October 2014

4G Americas' Recommendations on 5G Requirements and Solutions

4G americas™

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In the wake of the world-wide deployment of 4G technologies, the industry has already started laying the foundation for 5G. Much of this discussion has been outside North America and South America until recently.

The ITU has defined requirements and approved standards for earlier mobile-communication generations: IMT-2000 (“3G”) and IMT-Advanced (“4G”). It will play the same role in defining requirements for 5G and eventually approve standards based on technologies that meet these requirements.

Today’s conversation is centered on predicting what the mobile industry will look like in 2020 and beyond. Some countries have indicated that they would like to see 5G deployed even before 2020. The consensus emerging is that the industry must give enough time for technology breakthroughs that deserve the moniker 5G. Also, there is consensus that because LTE and LTE-Advanced are just being deployed, there is considerable life left in 4G. In fact, the LTE family of technologies should remain viable through at least 2020 because they’ll continue to evolve and advance in terms of higher speeds and greater capacity. Carrier aggregation (CA), small cell enhancements and device to device signaling are just some of the examples of how LTE is advancing.

When will 5G be deployed? What will characterize networks in 2020 and beyond? What are the likely solutions and technologies that will come in to play? These are some of topics currently being discussed among operators, the supplier community, research institutions, standards bodies, trade organizations and governments. Examination of 5G requirements and solutions is basically an exercise in planning a network evolution plan that spans six to seven years. While past generations have been identified by a major new technology step, such as the definition of a new air interface, the expectation is that 5G will be approached from an end-to-end system perspective and include major technology steps both in the radio access network and the core network. These steps can be evolutionary, or they can be revolutionary by introducing a completely new concept.

This paper examines the 5G market drivers, use cases, requirements, regulatory considerations and the technology elements. The 5G market drivers and use cases described in Section 2 include the Internet of Things (IoT), extreme video, PSTN sunset, public safety and context-aware services. Based upon these 5G market drivers and use cases, Sections 3 and 4 describe the requirements and regulatory aspects such as low latency, high throughput, mobility on demand, high reliability and resiliency and network flexibility. Section 5 describes potential technologies for 5G, including packet core, RAN and aspects applicable to end-to-end 5G systems. Section 6 discusses the spectrum aspects associated with 5G networks including both licensed and unlicensed spectrum options.

These 5G requirements and recommendations have been identified by 4G Americas and have been developed for the purpose of being considered for the further development of the end-to-end 5G system.

1. INTRODUCTION

Wireless has evolved since its inception and seen several generational changes to the service offerings. The generational term is meant to help delineate differences between technologies. Discussions are ramping up about fifth-generation (5G) wireless access. 5G is associated with the next step of IMT, IMT-2020, initial planning for which is currently under way in the ITU. Additionally, a number of other changes in the end-to-end system will be part of 5G evolution, both in the RAN and the core network.

There seems to be broad consensus that 5G will be introduced around 2020. However, the work on 5G is still in its early stages in terms of examining use cases, requirements and component technologies. A primary question is whether 5G will include another new air interface or a collection of air interfaces, each for a different scenario and use case.

Because the industry is working to define the 5G requirements for deployments in the 2020 timeframe without having actual technology developed, one should be cautious that some requirements may not be feasible in the expected timeframe.

The aim of this paper is to address the following areas from a North American perspective:

- What are the key use cases and corresponding key challenges and requirements for wireless beyond 2020?
- What are the key new technology components and solutions that can be used, in combination, to address these challenges and requirements?

It should be noted that the future of 5G wireless access as referred to above is much more than just about radio-interface technology. 5G wireless access should be seen as the overall future solution to providing wireless access to people and devices.

A clear definition of 5G or 5G requirements is not yet available. However, requirements such as support of large number of connected devices, “Always online,” energy efficiency and support of flexible air interfaces may not be achieved by just an evolution of current systems. Instead, those requirements may require 5G to have new protocols and access technologies.

2. MARKET DRIVERS AND USE CASES FOR 5G

3G and 4G technologies have mainly focused on the mobile broadband use case, providing enhanced system capacity and offering higher data rates. This focus will clearly continue in the future 5G era, with capacity and data rates being driven by services such as video.

But the future also will be much more than just enhancements to the “conventional” mobile broadband use case. Future wireless networks should offer wireless access to anyone and anything. Thus, in the future, wireless access will go beyond humans and expand to serve any entity that may benefit from being connected. This vision often is referred to as “the Internet of Things (IoT),” “the Networked Society,” “machine-to-machine communications (M2M)” or “machine-centric communications.” North American operators’ best customers are no longer humans; they’re increasingly machines such as smart utility meters, digital signage and vehicle infotainment systems.

As users begin to use more interconnected devices to play games and collaborate, the ability of the devices themselves will need to expand to create personal networks. Even machines that need to talk to wide-area networks for alarm monitoring, home health, fleet management and many other applications will continue to grow. It is not inconceivable for machine customers to outnumber human customers.

The industry foresees a future of wireless of an increasingly interconnected world where voice, video, medical, entertainment and other applications and services will be served by a highly integrated and automatically configurable network. Users will simply request the information they need, and the information will be delivered to their desired location and device.

Issues that need to be addressed are the ways in which the user community will need to interact with information in an environment where on-demand, high-speed mobile data has become a reality. That will mean some significant changes in the way that user interfaces are designed and higher-layer features are developed. The Internet model as we know it today may not be what best serves 5G users.

In summary, 5G is about enabling new services and devices, connecting new industries and empowering new user experiences. This will entail connecting people and things across a diverse set of scenarios.

This section describes several envisioned 5G use cases across multiple industry verticals. It should be noted that what is described is only a subset of the use cases that can be envisioned. Also, new yet unknown use cases will most likely emerge, and 5G should have the flexibility to adapt to them.

2.1 INTERNET OF THINGS (IoT)

A wide variety of cellular-enabled IoT applications will be prevalent by 2020, ranging from smart utility grids to earthquake/tsunami detection sensors that warn the public. All such applications can and are beginning to get deployed even on today's cellular networks. However IoT applications are predicted to grow at a much faster pace than perhaps what existing networks and cellular technologies can optimally handle.

To support possibly billions of IoT devices, a wireless network infrastructure is needed that's not only highly scalable in terms of its capacity, but can also optimally handle differing service needs of various IoT verticals. Examples of differing service needs include diverse requirements for mobility, latency, network reliability and resiliency. These diverse requirements may require re-architecting key components of the cellular network, such as to support mobility on demand only for those devices and services that need it. The following example use cases involving Machine Type Communication (MTC)[1] will become the norm in the 2020 timeframe.

2.1.1 SMART GRID AND CRITICAL INFRASTRUCTURE MONITORING

Today's societies depend on a wide array of critical infrastructure to function properly. Malfunction or damage to this infrastructure could result in huge financial impact, quality of living degradation and even loss of life. The 2003 power blackout in the Northeastern U.S. is an example of how infrastructure failure can bring an entire region and its economy to a halt. Other examples of such disruption include bridge and building structural failure leading to collapse, or water and sewer systems malfunctioning. It is therefore important to monitor the "health" of critical infrastructure reliably and cost effectively.

Critical infrastructure monitoring is an expensive undertaking, often requiring service levels achievable only by dedicated wire-line connectivity. For instance, in order to detect a fault in a high-voltage transmission line and be able to take corrective action to prevent cascading failures, the required communication latency is beyond what current wireless networks can achieve. Similarly, structural

---

[1] Within the industry several different terms are used to describe machine to machine communications, these terms include IoT, M2M, MTC, etc. This white paper uses these terms interchangeably.

monitoring requires the provisioning of a large number of low-data-rate, battery-powered wireless sensors, which today's wireless networks are not optimized to support both in terms of battery life and cost efficiency. 5G will be designed to support reliable low-latency communications among densely deployed devices that are subject to power constraints and wide-ranging data-rate requirements.

2.1.2 SMART CITIES

Massive urbanization is an ongoing trend around the world that's severely straining city services, resources and infrastructure. According to the World Health Organization, by 2030, six out of every 10 people will live in a city. Smart-City initiatives aim at improving cost, resource, and process efficiency of cities, while maintaining a high living quality for their rising populations. The following are three potential examples of 5G-enabled Smart City use cases:

Smart Transportation: Traffic congestion is becoming a major issue in many urban areas, and it's leading to productivity loss, environmental pollution and degradation of quality of life. 5G will enable real-time collection of massive amounts of data from vehicles, drivers, pedestrians, road sensors and cameras to help streamline traffic flow. For example, it can help optimize traffic lights and road usage, direct public transportation to where it is needed most, navigate vehicles to avoid congestion and raise tolls to limit traffic entering a congestion zone.

Smart Building: Urban buildings are major consumers of energy and resources. Streamlining building operations will lead to increased productivity and energy efficiency. For example, 5G-connected sensors/actuators can help optimize building temperature, humidity and lighting based on current activities inside them. They will also enable buildings to detect when hidden pipes and cables need repair, when unauthorized access takes place, when office supplies are running low and even when garbage bins are full. This information allows building management to take appropriate action in a cost-effective and timely manner.

Smart Home: Home security and automation applications constitute another M2M service area that is expected to grow significantly in the future. Examples include the transmission of home security alarms and home surveillance video data to commercial monitoring stations.

2.1.3 M-HEALTH AND TELEMEDICINE

Telemedicine is a major tool for improving healthcare access both in remote rural and urban areas. It can help reduce health care costs while improving health outcomes. A major enabler in this field is the comprehensive use of cloud-based electronic medical records that would serve as repositories of medical information about individual patients. With 5G, these records, which contain high-resolution medical images and video, can be made available to physicians and medical professionals anytime and anywhere. Remote, real-time general physician and specialist consultations would also contribute to cost savings, convenience and better and timelier medical outcomes.

A major hurdle in the realization of this scenario is the lack of a wireless infrastructure that would need to handle the voluminous nature of medical images and video with sizes ranging from hundreds of Megabytes to Gigabytes per instance. The increasing use of diagnostic tools such as 3D and 4D ultrasounds[2], CAT scans and MRIs, and the miniaturization of this equipment to a portable/hand-held form factor will lead to even higher demands being placed on wireless networks. In addition, massive improvements in the quality of low-cost mobile displays and application software now available to medical professionals make this field ripe for massive adoption. Bio-connectivity, which is the continuous and automatic medical telemetry (e.g., temperature, blood pressure, heart-rate, blood glucose) collection via wearable sensors, is another strong emerging trend that will add to the wireless communications requirements. 5G will enable these and other future medical applications through significant improvements to wireless data throughput and network capacity.

2.1.4 AUTOMOTIVE

Advanced Driver Assistance Systems (ADAS) and Autonomous Vehicles are emerging trends in the automotive space. Together, they bring a number of benefits, including better safety, fewer collisions, less congestion, better fuel economy and even higher productivity for the drivers. 5G wireless technologies supporting high-speed, low-latency vehicle-to-vehicle and vehicle-to-infrastructure communications are key enablers of ADAS and Autonomous Vehicles. In addition, today’s drivers and passengers are demanding richer infotainment options, which are adding to the strain on wireless networks. The following section describes several potential automotive use cases for 5G.

