5G for Mission Critical Communication

Achieve ultra-reliability and virtual zero latency

NOKIA

Contents

1. Executive Summary

Mobile networks must meet new demands as human communications changes from click and wait/background traffic, to interactive, real-time, haptic communication, and introduction of critical machine-to-machine type communications. The networks must provide significantly reduced end-to-end latency and higher reliability than is achievable today. Ultra-reliability is vital for safety. Low latency is crucial to ensure applications are usable and interactive whether human-to-human, human-to-machine or machine-to-machine communication. This white paper outlines the needs for ultra-reliability and virtual zero latency communication as well as the solutions to build 5G networks.

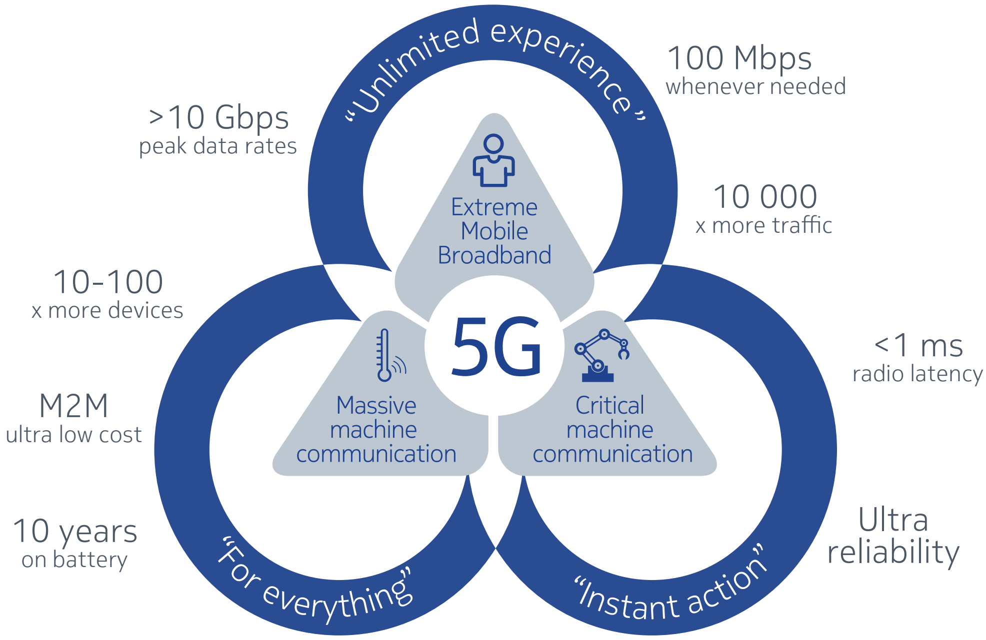

Figure-1: 5G Diversity of service, use case and requirements

The following developments are proposed:

- **Radio Access:** Ultra-reliability is achieved by designing the radio access to include Diversity through redundancy links (Massive MIMO) and Multi-Connectivity, as well as Interference Management and user/service-optimized retransmission mechanisms. The introduction of Flexible Frame Structure can bring radio latency down to milliseconds.
- **Programmable 5G multi-service architecture:** The key components come from Network Slicing, Programmable Networks, Network Resiliency and Mobile-Edge Computing. These building blocks will ensure that the network is flexible, reliable and optimized to bring content closer to users instantly.

- **Device to Device communication:** Direct D2D communication will be an important communication method in 5G. It is characterized by short distance between communication devices, no user-plane processing needed by the network elements, and bypassing transport networks, which helps to minimize delay. An additional D2D link can be used as a diversity path to increase reliability or to extend network coverage, helping to improve availability.

2. Why do we need ultra-reliability and low latency communication?

Minimizing latency and increasing reliability opens up potentially lucrative new business opportunities for the industry, arising from new applications that simply will not work properly if network delays are too high. Latency determines the perception of speed. Real-time functionality demands the lowest possible delay in the network. Reliability creates confidence in users that they can depend on communications even in life-threatening situations.

