project) in terms of a QCI (QoS Class Identifier) for 4G and a 5QI (5G QoS Indicator) for 5G. Figure 3-29 identifies the QCI / 5QI values typically used by VoLTE and VoNR respectively along with their associated attribute values.

QCI / 5QI	Resource Type	Priority Level	Packet Delay Budget	Packet Error Rate	Conversational Voice
1	GBR	High	100ms	10^{-2}	Conversational Voice
5	Non GBR	Very High	100ms	10^{-6}	IMS Signalling (SIP)
9	Non GBR	Low	300ms	10^{-6}	www, email, video etc.

Figure 3-29 Typical QCI / 5QI Values used in VoLTE / VoNR

In terms of network operation, once the device has set up the first EPS Bearer / QoS Flow (QCI / 5QI = 5) towards the IMS, it will use this connection to register on the IMS using the signalling protocol SIP (Session Initiation Protocol). Therefore, whenever the device wishes to initiate a call, or a call needs to terminate on a device, signalling will be exchanged between the device and the IMS. This in turn will trigger the setting up of a second EPS Bearer / QoS Flow (QCI / 5QI = 1) which will be used to carry the actual voice traffic. Note this is termed a GBR (Guaranteed Bit Rate) service and as such, it will only remain for the duration of the call. Interesting, although the call is using a packet switched network, calls are still billed based upon call duration / called party address / time of day, rather than volume of data transferred.

3.3.3 Deploying VoNR

Due to the way 5G technology is being deployed, it is not possible to just switch VoNR on and assume everything will be OK; mobile service providers will need to deploy VoNR carefully in order to ensure continuity of service. As such, they will typically follow the following stages.

3.3.3.1 Stage 1 – Voice in EN-DC (Option 3)

Returning to our discussion earlier in the course, we explained that most early 5G deployments were based on dual connectivity in which 5G cells (gNBs) were added to the existing 4G RAN (Radio Access Network) to supplement network capacity and to increase the achievable data rates. This was termed EN-DC (E-UTRA NR Dual Connectivity) but also referred to as Option 3. Moreover, the device remained under the control of the 4G EPC (Evolved Packet Core).

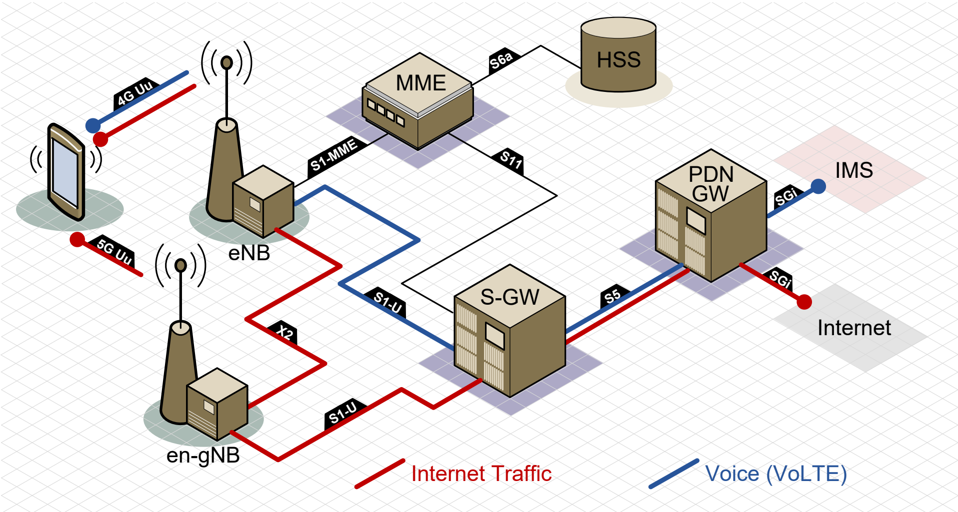


Figure 3-30 Stage 1 – Voice in EN-DC (Option 3)

As such, all voice calls will remain on the 4G network as VoLTE with only the Internet traffic being carried by 5G alone or more likely split between both 4G and 5G, as illustrated in Figure 3-30. Therefore, should the device move out of 4G coverage, the call could be transferred to the legacy circuit switched networks of 2G and 3G using SRVCC (Single Radio Voice Call Continuity). Alternatively, CSFB (Circuit Switched Fallback) may continue to be used to transfer the call.