Vehicular Internet/Infotainment: Increasing content consumption by vehicle occupants will greatly contribute to the need for wireless bandwidth and mobile network capacity. Typical infotainment options include video, audio, Internet access and upcoming applications such as augmented reality and heads-up displays. For these applications, vehicle occupants will expect a user experience comparable to those offered by their home and office networks. The vehicles themselves form another group of Internet users for map, traffic data and high-resolution picture download, as well as sensor data and image upload.

Pre-Crash Sensing and Mitigation: Collisions lead to injury and property damage, as well as time and productivity loss due to traffic congestion. Pre-crash sensing enables vehicles to sense imminent collisions and exchange relevant data among vehicles involved, allowing vehicles and drivers to take counter-measures to mitigate the impact of the collision. Pre-crash sensing requires highly reliable and extremely low latency vehicle-to-vehicle communications.

Cooperative Vehicles: Limited highway capacity in many cities often results in severe traffic congestion. Cooperative vehicles use Vehicle-to-Vehicle and Vehicle-to-Infrastructure communications to safely operate vehicles as a self-driving car train on a highway in order to improve highway capacity, reduce occurrence of driver error and achieve better fuel economy. To ensure safety and reliability while operating as a self-driving car train, reliable and very low-latency communications among vehicles and with the infrastructure are needed.

Inter-Vehicle Information Exchange: Peer-to-peer inter-vehicular communication using D2D cellular technology under the guidance of the operator policies can allow vehicles to communicate information related to road safety and traffic congestion directly in a mesh fashion, thus offloading data from the traditional RAN infrastructure. This is just one possible example of the type of information that can be exchanged.

---

[2] 4D Ultrasound are ultrasound videos with 3D images where the 4th dimension is time. Most 4D ultrasound videos today are less than 5 frames per second. For greater diagnostic utility, they would need to be much higher resolution and about 30 frames per second. For cardiac applications, greater than 100 frames per second would be desirable.

2.1.5 SPORTS AND FITNESS

Fitness-related applications, such as activity and body monitoring applications that track walking, running, and biking activities, metabolic rate, cardiovascular fitness, sleep quality, etc. will constitute a significant vertical market in M2M services. Some of these applications will utilize body or personal area networks to collect biometric information and then use cellular networks to transmit it back to centralized data acquisition sites.

2.2 EXTREME VIDEO, VIRTUAL REALITY AND GAMING APPLICATIONS

Future wireless communication systems will support extreme video and gaming applications that use features such as augmented and virtual reality. Such immersive multimedia services would require the use of technologies such as 3D audio, 3D video and ultra-high-definition formats and codec(s). Examples of such services include:

- Mobile telepresence with 3D rendering capabilities that will extend well beyond the traditional wired office environment.
- Internet gaming, including wirelessly delivered gaming control with high-resolution graphics and dynamic management of feedback mechanisms via smartphone to ensure an enhanced, augmented reality gaming environment.
- Adoption of higher resolution devices, head-mounted displays and wearables in fields such as emergency services, public safety, telemedicine, smart cities, professional services and retail is expected to place further demands on mobile networks.

This type of interactive experience will require the network to support much lower latencies and much higher bandwidths than are possible today.

2.3 EXPLOSIVE INCREASE IN DENSITY OF DATA USAGE

The use cases outlined so far in this section identify some general trends in the industry:

- The number of devices using cellular networks is expected to increase significantly in the coming years. In other words, the density of cellular devices (devices:area) will increase. A large part of this increase will be coming from M2M services.
- Some future services will require much higher data rates compared to what is typically achievable today. Examples of such services have been provided in the previous use case on extreme video and Internet gaming.

Concentration of devices using Ultra High Definition (4K and 8K) video and high-resolution picture and video-sharing applications occur at event venues such as stadiums. In addition, significant variations in UL:DL traffic ratios imply a need for an air-interface design that can more flexibly assign traffic capacity to the different transmission directions. The effect of these two factors on network traffic will be multiplicative, resulting in an explosive increase in data traffic demand per square mile of the coverage area, especially in urban environments. Some of this is validated by looking at existing traffic trends where data traffic density in urban environments such as stadiums, financial businesses, hospitals, universities and major transportation corridors has increased dramatically. To handle this surge of data traffic demand by 2020, network capacity would have to be increased by orders of magnitude.

2.4 PUBLIC SAFETY

The U.S. is planning to deploy an LTE broadband network for public safety at 700 MHz to leverage pricing of standardized commercial equipment. Canada is currently evaluating the use of an LTE broadband network for public safety at 700 MHz. It seems natural that future wireless broadband networks will also need to consider public safety in their fundamental design. Some of the “special” public safety needs include:

- **Mission-Critical Voice:** Allowing a public safety responder to push a button (push-to-talk) to communicate with other public safety responders. This needs to be extremely reliable, working both on and off network without any delay for dialing phone numbers. The feature needs to allow communication with one or more groups (e.g., local police, regional police, and local public safety) in real time. Public safety users must be able to monitor multiple groups simultaneously (scanning communication on different groups) and allow additional users to join an on-going group discussion.

- **Broadband Data:** Much of this will be IP traffic from a public safety device to a server, possibly in the cloud. Although this capability can be handled by existing LTE equipment, it is important that 5G consider the following public safety use cases:

  - High-resolution security cameras monitoring public spaces and property with the captured images/video analyzed to alert authorities when incidents occur or persons or interest are detected.

  - Drone- or robot-based surveillance systems to monitor remote areas.

  - Wireless sensors and tracking devices used for intrusion detection, bio and chemical hazard detection and emergency personnel tracking.

  The data generated by these and many other modalities will significantly strain 4G radio link and networks.

Besides these needs specifically for public safety officials, 5G systems will need to support legacy public safety features such as public warning systems (PWSs), emergency calling, multimedia emergency services (MMES) and lawful intercept. To support all such use cases, future wireless networks must provide a robust, highly reliable, resilient, and low-latency communication infrastructure.

2.5 PSTN SUNSET

The Public Switched Telephone Network (PSTN) sunset in North America is scheduled prior to 2020. With the general industry trend of migrating towards wireless communication, it is expected that in the 2020 timeframe and beyond, wireless broadband networks will be commonly used to replace the PSTN. Therefore, the 5G ecosystem must also serve today’s landline needs. For 5G networks to be considered a viable replacement to PSTN, they must exhibit the same levels of reliability and robustness. In addition, PSTN services primarily serve stationary customers that do not require support for mobility. The concept of mobility-on-demand that can simplify the packet core and make it scalable should be explored.

2.6 CONTEXT-AWARE SERVICES

The past decade has seen a tremendous rise in the use of always-on, Internet-connected devices. The users of such devices are consistently bombarded with information, most of which may not be relevant or actionable for them. For the most part, existing service models require users of such devices to reach out to the Internet to get the useful information and/or service that they desire.

In such a service model, amongst other things users first have to figure out the best match for their request and then find out how to get to it. With the ever-increasing amount of available information, it is quite evident that this service model is not scalable. A desired approach is for a service to be context aware and to be able to provide a seamless delivery of the right set of information at the right time using the right means. This approach can also be described as instead of the user going to the Internet and figuring out a way to fulfill its needs, the Internet comes to the user with the right information [1].

3. REQUIREMENTS FOR 5G

Requirements can be subdivided into two categories:

- User-driven requirements in terms of quality of experience, user satisfaction, reliability and speed of the connection.
- Network-driven requirements in terms of network operation and management.

3.1 USER-DRIVEN REQUIREMENTS

3.1.1 BATTERY LIFE

Several of the IoT applications involve battery operated sensor networks that are out in the field and transmit data only occasionally. Wide-scale deployment of 5G-based sensor networks would be possible only if much longer battery life and/or reduced energy consumption by such devices guarantees their unattended operation over a duration spanning years.

3.1.2 PER-USER DATA RATE AND LATENCY

Per-user data rate and latency attributes for a network define the typical data rate and round-trip delay, respectively, that users experience. Ultimately the values for these attributes determine the types of applications that can be supported on a network. It is estimated that by 2020, there will be a new class of data-rate-hungry services with low latency requirements. Previous work has shown that applications in the future such as augmented reality, 3D gaming and “tactile Internet” will require a 100x increase in achievable data rate compared to today and a corresponding 5x to 10x reduction in latency. 5G networks must therefore be designed to meet these data-rate and latency requirements.

3.1.3 ROBUSTNESS AND RESILIENCY

5G networks will increasingly be used as the primary source of communication as a replacement network for PSTN after its sunset. They also will support emergency communications and public safety, including during and after disasters. A key requirement for such use cases is for the network to be robust, reliable and resilient. This requirement would also need to ensure the ability to defend against security attacks such as denial of service (DoS) for mission-critical applications such as public safety, smart grids and natural gas and water distribution networks.

3.1.4 MOBILITY

5G systems are expected to support both very-high-mobility scenarios (e.g., high-speed trains, planes), as well as scenarios with low to no mobility for end devices. The technology therefore should be able to cope efficiently with such extreme situations by providing mobility on demand based on each device's and service's unique needs and capabilities.

At the same time, machine-type communicating devices can require nomadic access to the network with the purpose of sending reduced amounts of data in mostly static locations. For these nomadic devices, reliability and resilience could be most important network features than mobility support. Finally, the need for extreme data rates (or extremely low latencies) at specific situations can usually be satisfied with very stable channel conditions in stationary devices. Future 5G systems will have to cover such extreme cases, from no mobility to future high-speed trains or even possibly aircraft.

3.1.5 SEAMLESS USER EXPERIENCE

Current cellular systems provide very high peak quality, such as very high peak data rates. However, the quality often varies substantially over the coverage area of interest. For example, the achievable data rates can be substantially lower for devices far from the base station site or in indoor locations. 5G wireless access should deliver a much more consistent user experience, irrespective of the user's location. Thus, the achievable quality (e.g., achievable data rate, achievable latency.) should be the key quality indicator, where achievable quality should be defined as the quality experienced with perhaps 95 percent probability, rather than, for example, peak data rates.

5G will likely comprise a collection of layers, technologies and frequency bands that should seamlessly interwork when moving across networks, layers and/or frequencies. Interruption times of the order of a few milliseconds for both inter-RAT and intra-RAT handovers can be expected in this sense. Services such as ultra-high definition video or tactile Internet will require end-to-end latencies in the order of 1 millisecond. Interruption times well beyond that could destroy the attractiveness of these services.

3.1.6 CONTEXT-AWARE NETWORK

With MTC and greater diversity in human communication devices, it becomes increasingly important for the network to provide the correct resources to meet the unique needs of each application and device. This is possible only if the network is context-aware and hence can dynamically adapt to meet those needs. This means, for example, that full mobility need not be provided to MTC devices that are stationary, that 3GPP mobility management and paging need not be provided for services that require only device-initiated communication and that resources are configured to support long battery life, high reliability, low latency, low cost, secure communications, global roaming and the like as is truly needed. Optimizing resource allocation in this manner makes possible simpler, lower cost, application-tailored devices and lower network costs because only the necessary resources are used. It also enables a better end-user/device experience. Mobile operators gain additional abilities for creating service plans customized for individual customers, groups of customers or market segments. An example is creating one set of plans for MTC customers that use a limited amount of network resources, and a second set of plans for smartphones that use a large amount of network resources.