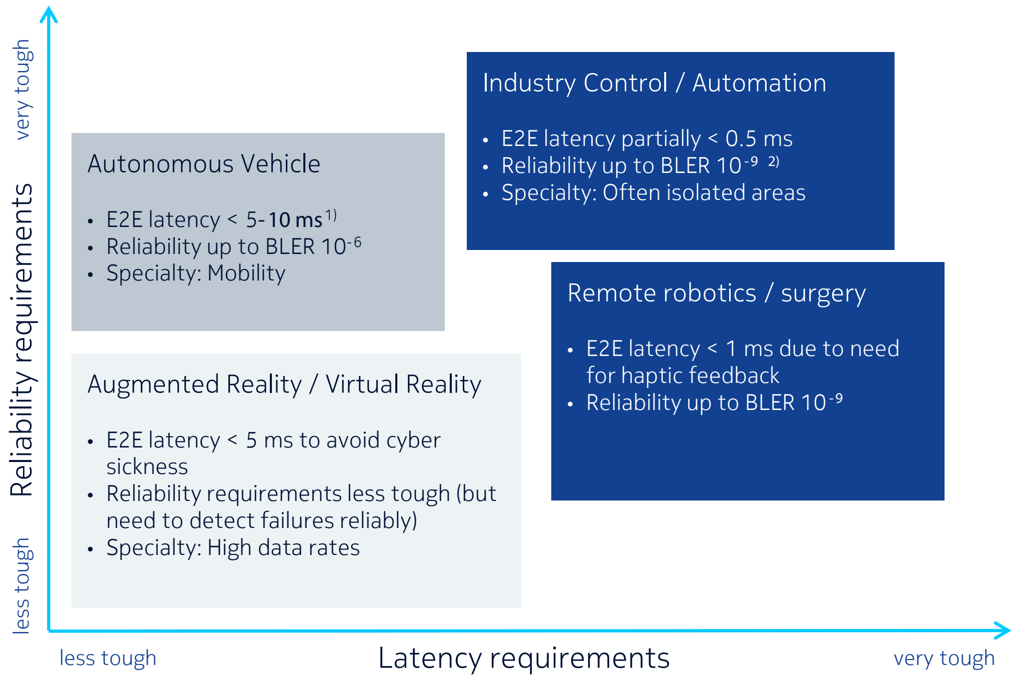

Figure-2: 5G Use cases requiring low latency and/or high reliability

2.1 Autonomous vehicles

Autonomous vehicles is a hot topic for many industry players from car manufacturers, consumers, and insurance companies to governments. The US Secretary of Transportation has said that driverless cars will be in use all over the world by 2025. The IEEE predicts up to 75 percent of vehicles will be autonomous in 2040[3]. While the autonomous vehicles developed today rely mostly on onboard sensors and systems, their performance and safety could be vastly improved through 5G communications.

Autonomous vehicles can reduce accidents and improve road utilization as vehicles can be driven closer to each other and more safely than human drivers can achieve. Transportation companies can take advantage of autonomous car fleets. The fleets can be utilized more effectively with fewer accident caused by human error. In addition, real-time, ultra-reliable communications between vehicles, infrastructure and smartphones could enable traffic to flow more smoothly, eliminating traffic jams. Commuting time can be used for other activities with the help of autonomous vehicles. This might save an hour per day for people living and commuting in cities.

The communication system needs to be extremely reliable as it involves human safety. The end-to-end latency requirement needs to be as low as 5-10 ms[1]

2.2 Augmented reality / virtual reality

Augmented Reality (AR) enhances a real-world view with graphics. Real-time information is displayed based on the user's location and/or vision.

Virtual Reality (VR) creates a totally new user experience with the user being in a fully immersive environment. The AR/VR device needs to track user movements accurately, process the movement and receiving image, then display the response immediately. An end-to-end latency of more than 5 ms would lead to cyber sickness, an uncomfortable and nauseating customer experience.

AR will enhance existing service experiences. For example, shoppers can experience how a dress would look on them without trying it on. AR can also be used in emergency situations, for example, firefighters could use AR to see ambient temperature, a building's layout, exits and potentially dangerous areas[4]. Police officers could use AR with facial recognition to identify a suspect in real-time from the police database before an arrest is made.

VR uses are extensive, not just gaming and entertainment. Students could learn inside a VR environment conducted by a remote teacher. Students can gain experiences as large as the inception of the universe or as small as how to split an atom. In product development, VR can be used to design and prototype products before they are built, shortening development time and cost.

NOKIA

2.3 Remote robotics / surgeries

Remotely controlling robots, rovers, devices or avatars in real time can help us to work safely outside dangerous places. Hospitals can arrange remote robotic surgeries via a customized 5G network as if the surgeon was physically present. For public safety, robots could be sent to work in dangerous situations, such bomb disposal or firefighting. The system needs to be extremely reliable with BLER up to 10^{-9} and end-to-end latency of less than 1 ms to support the necessary haptic feedback.

Many haptic screens and devices are being developed currently to respond to touch and provide tactile sensations by varying the friction between the user's finger and the screen. This creates an experience of "You feel what you touch (remotely)". An early example is the new iPhone, which introduces 3D human sensitive touch.

The combination of haptic interaction and 360° cameras feeding live video over a 5G network to a VR head mounted device will produce a powerful experience as though the user is actually in the remote location and in control.