3.3.3.2 Stage 2 – Voice with EPS Fallback

As the 5G networks evolve beyond EN-DC, two major things will happen. Firstly, the mobile service providers will increase their 5G RAN coverage by adding more gNBs and allocating more radio spectrum to 5G. Secondly, the networks will move to Standalone Operation with the building and roll out of the 5GC (5G Core) network.

However, even though a true 5G Standalone network exists, there is every chance that the 5G RAN will not necessarily have been dimensioned for voice traffic but instead still being targeted towards eMBB (enhanced Mobile Broadband) and the various IoT use cases (MMTC (Massive Machine Type Communications) and URLLC (Ultra Reliable Low Latency Communications)). As such, even though the device is fully operating on the 5G network for its data services, it will fall back to 4G during initial call establishment through either a redirect or inter-system handover procedure. As such, the call will be VoLTE.

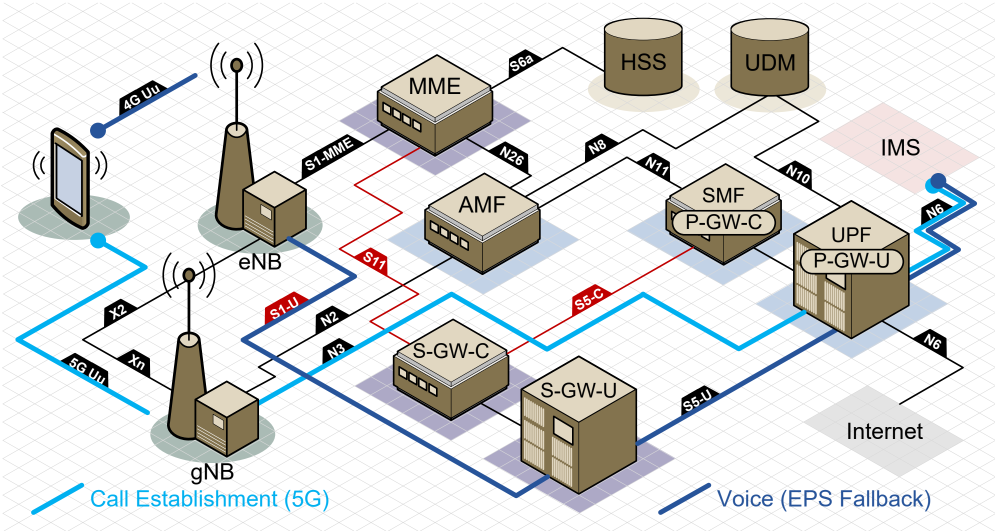


Figure 3-31 Stage 2 – Voice with EPS Fallback

3.3.3.3 Stage 3 – VoNR with Inter RAT Handover to VoLTE

As the networks continue to evolve and more radio spectrum is allocated or re-farmed to 5G, the mobile service providers will be able to ensure that the 5G RAN has been sufficiently well dimensioned to support voice services. Therefore, whist the device is operating within 5G coverage, there is no reason not to run voice services across VoNR negating the requirement to switch back to 4G as and when a call needs to be established. However, it is anticipated that 5G coverage may have a smaller footprint than 4G initially and thus, there will be situations when the device needs to transition back to the 4G network. As such, an Inter RAT (Radio Access Technology) handover would be required (SRVCC) to maintain the call as the device moves out of 5G coverage.