Context includes network awareness, such as the availability of alternative multi-RAT, small cell and macro networks of varying capabilities, application and device awareness with associated service requirements, subscription context such as operator preferences for providing service and subscriber analytics. Awareness of these attributes makes it possible for the network to dynamically adapt to the needs of devices and applications rather than have applications adapt to today's one-size-fits-all set of access characteristics.

3.2 NETWORK-DRIVEN REQUIREMENTS

3.2.1 SCALABILITY

Support for IoT use cases will be key to the success of 5G networks. An expected 10X-100X increase in the number of devices, primarily because of M2M services, requires network elements that can scale up gracefully to handle this growth. This requirement is true both for the user plane and the control plane. For example, the 5G network should be able to scale well to handle signaling traffic, such as for authentication/authorization for large numbers of IoT devices. Another example is that the user plane must be able to scale well to handle (in)frequent and small data transmissions from large number of devices.

When it comes to supporting a mix of traffic from IoT applications and more traditional services such as voice and video, the term "scalability" has an additional dimension. 5G networks should be able to support both high-data-rate/low-latency conventional services alongside M2M applications that require much lower bandwidths. Each M2M vertical will likely have its own unique traffic pattern. Both frequent and infrequent (bursty) data transmission, for example, will have to be supported in an efficient manner. For example, traffic pattern and transmission requirements (data rate, latency) from earthquake/tsunami warning sensors will be quite different than for traffic from a vending machine. Similarly, many devices in 5G networks will be stationary or nomadic and require no mobility support or only occasional mobility support. 5G network designs therefore should not assume mobility support for all devices and services but rather provide mobility on demand only to those devices and services that need it.

3.2.2 NETWORK CAPACITY

Experience suggests that we can expect a 1000x – 5000x [2],[3] rise in traffic over the next decade. To handle this explosive increase, a key requirement for 5G networks will be to increase traffic-handling capacity – which is defined as the total traffic that the network can handle – while still maintaining QoS.

3.2.3 COST EFFICIENCY

With an expected increase in the total network traffic, and the need to stay competitive, the next generation of mobile networks should provide a significant cost benefit over the current generation. The cost improvement should be at least as good as or possibly much better than what we experienced in going from 3G to 4G. In this context cost refers to both the OPEX and CAPEX of delivering a byte of data to the subscriber. Network Function Virtualization (NFV) will play a key part in achieving cost reductions.

3.2.4 AUTOMATED SYSTEM MANAGEMENT AND CONFIGURATION

In 5G deployments, the network density is expected to significantly increase for a number of reasons including higher data volume density and the use of higher frequency spectrum. To better manage the CAPEX and OPEX of running a network with a much higher number of network nodes, a key requirement is that 5G networks will be able to self-configure as much as possible.

3.2.5 NETWORK FLEXIBILITY

The 5G network architecture should allow the RAN and the core network to evolve and scale independently of each other. Changes/enhancements to one should not mandate changes/enhancements to the other. In order to achieve this goal, the RAN and the packet core should avoid mutual dependencies. Additionally, 5G networks must also support multi-RAT connectivity efficiently and effectively. This includes the ability to:

- Provide an access-agnostic packet core across multiple radio technologies (e.g., cellular, Wi-Fi) to support uniform authentication, session continuity and security.
- Provide plug-and-play capability where a new access technology may be attached to the packet core without any modifications.

Decoupling the packet core from the RAN will also help efforts to further flatten out and simplify the network. Flattening improves network scalability by enabling some of the functions to be pushed closer towards the user/edge.

Revisiting the current architecture for mobility and operator policy management can help minimize interdependency between the RAN and the packet core. As part of this effort new mobility protocols, the concept of mobility on demand and the use of Software Defined Networking (SDN) for providing and enforcing operator policies should be explored.

Recent trends in virtualization envision a separation of the control and data planes, as well as a decoupling of hardware and software so that network functions are mainly driven by software (with hardware equipment being as generic as possible). These principles would enable greater flexibility in deploying network functions on demand.

3.2.6 ENERGY EFFICIENCY

Rather than only maximizing spectral efficiency, there is an ever growing concern about the energy consumption per bit (expressed in Joules/bit) that represents a measure of the energy efficiency. Network functions should not convey excessive energy (both radiated and consumed by the network infrastructure). More interestingly, energy consumption could be adapted to the current traffic conditions to achieve significant energy savings in off-peak situations.

3.2.7 COVERAGE

Although coverage is ultimately limited by the band in use, 5G will take special care in improving coverage for IoT-related applications to make 5G viable for this emerging market. While it is clear that coverage largely depends on the frequency of operation and density of deployment sites, special actions can be taken so as to ensure optimal coverage for specific services such as IoT, public safety and other critical systems.

3.2.8 SECURITY

Mission-critical applications such as smart grids, telemedicine, industrial control, public safety and automotive, have strict security requirements to defend against intrusions and to ensure uninterrupted operations. 5G should address the following security objectives:

- **Integrity:** Ensure information is not tampered with either accidentally or deliberately during transit. This includes the ability to authenticate the source of the received information and the ability to authenticate the recipient.

- **Confidentiality:** Keep sensitive information away from unauthorized users. This includes proper user authentication, data protection through encryption, etc.

3.2.9 DIVERSE SPECTRUM OPERATION

To serve the use cases outlined in the previous section, 5G is expected to operate in a diverse set of spectrum bands. These include traditional sub-6 GHz cellular bands for coverage and low-power operation, to above-6 GHz bands including millimeter spectrum for ultra-high data rates. As the propagation characteristics and hardware implications of these bands are expected to be substantially different, 5G systems would need to accommodate these requirements in radio access, network architecture, protocol and modem design considerations.

3.2.10 UNIFIED SYSTEM FRAMEWORK

Use cases described in the previous section have very diverse, and sometimes conflicting, requirements. For example, sensor applications generally require low data rates and can tolerate high latencies. Meanwhile, evolving applications such as telemedicine require high data rates and low latencies. 5G should be as flexible and extensible as possible to support existing and future use case requirements, thus avoiding the need to introduce a dedicated system for each emerging use case.

3.3 ASPECTS OF 4G NETWORK ARCHITECTURE THAT CAN BE ENHANCED BY 5G

The basic principles of 4G network architecture were conceived several years ago prior to the explosion in mobile broadband usage. Many of the requirements that have emerged since then have been handled by incremental modifications to the basic architecture. Although capable of satisfying the requirements, 4G architecture does so by increasing the complexity of existing functional components and by adding new functional components. We describe some of the limitations that we see in 4G architecture that can potentially be improved in 5G architecture.

3.3.1 ENHANCEMENT OF NETWORKING FLEXIBILITY

With the advent of small cells in indoor environments such as offices, there is a need for some traffic to be routed locally while other traffic needs to access MNO or third-party services. For example, as a result of enterprise bring-your-own-device (BYOD) policies, devices increasingly have multiple “personalities,” with some applications communicating within the private office environment while other applications communicate with Internet-based consumer services. For local offloading, 4G requires a separate mobile packet gateway (PGW) deployed locally largely because mobile-network-specific tunneling is employed for all traffic.

Advances such as content distribution network (CDN) virtualization are being made that enable content caching closer to the device at various locations in the transport network between the base station and the core network. Intelligent-content-request routing mechanisms are being proposed in the context of content-centric networking, some of which may be applicable to traditional CDNs, as well. In order for applications on the mobile device to leverage such innovations, a local mobile gateway has to be deployed, and devices need to support multiple access point name (APN) connectivity or the network has to initiate an APN switch, which may result in some disruption of ongoing services.

In addition to deploying local gateways, new APNs have to be provisioned in the network and devices, which is sometimes a complex process for MNOs. A solution that eliminates the need for such a specialized mobile specific local gateway, new APN provisioning and associated signaling is preferable.

3.3.2 ADDITIONAL SUPPORT FOR ESSENTIAL FUNCTIONS AS FUNDAMENTAL ATTRIBUTES OF NETWORKING LAYER

Like mobile architecture, Internet architecture continues to evolve to support new use cases. One challenge is that IP was designed at a time when the network's fundamental objective was to transport data packets between fixed communication hosts quickly and efficiently. With packet headers naming the communication hosts via the IP addressing scheme, the network task has been simply to forward the packets hop-by-hop from one host to the other. Significant elegance and operational efficiency derived from routing protocols have enabled automatic topology mapping and pathology-free routing without significant operator intervention.

In response to the emergence of more elaborate use cases and usage demands (e.g., mobility, content distribution, security), the networking community has incrementally added new functions as either overlays on the existing network or as specialized elements in the network to address the need. The key point is that the importance of capabilities considered essential today was not fully appreciated at the time the original Internet design was conceived. As a result, these capabilities were not incorporated as fundamental elements in the original design. So over the years, the cost of managing and operating the network has progressively increased, largely due to the added complexity introduced by the persistent stream of functionality patches and overlays (support for mobility amongst them).

5G architecture should re-examine the mobile network architecture from the perspective of how it stands to benefit from research in the Future Internet Architecture, and in particular network architectures and protocols that implicitly support mobility, security and content caching (storage) as fundamental components of the network design.

3.3.3 PROVIDING MORE FLEXIBLE MOBILITY SOLUTIONS

It is well known that even in a mobile network, many devices are stationary. For example, video constitutes 55 percent of the traffic, and people are generally static when watching video. Many M2M devices, such as utility meters, are stationary. Furthermore, even if the device is moving, maintaining the same IP address is not required for proper functioning of many applications.

An example is HTTP adaptive streaming (HAS). This application works by downloading 2 seconds' worth of chunks that are buffered in the client with a deep buffer of tens of seconds. If downloading of a chunk is interrupted by a handoff, a new TCP connection with a new IP address can be set up to download the same chunk again. Another example is tracking devices on pets/people where the device is periodically updating the network about the wearer's location. If an active session is interrupted a new session can be set up to perform the location update.

The functions that handle device mobility in 4G are oblivious to the specific requirements of the applications and devices. The same seamless mobility is always provided, incurring additional signaling, processing, memory and bandwidth overheads. More specifically, inefficiencies in 4G include the following:

1. Establishing and modifying the tunnels when they are not needed incurs a number of signaling messages between various network elements. The RAN signaling protocol used in LTE to track both idle and active UEs is an adaptation of legacy 2G circuit voice systems designed for relatively low volumes. Voice calls are typically of longer duration and happen much less frequently than data interactions, especially when data transactions are associated with large volumes of M2M devices. As a result, today's 3GPP handling of mobility entails significant RAN and core network signaling overhead that is unnecessary for devices and applications that are primarily static or nomadic.

2. There is overhead associated with additional headers that are added to every packet, and this consumes additional bandwidth on the backhaul links. The overhead can be significant when packets are small. Processing resources are unnecessarily consumed at the mobile gateway nodes and the base station to encapsulate and de-encapsulate packets in tunnels.