2.4 Industry control / automation

Industrial networks have stringent requirements because they require fast machine-to-machine communication and ultra-reliable connectivity. A system failure could mean loss of equipment, production, or even loss of life. Time-critical process optimization is a key requirement for factories-of-the-future (FoF)[5]. The need for wireless ultra-reliability and virtual zero latency will be driven by uses that include instant optimization based on real-time monitoring of sensors and the performance of components, collaboration between a new generation of robots, and the introduction of wireless connected wearables and augmented reality on the shop floor.

Machines can receive, analyze and execute tasks much more quickly than humans. Therefore, machine-to-machine communication requires extremely low latency, for example closed-loop control applications for industry automation need lower than 1 ms latency. High reliability (packet error rate $< 10^{-9}$) is important to maintain close synchronization and high availability. Furthermore, the overhead should be kept to a minimum to ensure a tolerable spectral efficiency with small packet payloads[2].

Indoor traffic control and indoor mobility control of shop floor equipment typically have cycle times around 1-10ms. The highest demands are from actuators and sensors requiring cycle times of less than 1ms with a jitter of less than 1μs. While today's wired systems meet these requirements, 5G will create a unified platform that addresses a wide range of needs from the company supply chain, to inter-enterprise communication, to the control of actuators/sensors on the factory floor. This will reduce administrative costs compared to maintaining multiple systems, eliminate the cost to install wiring and increase flexibility to change production flow in the factory.

3. Solution Description

5G introduces new aspects to address ultra-reliability and low latency by evolving radio access as well as network architecture.

3.1 Radio access

Radio access is close to the user and has a significant impact on reliability and latency. While LTE supports today's broadband traffic, more advanced technology will be needed to provide the ultra-reliability and low latency required by new use cases and applications. Nokia recommendations for 5G radio access evolution include the following.

3.1.1 Diversity and interference management for high reliability

The quality of the radio link between the base station and mobile terminal direct affects overall system reliability. Signal to Interference and Noise Ratio (SINR) is used to measure the quality of the radio link. The higher the SINR, the lower packet error probability, which results in higher reliability. SINR can be increased by enhancing the signal, for example with redundancy, and/or to reduce interference and noise via interference management.

Microscopic diversity

More antennas provide better coverage (or space availability). 2x2 antenna schemes are most commonly used today. The latest 3GPP releases support 8x8 and full dimensional Multiple-Input Multiple-Output (MIMO) schemes. In general, increasing the number of antennas (higher order of MIMO) results in a better signal and higher SINR. Exploiting this principle, massive MIMO with hundreds or thousands of antennas is recommended for 5G systems.

Macroscopic diversity (multi-connectivity)

Multi-connectivity can also ensure ultra-reliability and low latency. There are several ways to implement multi-connectivity, including multi-RAT combination of LTE and 5G, and multi-cell/multi-node coordinated tx/rx techniques. For the latter, multiple base stations transmit the same data synchronously, which is then combined at the receiver.

Performance can be improved by tackling shadowing effects through diversity and redundancy and by increasing the total received power of the desired signal. Macroscopic diversity can be used for Mission Critical Communications (MCC), but also for enhancing mobility procedures critical to ensure ultra-reliability and low latency. The break-before-make method of LTE might result in data interruption times of up to 55 ms, which become an obstacle for reliable communication. Virtually zero interruption time must be the target for 5G, with make-before-break at every handover for always available data connectivity.

Interference management

Mitigating interference by either network-based or terminal-based techniques has been identified as a promising complementary solution to improve the SINR[6]. Reducing the received interference from neighboring base stations or terminals improves SINR. As a rule of thumb, cancelling the strongest or two strongest interferers is usually enough to achieve most of the potential gain.