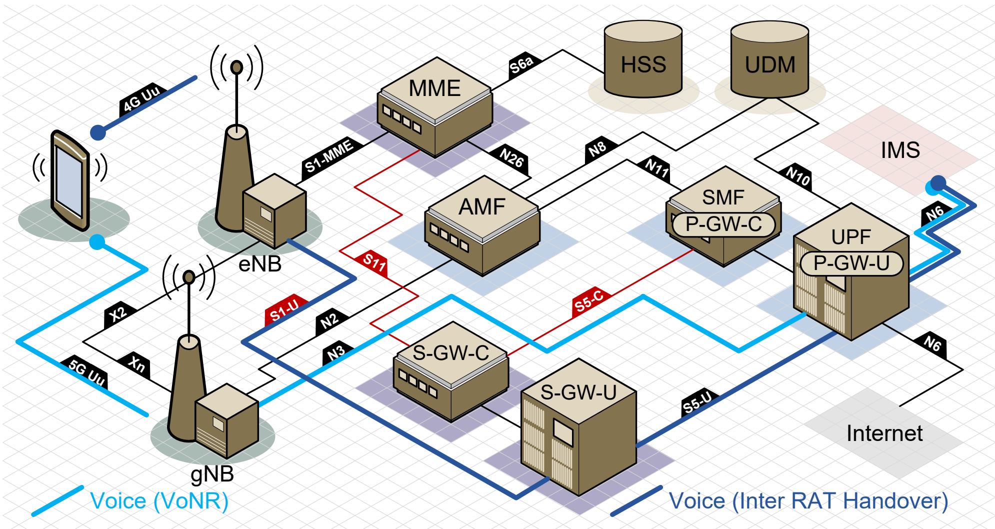


Figure 3-32 Stage 3 – VoNR with Inter RAT Handover to VoLTE

3.3.3.4 Stage 4 – VoNR

Finally, when 5G coverage becomes universal, then there is every indication that the call will remain on the 5G network for its entirety and thus VoNR. However, in rural areas, there may still be a need to support both EPS

Fallback if the 5G radios have not be dimensioned for voice services or Inter RAT Handovers if 5G coverage is patchy so in essence, mobile service providers may well be supporting Stages 2 to 4 simultaneously across their network for some years to come.

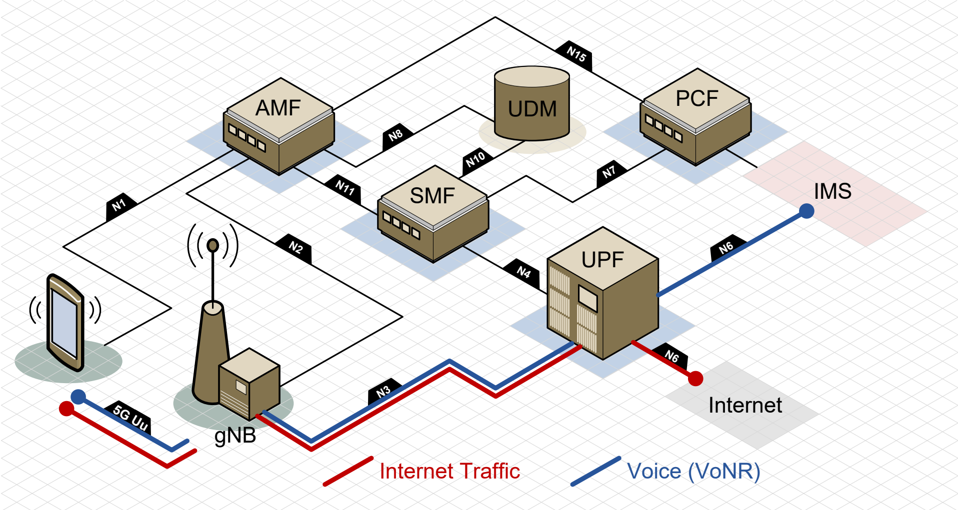


Figure 3-33 Stage 4 – VoNR

3.3.4 VoNR and Roaming

Just as we have come to expect to roam between networks and continue to make and receive calls on 2G, 3G and 4G networks, we will expect to do the same on 5G. To this end, VoNR needs to be able to operate on Visited PLMN's. Figure 3-34 illustrates the simplified 5G roaming network architecture and the path IMS signalling (SIP) and voice traffic will take between the device and the IMS. Note a new network function termed the SEPP (Security Edge Protection Proxy) has been included between the Visited PLMN and the Home PLMN and provides border security for control information passing between the two networks. Examples include message filtering, policy control and topology hiding etc for communication between the AMF (Visited PLMN) and the UDM (Home PLMN) during say registration, and between the V-SMF (Visited – Session Management Function) and the H-SMF (Home – Session Management Function).