3. Because the tunneling encapsulation and de-encapsulation can occur only at special router nodes designed to handle the associated signaling messages, packets cannot always be routed using the shortest path. In particular, if the destination is closer to the base station to which the device is attached compared to the SGW/PGW nodes, then a specialized solution (SIPTO/LIPA) with a local SGW/PGW is required to avoid triangular routing. Additional cost is incurred to deploy this local SGW/PGW. And if the device has two applications, one that is local and one that is not, then the device needs to support two PDN connections and needs to attach to two PGWs. This essentially doubles the signaling involved.

3.3.4 EXPANDED FORM OF MULTI-RAT INTEGRATION AND MANAGEMENT

Wi-Fi/4G interworking is becoming increasingly prevalent, resulting in devices communicating over multiple radio access technologies. This interworking also has produced network topologies, such as "trusted non-3GPP access," that support seamless access selection, authentication, bearer plane inter-operation and in some cases seamless mobility. We expect this trend to continue with devices communicating over multiple air-interface types.

However, because the supporting network architectures for the different air-interface types were defined independently by different standards bodies, today there is little commonality in network functions and procedures. This fragmentation leads to a coarse level of interworking based on redirecting the mobile to an alternative access technology, where it exercises a different set of network procedures. For example, signaling on different access links is largely separate resulting in duplication of mobility, authentication and policy signaling for each access technology. Moreover, the selection and routing of traffic over the different technologies is left to the device, with some policy-based guidance functions such as the ANDSF. However, this limits the potential to steer traffic across different technologies based on dynamic criteria such as network status. For example, the network cannot alter whether a users' video should be sent over Wi-Fi versus over cellular as a function of network conditions and steer traffic accordingly.

3.3.5 ENHANCED EFFICIENCY FOR SHORT-BURST OR SMALL-DATA COMMUNICATION

Smartphone applications and many M2M devices frequently exchange short bursts of data with their network-side application. When there are no other communication needs, the devices have only a small amount of data to send but nevertheless have to go through a full signaling procedure to transmit the data. This wastes battery life, spectrum and network capacity.

To handle this type of transaction more efficiently, the network needs to support a truly connectionless mode of operation, where devices can simply wake up and send a short burst of data. Upon reception of the short burst, device and application-related state information can be retrieved from a controller function and resources to handle the packet allocated accordingly. Some attempts have been made to address this in 4G through some tweaks. 5G will offer an opportunity to include the requirement upfront in the design, thereby leading to superior solutions.

3.3.6 EXPANDING CONTEXT INFORMATION KNOWN TO THE NETWORK

Today 3G/4G access and the mobile core network have limited knowledge of the device and even less knowledge of applications that are requesting access. Static device and subscription information is available in the 3GPP Home Subscriber System (HSS) data base. This typically includes information about device type and subscribed services (e.g., SMS, voice). Standardized use of this information has been largely limited to determining the authorized PDN connections for the UE and authorized QoS. The network knows even less about the applications being used on devices. Almost all smartphone applications are provided “over the top” by third parties. Unless there are add-on deep packet inspection or analytic tools, the network has no visibility into these applications or their needs.

This contrasts sharply with knowledge obtained by applications and devices that produce big-data-based subscriber analytics, which enable targeted advertising and context-relevant subscriber offers. 5G needs a richer and more flexible method for the network to obtain and utilize information relevant to deciding how network resources should be allocated in the context of operator policy.

4. REGULATORY CONSIDERATIONS

The regulatory environment shouldn’t be overlooked as 5G technologies are being developed. It is anticipated that existing regulations will be applicable to the 5G environment. The following challenges need to be addressed to support regulatory requirements in 5G:

- **Location Accuracy** – Location accuracy requirements for emergency calls (e.g., 911) continue to evolve, with tighter and stricter regulations being imposed by the regulators. The 5G network must comply with the existing location accuracy requirements, as well as the anticipated stricter location accuracy requirements to be imposed within the next few years. The stricter location accuracy requirements could include both indoor and outdoor environments and an altitude component, and they could be applicable to dense urban, suburban, rural and remote environments. Specifically for indoor locations, where GPS cannot fulfill even current requirements, access subsystem in 5G networks will have to assist in achieving the needed accuracy requirement. In addition to location accuracy requirements for emergency calls, location accuracy may also be applicable to the public safety first responders, such as firefighters inside a burning structure.

- **Lawful Intercept** – It is anticipated that existing requirements for lawful intercept capabilities will continue and may be expanded as communication options and choices evolve. The existing lawful intercept architecture is based upon the principles that communication paths traverse centralized network elements, which can be monitored for lawful intercept purposes. However, the 5G network has the potential for communications paths that do not traverse centralized network elements (e.g., direct device-to-device communications, mesh network communications). So one technical challenge is to develop a 5G architecture that enables communications without transversal of centralized network elements while complying with lawful intercept regulations.

- **Tower Sharing** – 5G networks will have to support multiple radio technologies. Today's cell sites typically have one set of antennas – increasingly MIMO – for each RAT. This design becomes an issue at sites shared by multiple operators and for tower companies. As this may lead to the inability of deploying 5G services, with ripple effects on other licensed services, the 5G architecture should support solutions to minimize the number of antennas in shared multi-RAT environments. Additionally, regulators should be encouraged to limit the scope of tower sharing because this undermines operators' ability to innovate and shield those innovations from rivals.

- **Flexible Spectrum Use** - Even when fiber is available near a cell site, wireless backhaul allows faster deployment until the fiber can be installed. In many cases, both rural and urban, wireless backhaul often becomes the only technologically or economically feasible alternative. On the other hand, it is expected that 5G will use spectrum above 6 GHz. Therefore, ideally any spectrum used for 5G access should be flexible enough to also be used as backhaul. Spectrum licenses should be flexible enough to allow operators to meet the rollout demand while being capable of using 5G spectrum for backhaul when appropriate.

- **Mandated Digital Roaming** - There are two different types of domestic roaming. In the first, customers can roam to get service when they're outside their operator's coverage area. In the second, customers can roam to get service when they're inside their operator's coverage area but their operator has weak or no signals in that particular spot, what's known as a "black hole." The second type is much more demanding on both networks' resources. It is important to note that some countries mandate domestic roaming on digital technologies. Therefore, 5G networks should be capable of coping with such demand, while regulators should be aware about the more stringent requirement imposed on the latter scenario.

- **Critical Infrastructure** – Most telecom networks are classified as critical infrastructure because mobile broadband services are becoming an essential part of daily life. That responsibility will only increase while supporting IoT services. In order to fulfill this mandate, 5G networks must be robust, reliable, resilient and secure, as specified in section 3.1.3. Therefore, self-healing functionality, such as domestic roaming and network sharing, should be considered. Specifically in the case of network sharing, multiple carriers support is an essential requirement, one that's not really addressed today. This would ensure coverage and backhaul backup, which are often the weakest link of a wireless network. Those improvements should also be recognized and supported by adequate actions from regulators.

- **Emergency Telecommunication Service (ETS)** – This service gives government users priority access to the next available channel in crisis/disaster situation, when networks often get congested. For example, some countries (e.g., the U.S., Canada) have set up Wireless Priority Service (WPS) in partnership with mobile operators. Therefore, 5G networks should be capable of supporting these essential services. In addition, roaming with public safety mobile broadband networks, such as FirstNet in the U.S., may be required.

- **Public Warning System (PWS)** – 3GPP TS 22.268 already provides requirements for implementation of emergency alert systems around the world. Examples include Earthquake and Tsunami Warning System (ETWS), Commercial Mobile Alert System (CMAS) in the U.S., EU-ALERT in Europe and Korean Public Alert System (KPAS). It is important that such functionality be maintained while deploying 5G networks. It will also be important for 5G networks to be capable to fulfill another important PWS aspect not fully covered currently, support for multiple languages.

- **Accessibility** – Mobile broadband services are part of daily life, so 5G services must be accessible to people with disabilities, as is the case with 3G and 4G.

- **Use of SIM, E164 and TAC** – From a regulatory perspective, as 5G networks are going to all-IP, it may be appropriate to consider alternatives to the use of SIM, E164 (international public telecommunication numbering plan) and Type Allocation Code (TAC), which uniquely identifies each mobile device's model number and version. For example, is IMEI still the way to capture the right device for legal interception, or should a new approach be taken? TAC (a subset of the IMEI), was developed to uniquely identify a wireless device. However, it is now found that it may not be appropriate anymore for an operator to identify a specific version of a model number. In the case if SIMs, should it be further improved or a new approach taken, although use of IP certificate may not be an improvement?

5. POTENTIAL TECHNOLOGIES FOR 5G

This section describes potential technologies for 5G to address the market drivers and use cases described in Section 2, the requirements for 5G identified in Section 3 and the regulatory considerations defined in Section 4. Specifically, this section will discuss the following potential technologies:

- Massive MIMO
- RAN Transmission at Centimeter and Millimeter Waves
- New Waveforms
- Shared Spectrum Access
- Advanced Inter-node Coordination
- Simultaneous Transmission Reception
- Multi-RAT Integration and Management
- Device-to-Device Communications
- Efficient Small Data Transmission
- Wireless Backhaul/Access Integration
- Flexible Networks
- Flexible Mobility
- Context Aware Networking
- Information Centric Networking (ICN)
- Moving Networks

5.1 MASSIVE MIMO

MIMO employs multiple antennas at the transmitter and receiver, and is a well-known technique to increase the spectral efficiency of a wireless link. When devices have only a few antennas, multiple base station antennas can be used to simultaneously serve multiple users using the same time frequency resource. This requires knowledge of the channel between the base station antennas and the receiver antennas so that appropriate pre-coding can be employed to eliminate interference between signals transmitted to different users. Single-user MIMO and multi-user MIMO are both part of the 4G standards.

Massive MIMO extends the multi-user MIMO concept by dramatically increasing the number of antennas employed at the base station to be significantly larger than the number of users being served simultaneously in the same time-frequency block. With hundreds of antennas serving tens of users simultaneously, spectral efficiency can increase 5x to 10x, while users on a cell's fringes can maintain high throughput [4]. Furthermore, the pre-coding required for each user's signal reduces to simple conjugate beamforming. The major challenge of acquiring channel information at the transmitter is solved by employing time-division duplexing, where the same spectrum is used in both the DL and the UL, ensuring that the DL channel is nearly the same as the UL channel. Users transmit orthogonal pilots on the UL from which the UL channel is estimated and then used for the conjugate beamforming on the DL [5]. UL pilot transmissions may not necessarily be orthogonal across cells because only limited resources can be devoted to pilot transmissions. This results in what is called “pilot contamination,” producing channel estimation errors. Pilot contamination is mitigated using pilot reuse, where the pilot sequences are reused only in cells outside of the immediate neighborhood of the cell where it is used.

Advances in radio and antenna technology are required to cost-effectively deploy a large number of antennas at the base station. While channel propagation is reciprocal, the receive and transmit paths may not be. Thus antenna calibration may be required to account for any substantial difference that may arise between DL and UL. It should be noted that Massive MIMO does not significantly increase the peak rate to a single user as it inherently needs multiple users to be served simultaneously to achieve the high spectral efficiency.

5.2 RAN TRANSMISSION AT CENTIMETER AND MILLIMETER WAVES

Mobile-communication networks have, until now, almost exclusively operated on frequencies below 3 GHz. However, extension into higher frequency bands, including frequencies above 10 GHz, is being considered for 5G. Frequencies of 3 GHz to 30 GHz are in the centimeter wavelength band, and frequencies of 30 GHz to 300 GHz are in the millimeter wavelength band.

The main benefit of frequencies above 10 GHz is the potential availability of large amount of spectrum and, perhaps even more important, large continuous spectrum chunks. The latter is needed to enable the very wide transmission bandwidths, such as several hundred MHz. Such transmission bandwidths are needed for efficient support of multi-Gbps data rates.

The main drawback of using higher frequencies is higher path loss. This can be partly compensated for by the use of more advanced antenna configurations, making use of the reduced size of the basic antenna elements at higher frequencies. Indeed, by keeping the size of the overall antenna configuration the same on both the transmitter and the receiver side, in combination with beam-forming, the overall path loss may actually be reduced as the frequency increases. This is one of the reasons for the current use of higher frequencies, in combination with highly directional antennas, for wireless backhaul.

However, this is true only for line-of-sight conditions. In non-line-of-sight conditions, which is the typical propagation situation in mobile communication, there are additional path-loss-degrading factors such as:

- Reduced diffraction, leading to higher path loss due to shadowed locations.
- Higher attenuation when propagating through walls, for example, leading to higher path loss in indoor locations covered by outdoor base stations.

Despite this, recent studies have shown that higher frequencies, up to at least 30 GHz, can be used for wireless access in non-line-of-sight conditions, assuming relatively short-range (100-200 meters) links. Even higher frequencies are being considered for ultra-dense deployments with even shorter access-node inter-site distance.

It seems likely that the use of higher frequencies will be one important component of 5G wireless access. However, higher frequencies can serve only as a complement to lower frequencies, providing high capacity and high data rates in dense urban environments. Lower frequencies should remain the backbone providing full wide-area coverage.

World Radio Conference (WRC) 2015 will focus on new spectrum mobile communication below roughly 6.5 GHz. Identification and assignment of spectrum above 10 GHz for mobile communication is expected to be on the agenda for WRC in 2018/2019. Thus, higher frequencies above 10 GHz may be available at the time of 5G initial deployments, which are expected around 2020.

5.3 NEW WAVEFORMS

The new waveforms in 5G include advanced multi-carrier transmission and non-orthogonal transmission.

5.3.1 ADVANCED MULTI-CARRIER TRANSMISSION

LTE radio access is based on OFDM transmission in both DL and UL: conventional OFDM for the DL and DFT-precoded OFDM for UL. OFDM transmission, which is a kind of multi-carrier transmission scheme, is also a candidate for 5G radio access. However, several other/modified multi-carrier transmission schemes are also under consideration for 5G radio access. These include (see e.g., [6] for more details):

- Filter-Bank Multi-Carrier (FBMC) transmission
- Universal Filtered Multi-Carrier (UFMC) transmission
- Generalized Frequency-Division Multiplexing (GFDM)

Common for these transmission schemes is that they, at least in principle, can provide a more confined spectrum compared to conventional OFDM. This is relevant for spectrum-sharing scenarios. It should be noted that the confined spectrum is a property of the fundamental waveform and that transmitter nonlinearities may cause additional spectrum spreading that may reduce these waveforms' benefits.

The more confined spectrum is also assumed to make the above transmission schemes less reliant on-time synchronization to retain orthogonality between different transmissions. This may be valuable especially for UL transmission with requirements on very low access latency as the need for time-consuming synchronization procedures may be relaxed or even avoided.

5.3.2 NON-ORTHOGONAL TRANSMISSION

4G radio access is based on orthogonal transmission for both DL and UL. Orthogonal transmission avoids interference and leads to high system capacity. However, for rapid access of small payloads, the procedure to assign orthogonal resources to different users may require extensive signaling and lead to additional latency. Thus, support for non-orthogonal access, as a complement to orthogonal access, is being considered for 5G. Examples include Non-Orthogonal Multiple Access (NOMA) [7] and Sparse-Code Multiple Access (SCMA) [8].

5.4 SHARED SPECTRUM ACCESS

The Federal Communications Commission (FCC) has been aggressive in its efforts to make new spectrum available for mobile communications. The FCC's 2010 National Broadband Plan concluded that 1.2 to 1.7 GHz of new spectrum is required to sustain capacity expansion required to meet anticipated growth in wireless data traffic. The FCC is pursuing shared spectrum access as a way to free up spectrum. Secondary users would be allowed to use the spectrum when the primary or incumbent is not using the spectrum in a given geography at a given time. The FCC has published a Notice of Proposed Rule Making (NPRM) for three-tier spectrum access in the 3.5 GHz band: incumbent, protected and general authorized access. A spectrum access server (SAS) manages the allocation of spectrum between these tiers.

Some new technologies are needed for mobile networks to use the shared spectrum:

- The RAN must be capable of interfacing with the SAS to request and receive spectrum allocations, and to provide the SAS with spectrum-sensing information from the base stations.
- The base stations must be spectrum agile and capable of spectrum sensing. Spectrum sensing means monitoring the frequency channel for use by the primary. With spectrum agility, users served in one channel are migrated to a new channel and then, to protect the incumbent, stop transmitting in the first channel when the SAS says to. This can be achieved with wideband adaptive radios. The migration should be designed in such a way that the users do not see an interruption when the switch to the new channel happens. That seamlessness can be accomplished through channel aggregation schemes so that when one of the channels is temporarily blocked, communication can continue on the remaining channels.

5.5 ADVANCED INTER-NODE COORDINATION

Interference between radio access nodes is the limiting factor of current wireless networks. In addition, the extreme densification of 5G radio nodes necessitates specific interference-avoidance solutions, both from the network and the device sides. One such solution is based on exchanging information between schedulers at the network side in order to avoid interference. This technique leverages current inter-cell interference coordination schemes (LTE Rel. 8) as well as Coordinated Multi-Point (CoMP) (LTE Rel. 11 and 12). However these solutions are currently not so effective because of a number of drawbacks. One drawback is the upgrades in backhaul transport that are required to support the necessary exchange of information with very small delays. Another significant drawback of CoMP-based solutions is inter-cluster interference: Cells are grouped into clusters for resources coordination, but the optimal cluster configuration that minimizes inter-cluster interference at a reasonable coordination complexity is still an open question.

Research in 5G is being conducted to overcome the aforementioned issues and facilitate inter-node coordination even with legacy backhaul networks. One such research area, originated in LTE Rel-12 and anticipated for Rel-13, deals with pre-compensating any foreseeable delay and jitter impairments of the backhaul. Other approaches try to relax inter-node coordination burden by allowing some degree of interference between nodes that must be handled at the receiver.

Centralization of radio processing functions reduce the signaling burden and would therefore be a driver for efficient inter-node coordination. However, the required capacity for the front haul network (transport network between the central baseband unit and the remote radio heads) would be one of the challenges when aggregating large numbers of base stations.

5.6 SIMULTANEOUS TRANSMISSION RECEPTION

Today's wireless systems dedicate spectral or temporal resources to UL and DL channels. Simultaneous transmission reception would enable sharing of the available resources between both directions of communication. Clearly, this would be a more efficient use of the available spectrum and in theory could double the current link capacity. This may seem a small gain compared to the 5G capacity needs and compared to what can be achieved through techniques like massive MIMO. However, the value of simultaneous transmission reception for 5G may not necessarily be in its capacity gains but in possible improvements in signaling and control layers. Removing the fundamental assumption of having separate

UL and DL could allow 5G systems to be designed with a new approach and possibly facilitate the achievement of 5G goals in ways that may not be obvious now [9].

It is generally not possible to only use the same channel for simultaneous transmission of UL and DL signals. The key challenge is the large power differential between the strong self-interference due to the device's own transmissions and the weak signal of interest coming from the distant transmitter. However, this has not deterred researchers from pursuing the goal of simultaneous transmission reception. Experimental demonstrations of simultaneous transmission reception for wireless communication systems have been reported since 1998 [10-17]. The work so far has been successful in reducing this interference by up to 85 dB [18], which has generally been achieved through a combination of analog cancellation, hardware cancellation and digital cancellation techniques. The interference-reduction levels achieved so far are sufficient for Wi-Fi-type systems over very short distances, where the received signal is strong enough to make the difference between the device's transmit power and received signal. These levels of interference reduction are clearly not sufficient for current cellular systems, which operate with much higher transmit powers and in greater path-loss environments. However, the necessary interference reduction levels for a 5G system operating in the millimeter wave range may be more realizable due to lower transmit powers.

The current experimental success in achieving simultaneous transmission reception in Wi-Fi-like scenarios is encouraging. However, it is by no means close to practical implementation in a cellular environment. Further refinement of the interference cancellation techniques and possibly completely new approaches to interference cancellation are necessary to make simultaneous transmission reception a reality in practical systems. The interest in simultaneous transmission reception as a possible feature of 5G systems could drive further work in this area that would bring the current experimental achievements closer to practical implementation.

5.7 MULTI-RAT INTEGRATION AND MANAGEMENT

The ever-increasing number of RATs to be supported in a given deployment makes it crucial to consider multi-RAT integration and management issues. The objective is to facilitate uniform multi-RAT management and convergence among disparate technologies, both 3GPP and non-3GPP, such as Wi-Fi. Operation efficiency and user experience would be dramatically improved by automatically steering devices to the most suitable RAT in a seamless way. While multi-RAT management has been an important aspect in previous mobile generations, 5G's user-driven requirements foresee a seamless user experience when moving across networks. That makes multi-RAT integration more critical, particularly for services such as ultra-high-definition video or tactile Internet.

Given the likelihood of having multiple, heterogeneous wireless access points available in ultra-dense scenarios (e.g., 5G, LTE, 3G and Wi-Fi), some kind of decoupling between the user and control planes should be provided in order to separate the user payload from the necessary signaling. Multi-RAT integration may also consider simultaneous connection to multiple RATs in an opportunistic manner.

The expected impact on the network of such schemes is the introduction of a logical entity that coordinates resources among multiple RATs. To this end, the introduction of virtualization techniques may facilitate this point by enabling the instantiation of network functions upon demand, without having to change the network topology and/or architecture. Software-defined network functions can cope with different RATs by instantiating the necessary network functions upon demand, without the need to physically deploy additional network nodes for multi-RAT management.

5.8 DEVICE-TO-DEVICE COMMUNICATION

The introduction of direct device-to-device (D2D) communication for LTE began during 3GPP Release 12. At this stage, the considered D2D functionality was relatively limited, mainly focusing on D2D communication for public safety communication and D2D proximity detection for more general commercial applications.

However, as part of overall 5G discussions where technology components such as highly integrated backhaul/access (see Section 5.10) and more general multi-hop communication is being considered, D2D communication should definitely also be included as a possible technology component. One should then consider direct D2D communication as a more general tool that is a well-integrated part of the overall wireless-access solution. Besides what's depicted in Figure 1, this should include:

- The use of direct peer-to-peer D2D communication as an overall more efficient mode of transmission when nearby devices have end-user data to convey between each other.
- The use of direct D2D communication as a means to extend coverage beyond the reach of the conventional infra-structure (device-based relaying).
- Cooperative devices where high-speed inter-device communication provides means for "joint" transmission and/or reception between multiple devices, thus opening up for more efficient communication with the network-device communication. Note that this can be seen as a kind of coordinated transmission/reception ("CoMP") but on the device, rather than the network side.