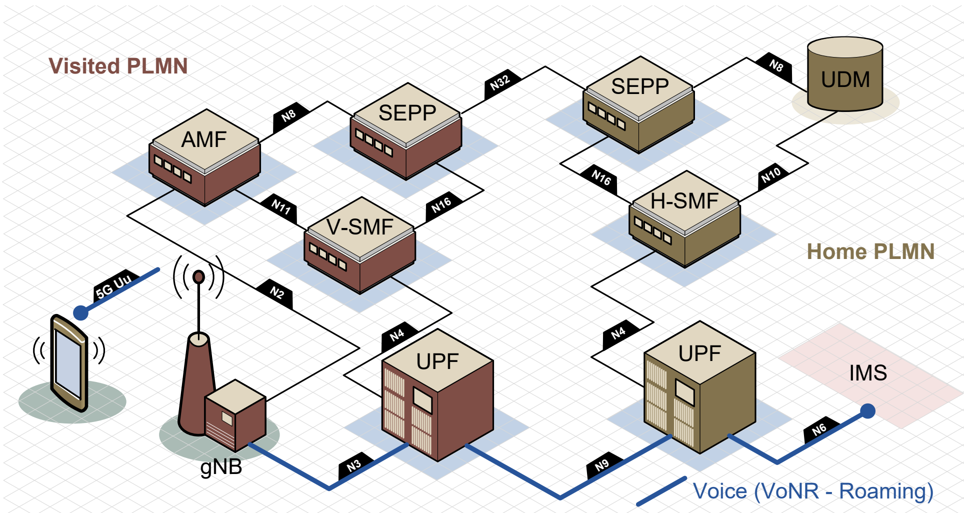


Figure 3-34 VoNR and Roaming

4 Glossary

3GPP (Third Generation Partnership Project)
5GC (5G Core)
5QI (5G QoS Indicator)
AI (Artificial Intelligence)
AKA (Authentication and Key Agreement)
AMF (Access and Mobility Management Function)
API (Application Programming Interface)
AR (Augmented Reality)
AS (Application Server)
AV (Augmented Reality)
AWS (Amazon Web Services)
CIoT (Cellular Internet of Things)
C-RAN (Centralized RAN)
CSFB (Circuit Switched Fallback)
CU (Central Unit)
CUPS (Control and User Plane Separation)
DN (Data Network)
DU (Distributed Unit)
eMBB (enhanced Mobile Broadband)
eNB (Evolved Node B)
EN-DC (E-UTRA – New Radio Dual Connectivity)
EPC (Evolved Packet Core)
EPS (Evolved Packet System)
E-UTRAN (Evolved – Universal Terrestrial Radio Access Network)
FDD (Frequency Division Duplex)
FWA (Fixed Wireless Access)
GBR (Guaranteed Bit Rate)
gNB (New Radio Node B)
GPRS (General Packet Radio Service)
GSM (Global System for Mobile communications)
H-SMF (Home – Session Management Function)
HSPA (High Speed Packet Access)
IMSI (International Mobile Subscriber Identity)
IMT (International Mobile Telecommunications)
IoT (Internet of Things)
ITU-R (International Telecommunications Union – Radiocommunications)
LTE (Long Term Evolution)
MANO (Management and Orchestration)
MEC (Multi access Edge Computing)
MICO (Mobile Initiated Connection Only)
MIMO (Multiple Input Multiple Output)
MMTC (Massive Machine Type Communications)
MTC (Machine Type Communications)
N3IWF (Non 3GPP Interworking Function)
NEF (Network Exposure Function)
NF (Network Function)
NFV (Network Functions Virtualization)
NFVI (NFV Infrastructure)
NIDD (Non IP Data Delivery)
NR (New Radio)
NSO (Non Standalone Operation)
PCF (Policy Control Function)
PDN (Packet Data Network)
PDN-GW (Packet Data Network Gateway)
PDU (Protocol Data Unit)
PLMN (Public Land Mobile Network)
PSTN (Public Switched Telephone Network)
QCI (QoS Class Identifier)
QoS (Quality of Service)
RAN (Radio Access Network)
RAT (Radio Access Technology)
RF (Radio Frequency)
SBA (Service Based Architecture)
SBI (Service Based Interface)

SDN (Software Defined Network)
SEPP (Security Edge Protection Proxy)
S-GW (Serving Gateway)
SIM (Subscriber Identity Module)
SIP (Session Initiation Protocol)
SLA (Service Level Agreement)
SMF (Session Management Function)
SRVCC (Single Radio Voice Call Continuity)
UDM (Unified Data Management)
UMTS (Universal Mobile Telecommunications System)
UPF (User Plane Function)
URLLC (Ultra Reliable Low Latency Communications)
V2X (Vehicle to Everything)
VNF (Virtualized Network Function)
VoIP (Voice over IP)
VoLTE (Voice over LTE)
VoNR (Voice over New Radio)
VR (Virtual Reality)
V-SMF (Visited – Session Management Function)

MPIRICAL

Innovating Telecoms Training

Contact Us

Tel: +44 (0) 1524 844669

Email: enquiries@mpirical.com

www.mpirical.com