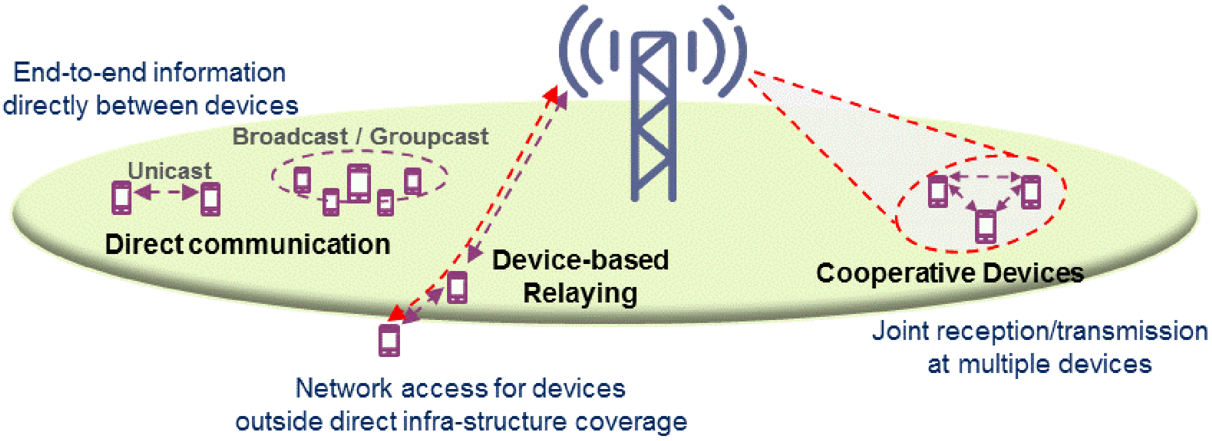

Figure 1. Device-to-Device Wireless Access Solution.

The key thing should be to see D2D communication as well-integrated part of the overall wireless access solution that is being taken into account already from the start of the 5G definition, rather than being a later-introduced "add-on."

D2D communication can be carried out in licensed spectrum under network control. D2D communication can also be carried out in, for example, unlicensed spectrum. Also in that case, an overlaid network, operating in the same or other spectrum, can be used to control/assist the D2D communication for enhanced efficiency and performance. However, D2D communication should also be possible in scenarios where there is no network coverage available, in which case the D2D link has to be possible to establish without network control/assistance.

5.9 EFFICIENT SMALL DATA TRANSMISSION

Small data bursts are most efficiently handled by connectionless access whereby a device wakes up and sends a short user plane data burst on a common 5G carrier using contention based access, eliminating dedicated radio resource control and higher layer signaling between the mobile and the network. Connectionless access reduces device power consumption and saves network resources when the size of the data burst is small compared to the overhead needed to set up bearers and establish security and device context for scheduled access. For 5G, the radio system must support connectionless contention based access multiplexed with scheduled access on a common carrier, with flexible allocation of resources that may vary along with traffic demand.

Smartphones and other devices capable of both scheduled and connectionless access may use connectionless access to send short data bursts and scheduled access for longer transactions. Example applications for connectionless access include the polling of servers (e.g., for email), sending instant messages (IM) and sending TCP keep-alives by applications that require long-term connectivity through NATs. In contrast, MTC sensor devices may only support connectionless user plane access, saving complexity and expense in the device in addition to further extend battery life. The savings in network resource usage by these connectionless-only devices may enable a lower cost tier of service, expanding the range of applications for which mobile wireless access makes economic sense.

To support connectionless access, a device must attach or otherwise establish a context in the network that allows the receiving base station to validate the packet as belonging to a legitimate, authenticated device. A security context must also be established whereby the network can ensure user plane integrity and privacy for the transmitted burst. Public/private key infrastructure, or an extension of the 3GPP security architecture, with the context information retrieval keyed from the user plane packet header may be used for this purpose.

Packet processing for connectionless access may be modeled on OpenFlow Software Defined Networking (SDN) procedures that handle the processing of unidentified packets arriving at a switch. For example, the following steps may be taken after the device attaches to the network, is authenticated, and establishes with a controller a security context and a device context indicating support for connectionless access.

1. The device (at a future time) selects a base station and sends a user plane data burst.
2. The base station sends the packet or packet header to the controller for validation, much as an OpenFlow switch would send a packet for which it had no flow table entry to an OpenFlow controller.
3. The controller authorizes access and sends security context information to the base station along with forwarding instructions, much as an OpenFlow controller would provide a flow table entry to a switch.
4. The base station deciphers the packet contents and forwards the packet according to the received instructions.

After a data burst has been sent, the device may wait for a response from the network and transition to a connected state as warranted by further transmissions.

5.10 WIRELESS BACKHAUL/ACCESS INTEGRATION

In many places around the world, wireless backhaul constitutes a major part of cellular’ s overall backhaul portfolio, especially in less populated areas, where other backhaul solutions are not economically and/or technically feasible. Wireless backhaul is typically based on a proprietary radio-link technology operating under line-of-sight propagation conditions. In the early days of mobile communication, wireless backhaul often used frequencies in the 2 GHz range and even lower. Today backhaul is being concentrated in spectrum above 6 GHz, including the millimeter band (above 30 GHz). A few years ago, wireless backhaul started to be deployed in the 70-80 GHz spectrum (E-band), where the backhaul can transport up to 5 Gbps. With the increased interest in small cells, including those deployed indoors, there is an increased interest in wireless backhaul solutions that can operate under non-line-of-sight conditions. There is still an assumption of a backhaul-specific radio-access technology operating in spectrum separate from the spectrum used for the access (BS-UE) link. The backhaul link may, in these cases, operate in millimeter bands. Lower frequencies also are being discussed as possibilities for small cell backhaul.

There is no fundamental reason why the basic radio-access technology needs to be different between the (wireless) backhaul link and the access link. This is especially the case taking into account that:

- The use of higher frequencies is being considered also for the access link (see Section 5.2).
- Lower frequency bands are being considered for wireless backhaul, as discussed above.
- Flexibility to use spectrum for both mobile and backhaul.

The radio conditions for a backhaul link to a small “under-roof-top” base station are more similar to a BS-UE link than to a conventional macro-cell backhaul link.

In the case of *integrated backhaul/access*, one sees the access and backhaul link as just two links in an overall unified wireless access solution:

- The same (set of) radio-access technology(ies) can be used on the access (BS-UE) link and the backhaul link.
- The backhaul and access link relies on the same spectrum pool. Note that this does not necessarily mean that the same carrier frequency is used on the two links at the same time and in the same location.
- Resource management, QoS optimization and other tasks are carried out jointly for the backhaul and access links.

The aim of access/backhaul integration is to achieve more efficient use of technology and spectrum and to improve the overall performance/quality of the overall end-to-end link. It should be noted that, from a radio-link point-of-view, integrated backhaul/access can be seen as a kind of multi-hop communication.

5.11 FLEXIBLE NETWORKS

Elastic computing and storage in data centers has ushered in a new need for highly dynamic networking that is being addressed by SDN. The benefits of SDN in data center networking, both within and between data centers, have been firmly established. Extending beyond the data center, creation of on demand virtual private networks with dynamically configurable capacity in the optical links between enterprise premises and the data center is also being explored.

It should be noted that the narrow view of SDN includes only OpenFlow switches and OpenFlow controllers that connect to the OpenFlow switches to centrally manage packet forwarding, with management support from an OpenFlow Config. However, the broader view of SDN is one where the emphasis is on network programmability through open northbound interfaces, a configurable policy framework, resource discovery and optimization and an SDN control framework that is separate from the forwarding plane.[3] The SDN controller may exercise control over the network elements not only through OpenFlow but also through other interfaces such network management interfaces. We take this broader view of SDN when applying it to the domain of wireless networks.

The application of SDN in 5G may extend beyond control of transport resources to include the wireless network's policy framework. Examples include how functions such as QoS management and application of forwarding plane functions via service chaining are applied in an SDN-enabled packet core. More broadly, an SDN framework can also be applied to wireless control functions where, for example, mobility management, security, charging and optimization become applications subtended by an SDN controller.

NFV virtualizes network applications and network functions such as those provided today by dedicated core network components, media servers and management functions. Virtualization enables these functions to run on hypervisors using shared, off-the-shelf data center computing infrastructure. NFV is already employed in 4G implementations and will be commonplace well before 5G. Going further, NFV can be exploited as technology fundamental to the design of 5G, where we expect the majority of the wireless network functions in the RAN and the core to be virtualized, thus increasing flexibility to meet varied demands on control and data processing from a large variety of applications. Interconnection amongst these functions and programming of the forwarding plane to support mobility is most flexibly achieved using SDN techniques.

Closely related and complementary to NFV is orchestration technology that enables automated deployment and management of network functions and network services. The orchestration framework is key to realizing the cost savings of moving to virtualized implementations. Latency and networking requirements may necessitate placement of virtualized functions in close proximity to each other or at local data centers close to the subscriber. In some cases, network functions key to achieving very low latency may be better implemented without NFV. The MANO work group within ETSI NFV forum has defined an orchestration framework that can be applied to 4G and more deeply integrated into the 5G network architecture.

NFV and the major principles of SDN, such as the separation of control plane and data plane, network abstraction and programmability of the network by external applications all should be broadly applicable to wireless networks. This potentially enables new services, reducing the time and effort needed to implement those services and reducing network costs.

The ability to rapidly establish network resources based on application demands and operator policy has many benefits. Rather than manually configuring fixed resources on dedicated hardware, which may not match dynamically changing application demand, operators can shift bearer plane assets as needed and allocate to users only those resources needed to satisfy device and application requirements. This enables network operators to better monetize their network assets and quickly and more easily roll out new end-user services. In addition, CAPEX savings comes from more efficient use of network assets and the use of simpler, lower cost bearer plane networking infrastructure. OPEX savings is obtained from

---

[3] http://www.opendaylight.org/project/faq#1

auto-configuration and simpler network configurations without wireless-specialized forwarding plane elements.

As distributed network functions running on dedicated hardware are phased out in favor of NFV, we expect SDN for wireless networking to facilitate meeting data center networking requirements for the virtualized wireless RAN, core network and applications. In the data center, SDN is needed to provide dynamic configuration and scalability of network resources as wireless network functions are instantiated, VMs are reassigned, traffic grows and new radio infrastructure is deployed outside of the data center. For wireless access, SDN programmability can be used to simplify access selection and packet routing, with forwarding rules directing packet flows both in the mobile and base stations. This real-time programmability realizes the goals of context-aware networking, so intelligence information is actually applied in the selection of packet routes, particularly when multiple air interface options are available.

5.12 FLEXIBLE MOBILITY

Flexible mobility for 5G embraces a selection of options, which may be dynamically assigned to a device or application according to the device and application context, or statically configured for specialized devices and applications. Flexible mobility consists of two components: one for managing mobility of active devices and a second for tracking and reaching devices that support a power-saving idle mode. Assigned mobility may range at one extreme, beginning with no “active mode” mobility, with no support for idle mode as is typical of Wi-Fi access today. The other extreme is full support for active and idle mode mobility similar to mobile 3G/4G. Gradations of flexible mobility bridge the gap between these extremes, allowing for independent assignment of idle-mode mobility on a per-device basis, and active mode mobility on a per-application basis as shown in the examples in Table 1.

Table 1. Examples of Flexible Assignable Mobility for 5G.

Idle Mode Mobility	**Active Mode Mobility**	**Application**	**Current Tech**
Yes	Yes	Smartphones, Tables for mobile communication	3G/4G
None	None	Always-on, battery in-sensitive	Wi-Fi
None with sleep/ coma mode in device	None	When the device need not be reachable from the network after idle transition and only the device initiates transactions (e.g., MTC long battery life)	--
Yes	None	For nomadic, long-battery life devices that must be reachable from the network	--
None	Yes	Always-on, battery insensitive with seamless continuity when active	--
None with sleep/ coma mode in device	Yes	When the device need not be reachable from the network after idle transition with seamless continuity when active	--

Here idle mobility involves the tracking of the device location while in “sleep mode.” Active mobility refers to establishment of IP anchors, where tunnel endpoints are updated when the UE moves between base stations so that seamless IP session continuity may be maintained.

Offering a range of mobility options not available today in 4G enables a better match between the needs of the device and application and network resources. This has advantages for both the network and user/device. Specifically:

1. “**No idle mode mobility**” means that context and state information for tracking the device need not be stored in the network, saving resources when applied to a plethora of MTC, sensor-type devices. There are two cases for no idle mode mobility: The device may not support idle mode and be always-on, similar to Wi-Fi, or it may support a sleep/coma mode whereby the device de-allocates Tx and Rx resources, hibernates until it wishes to initiated a transaction and is not reachable from the network until that point. The latter case is particularly useful for a sensor-like device where battery life is of paramount importance and network-initiated communication is not required.

2. “**No active mode mobility**” means that the network need not establish and maintain user plane tunnels and store related state information. IP addresses are allocated local to the base station, allowing for more efficient routing of locally available content. The user/device benefits from lower latency due to more direct routing when compared to active mobility with a centralized anchor. This mode is most appropriate for stationary and nomadic devices.

With active mode mobility, the network may flexibly assign an IP anchor to one of several network elements, such as a base station router, aggregation router, edge router or, in the case of “no active mobility,” a local prefix is assigned with no anchor. A device such as a smartphone may have multiple active mode mobility contexts for different applications with, for example, a central anchor point assigned for a mobile communication application, and no anchor assigned for HTTP Internet access, where IP session continuity is not required and content can be provided locally.

Flexible mobility is enabled by two critical technologies. The first is context awareness. The network needs context information about the application’s needs and the device’s capabilities to determine the appropriate level of active and idle mode mobility to assign. The second technology is SDN control of the transport path. Flexible active mode mobility is possible only if mobile anchors can be dynamically assigned at a central point, close to the device, in between or not at all at the time an application is invoked. This requires a context-aware controller that can program transport elements to establish tunnels where needed and forward traffic accordingly.

5.13 CONTEXT-AWARE NETWORKING

The network cannot provide resources tailored to serve a wide range of devices and applications without context information that goes significantly beyond that available in 4G. Context awareness allows the network to adapt to the needs of applications within the framework of network constraints and operator policy. This is preferable to the alternative, where applications adapt to the constraints of a one-size-fits-all set of service characteristics on a default bearer as is typical in 4G.

In addition, the ability of both the network and device to use context awareness (e.g., location, historical usage pattern, subscriber preferences) can help further enhance the user experience. This ability also enables the concept of the Internet coming to the user with the most relevant and timely information rather than the user having to go to the Internet to retrieve information and then to filter out the irrelevant pieces of information. Context awareness includes awareness of the following:

1. **Network Analytics**, including alternative RATs, network layers (macro cell, mmWave, small cell, Wi-Fi) and corresponding congestion levels, capabilities and performance characteristics.

2. **Subscriber Analytics**, including subscription attributes, wireless activity level, loyalty management status, experience analytics, historical subscriber activity, location history, current location, subscriber contacts, location context (e.g., work, home, mall) and application usage.

3. **Device Attributes and Capabilities**, including information on single function vs. multi-function devices, device support for specialized applications, MTC vs. subscriber devices and radio and network optimization capabilities.

4. **Application Requirements**, including QoS requirements (e.g., delay, throughput, latency), connection reliability, access price, power consumption and security level.

5. **Subscriber Preferences**, including preferred access options, power savings vs. performance and access cost.

6. **Operator Policies and Subscription Context**, including allowed services, service attributes and QoS.

Context information may be gathered from the device, network monitors, network elements, network data bases and analytics platforms. It is processed by the network when a device attaches or an application is invoked and results in a determination of service attributes that govern how the device and application will be treated by the network. The service attributes for access may for example include cost, reliability, power consumption, security level, QoS and mobility. The service attributes for access may be mapped to configurable 5G features, which are then assigned by the network.

For example, context information may determine that low-cost access, with no support for active mobility and long battery life, is best for providing service to a nomadic sensor device that attaches to the network. The network as a result configures connectionless access with low priority, simple IP networking with no tunneling and an idle-mode wake-up period of one day.

5.14 INFORMATION CENTRIC NETWORKING (ICN)

As discussed in Section 3.3.2, 5G should be based on new network architectures and protocols designed specifically with support for mobility, security and content caching as fundamental design criteria. Information Centric Networking (ICN) – for example, as realized in the Named-Data Networking (NDN)[4] and Content-Centric Networking (CCN)[5] programs – is emerging as a leading architecture that can meet such design criteria. ICN approaches are focused on the support of future Internet evolution and, in particular, support of new communication models that focus on the distribution of information rather than the communication of data packets between endpoints.

Core to this work is the rethinking of the key design principles of the Internet to incorporate the requirements of new application models such as scalable content distribution, mobility, security and trust as fundamental architectural design features. As such, key (networking) areas that are touched are summarized as follows:

- Naming – focus on what content is of interest rather than where it is.
- Routing – based on (hierarchical) names rather than addresses.
- Mobility – now an intrinsic capability of networking layer.
- Caching/Storage- information resides anywhere in the network at any time.

---

[4] http://named-data.net/project/archoverview/

[5] http://ccnx.org/what-is-ccn/

- Security – secure content rather than communication channels.

ICN (NDN/CCN) enhances the role of the networking layer replacing its task from simply providing a pair-wise packet-delivery channel between communicating hosts to delivering named-information to an endpoint that expresses an interest in it without explicit direction as to where that content is stored. By naming information at the network layer, ICN (NDN/CCN) facilitates the deployment of in-network caching (and storage in general), simplifies multicast and enables a security model where content itself is secured rather than the channel over which the content is conveyed. Expressions of interest, implemented via a request-response mechanism, support mobility and congestion control/avoidance mechanisms implicitly so the need for an overlay to support mobility may be avoided. The strategy mechanism provides a flexible mechanism to realize intelligent, context-aware control of content storage and delivery.

The basic concepts being explored in the NDN/CCN community have direct relevance to the design of the 5G core network and can result in significant simplification. It will serve the cellular community well to carefully study this work to understand how it can be leveraged in 5G and how it can be shaped to better serve longer-term cellular industry interests. The basic principles may serve to alter the evolution path of the current cellular core and enhance the service provider's role in the information delivery economic chain.

5.15 MOVING NETWORKS

Mobility support in 5G will likely extend to very high speeds, such as 350 km/h and beyond, and even aircraft communications, although this possibility still remains unclear. However, both the device and cell can be moving like in D2D, V2V or V2I scenarios. Ultimately, the concept of a cell becomes blurred in favor of a more general concept of connectivity, where the network follows the movement of the user rather than the opposite (as usually conceived in previous mobile generations).

The management of nomadic and moving cells presents a number of issues – such as activation/deactivation of cells, trajectory prediction and handover optimization – because users will rapidly traverse multiple cells in a very short time. Additionally, the Doppler shift caused by very high relative movement between transmitter and receiver can challenge the use of millimeter waves.

6. SPECTRUM

Traffic growth continues to increase at exponential rates. The old adage that "if you build it, they will come" does not hold true. Users are coming faster than the infrastructure can be built. The trend that started in the 1990s of merging IT and cellular systems continues. Data services are expected anywhere and at any time and at an ever-faster rate. Low latency is not a luxury, but a basic requirement.

The industry has estimated capacity resources to start becoming saturated around the 2018 timeframe. These estimates are based on existing technology and do not take into account incremental steps of air interface capacity improvements. When improvements such as beamforming, load balancing, channel efficiency and CA are taken into account, estimates show that perhaps it will be 2020 when new technologies will make their commercial debuts.

Besides just technology advances and system architecture evolution, it's clear that additional spectrum allocated for mobile broadband will be required to meet the projected demand[6]. The industry recognizes that new spectrum below 6 GHz will be difficult to obtain and that spectrum above 6 GHz may not have the desired propagation characteristics for wide-area coverage, although in many cases it will be suitable for high-density system deployments. The simple truth is that future networks must have the ability to utilize the entire range of spectrum ranging from below 1 GHz up to 6 GHz and then well into the millimeter wave ranges efficiently and seamlessly.

Mobile operators reacting to the demand of the marketplace, the advancement of enabling technologies and the convergence of communication, information and entertainment are offering an increasingly vast range of personal, business, public service, wide area data, location and MTC across a wide variety of environments.

In order to meet the challenges, both the application of advanced technology and the use of appropriate higher spectrum bands with larger channel bandwidth, are anticipated from a data rate perspective to provide support for the large factor increase projected for 5G system traffic. From a capacity viewpoint, the use of higher spectrum bands may also support new architecture models promoting increased effectiveness of the systems.

As noted previously, spectrum above 6 GHz (principally in the multi-GHz ranges) is already being investigated for its technical suitability for mobile broadband deployments to support 5G. There's also consideration of use of these bands at national levels and/or at the World Radio Conference level (targeting WRC-19).

One-size-fits-all air interfaces, which have been the typical generational solutions of the past 20-plus years, may no longer be the total solution. Although subsequent generations of wireless may indeed use new air interfaces, the industry fully expects that future technology will depend more on heterogeneous network deployments utilizing multiple air interfaces in order to meet throughput, coverage and capacity needs. Additionally, the concepts of spectrum sharing and unlicensed operations must be part of any 5G vision. Wi-Fi will continue to grow in importance.

Spectrum is already highly fragmented and is not likely to change in the future. Roaming, a problem once thought solved, is becoming a resurging issue with all of these new bands. The concept of a 50-band mobile is needed from a global ecosystem viewpoint. But in practical terms, it is not possible. Hence the continued need to attain globally harmonized frequency bands in future spectrum allocations is still an urgent and pressing quest.

Furthermore, the ability to deliver the best QoS and utilize the best radio band available consistent with the QoS requirement is a key feature of mobile equipment of the future. It may be required in future equipment (and hence, supported by system specifications) to operate in reasonable and harmonized ways not only on dedicated spectrum bands, but also to incorporate operation under licensed assisted access schemes, shared spectrum arrangements or other concepts.

---

[6] See Report ITU-R M.2290-0 (12/2013) "Future spectrum requirements estimate for terrestrial IMT"

7. CONCLUSIONS AND RECOMMENDATIONS

An end-to-end 5G system has to be architected to meet the expected demand in 2020 and beyond. Figure 2 illustrates a comprehensive view that must be considered in the initial planning process for 5G.

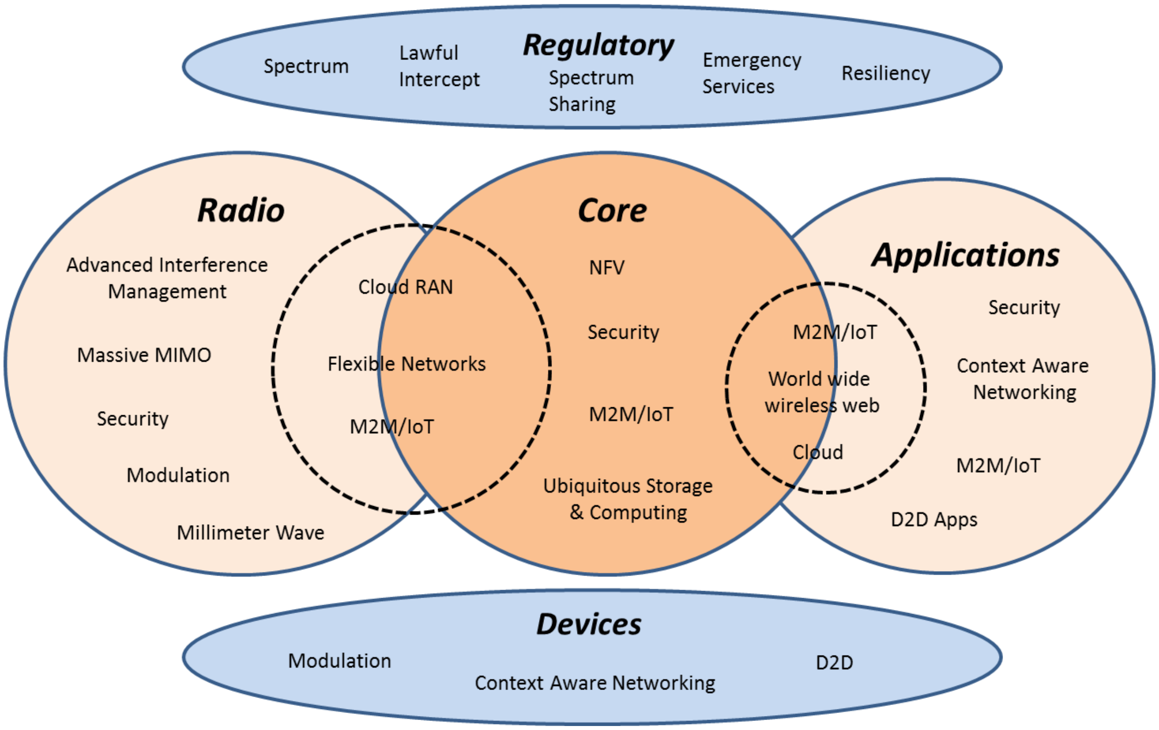

Figure 2. A Preliminary View of an End-to-End 5G Ecosystem.

The key 5G principles highlighted in this white paper are:

- As 5G is defined and requirements developed, it must include the entire 5G ecosystem (e.g., air interface, devices, transport, packet core).
- 5G development should provide global harmonization under a single framework and allow time for true advances of technology, feasibility studies, standardization and product development.
- It is critical that the countries of the Americas invest in 5G research.
- Avoid debate (at least initially) on what 5G is. 5G does not (yet) describe any particular specification in any official document published by any standardization body.
- 5G planning should consider all major technology advances on the road to 5G.
- Wherever feasible, features being discussed as 5G requirements should be implemented as LTE-Advanced extensions before the full 5G is available. This will also give time to recoup the investment in 4G.
- There are ongoing enhancements in LTE-Advanced that will continue through 2018. 5G is envisioned to have initial deployments around 2020. It must be recognized that significant breakthroughs in new radio transmission interfaces may be accompanied by a break in backward compatibility.

In conclusion, it is recommended that the market drivers, use cases, requirements, regulatory considerations and the technology elements described in Sections 2, 3, 4, 5 and 6 all be taken into consideration in the further development of the end-to-end 5G system.

APPENDIX A: ACRONYM LIST

3D	Three Dimensional
3G	Third Generation
3GPP	Third Generation Partnership Program
4D	Four Dimensional
4G	Fourth Generation
5G	Fifth Generation
ADAS	Advanced Driver Assistance Systems
ANDSF	Access Network Discovery and Selection Function
APN	Access Point Name
BS	Base Station
BYOD	Bring Your Own Device
CAPEX	Capital Expense
CAT	Computerized Tomography
CCN	Content-Centric Networking
CDN	Content Distribution Network
CMAS	Commercial Mobile Alert System
CoMP	Coordinated Multi-Point
D2D	Device to Device
DFT	Discrete Fourier Transform
DL	Downlink
DoS	Denial of Service
ETS	Emergency Telecommunication Service
ETWS	Earthquake and Tsunami Warning System
FBMC	Filter-Bank Multi Carrier
FCC	Federal Communications Commission
Gbps	Gigabits Per Second
GFDM	Generalized Frequency Division Multiplexing
GHz	Gigahertz
HAS	HTTP Adaptive Streaming
HSPA+	High-Speed Packet Access Plus
HSS	Home Subscriber System
HTTP	Hypertext Transfer Protocol
ICN	Information Centric Networking
IM	Instant Message
IMEI	International Mobile Equipment Identity
IMT	International Mobile Telecommunications
IoT	Internet of Things
IP	Internet Protocol
IT	Information Technology
Km/h	Kilometers per hour
KPAS	Korean Public Alert System
LIPO	Local IP Access
LTE	Long Term Evolution
M2M	Machine to Machine
MANO	Management and. Orchestration
MHz	Megahertz
MIMO	Multiple Input, Multiple Output
MMES	Multi-Media Emergency Services
mmw	Millimeter wave
mmWave	Millimeter wave
MNO	Mobile Network Operator
MRI	Magnetic Resonance Imaging
MTC	Machine Type Communication
NAT	Network Address Translation
NDN	Named-Data Networking

NFV	Network Function Virtualization
NOMA	Non-Orthogonal Multiple Access
NPRM	Notice of Proposed Rule Making
OFDM	Orthogonal Frequency Division Multiplexing
OPEX	Operating Expense
PDN	Packet Data Network
PGW	Packet Gateway
PSTN	Public Switched Telephone Network
PWS	Public Warning System
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
Rx	Receive
SAS	Spectrum Access System
SCMA	Sparse-Code Multiple Access
SDN	Software Defined Networking
SGW	Signaling Gateway
SIM	Subscriber Identity Module
SIPTO	Selected Internet IP Traffic Offload
SMS	Short Message Service
TAC	Type Allocation Code
TCP	Transmission Control Protocol
TS	Technical Specification
Tx	Transmit
UE	User Equipment
UFMC	Universal Filtered Multi Carrier
UL	Uplink
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
Wi-Fi	Wireless Fidelity
WPS	Wireless Priority Service
WRC	World Radio Conference

APPENDIX B: REFERENCES

[1] “5G: Personal Mobile Internet beyond What Cellular Did to Telephony”, *Gerhard Fettweis, Siavash Alamouti,* IEEE Communication Magazine, February 2014.

[2] “Evolutionary & Disruptive Visions Towards Ultra High Capacity Networks” IWPC‘s MoGIG White Paper Proposal, 2014.

[3] “Scenarios, requirements and KPIs for 5G mobile and wireless system”, METIS Document number: ICT-317669-METIS/D1.1.

[4] T. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Transactions on Wireless Communications, vol. 9, no. 1, pp 3590-3600, Nov 2010.

[5] T. Marzetta, “How much training is required for multiuser MIMO,” Proceedings of Asilomar, CA, Oct. 2006.

[6] ”5G Waveform Candidate Selection”, 5GNOW deliverable D3.1, http://www.5gnow.eu/node/52

[7] “Non-orthogonal Multiple Access (NOMA) for Cellular Future Radio Access”, Vehicular Technology Conference (VTC) Spring), 2013.

[8] “Sparse-code multiple access”, PIMRC, 2013.

[9] Jung Il Choi, et al, “Beyond full duplex wireless”, Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on, 4-7 Nov. 2012,pp 40 – 44.

[10] S. Chen, M. Beach, and J. McGeehan, “Division-free duplex for wireless applications,” In IEEE Electronics Letters, 34, 147–148 (1998).

[11] D. W. Bliss, P. Parker, and A. R. Margetts, “Simultaneous Transmission and Reception for Improved Wireless Network Performance,” In IEEE/SP 14th Workshop on Statistical Signal Processing, pp. 478–482 (2007).

[12] B. Radunovic, D. Gunawardena, P. Key, A. P. N. Singh, V. Balan, and G. Dejean, “Rethinking Indoor Wireless: Low power, Low Frequency, Full-duplex,” Technical report, Microsoft Research (2009) .

[13] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, “Achieving single channel, full duplex wireless communication,” In Proceedings of ACM Mobicom, pp. 1–12 (2010).

[14] M. Jain, J. Choi, T. M. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, “Practical, real-time full duplex wireless,” In Proceedings of ACM Mobicom, pp. 301–312 (2011).

[15] M. A. Khojastepour, K. Sundaresan, S. Rangarajan, X. Zhang, and S. Barghi, “The case for antenna cancellation for scalable full-duplex wireless communications,” In Proceedings of Hotnets, pp. 1–17 (2011).

[16] A. Sahai and G. Patel and A. Sabharwal, “Pushing the limits of Full-duplex: Design and Real-time implementation,” In arXiv.org:1107.0607, (2011).

[17] A. Sahai, G. Patel, and A. Sabharwal, “Asynchronous Full-duplex Wireless,” In Proceedings of COMSNETS, pp. 1–9 (2012).

[18] Melissa Duarte, “Full-duplex Wireless: Design, Implementation and Characterization,” Ph.D. Thesis, Rice University, April 2012.

ACKNOWLEDGEMENTS

The mission of 4G Americas is to advocate for and foster the advancement and full capabilities of the 3GPP family of mobile broadband technologies, including LTE-Advanced, throughout the ecosystem's networks, services, applications and wirelessly connected devices in the Americas. 4G Americas' Board of Governors members include Alcatel-Lucent, América Móvil, AT&T, Cable & Wireless, Cisco, CommScope, Entel, Ericsson, HP, Intel, Mavenir Systems, Nokia, Openwave Mobility, Qualcomm, Rogers, Sprint, T-Mobile USA and Telefónica.

4G Americas would like to recognize the significant project leadership and important contributions of Brian K. Daly and DeWayne Sennett of AT&T, as well as representatives from the other member companies on 4G Americas’ Board of Governors who participated in the development of this white paper.