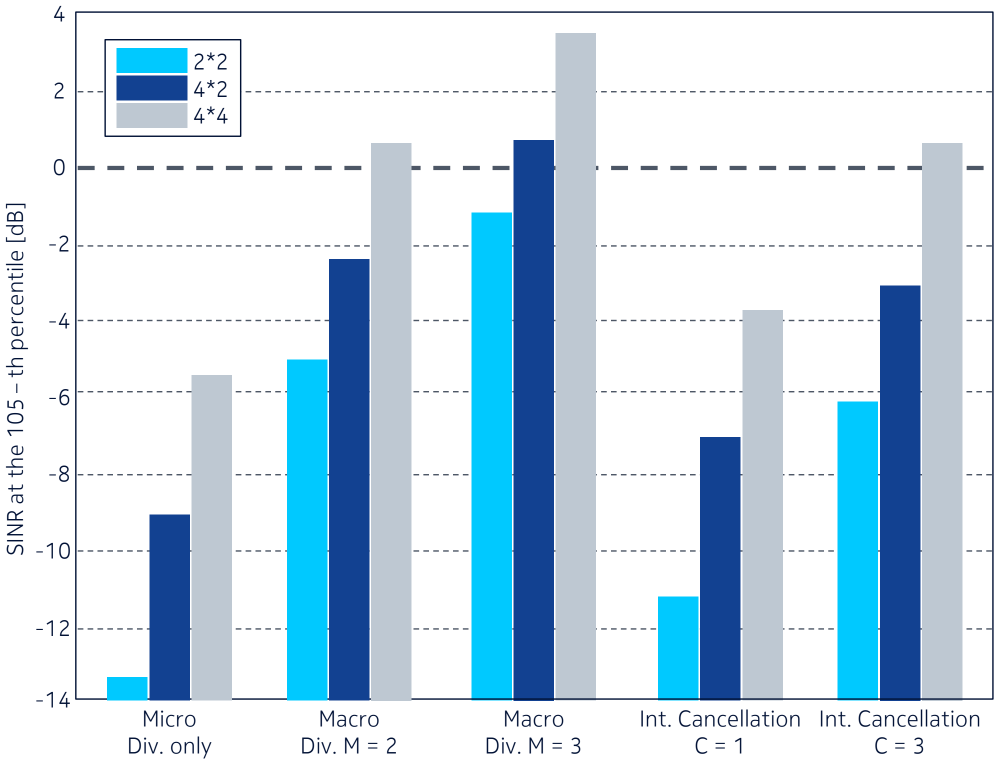

Figure 3: SINR for very stringent reliability by using different orders of MIMO, Macroscopic diversity (M), and Interference Cancellation (C) [7]

LTE solutions, for example ICIC and eICIC, are rather proactive and semi-statically configured. In 5G, more dynamic solutions tailored for increasing the reliability are recommended instead of increasing the average capacity as in LTE. For example, selective blanking of the strongest interferers during retransmissions can greatly increase the probability of success, as shown in Figure 4

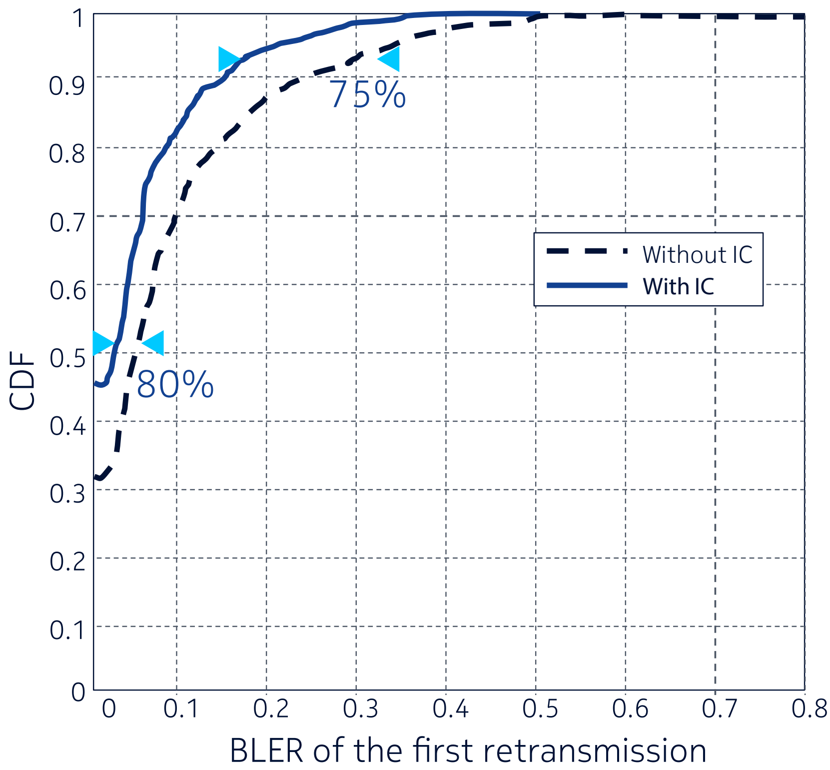

Figure 4: Interference coordination for high reliability

3.1.2 Flexible frame structure for low latency

The 1 ms transmission time interval (TTI) in LTE and 8 ms waiting time at every retransmission results in end-to-end latencies of 20-40 ms[8]. For 5G, a shorter frame and faster processing time at retransmissions will produce lower latency. One option for the coexistence of Mission Critical Communication (MCC) and Mobile Broadband (MBB) is flexible multiplexing of users on a shared channel with dynamic adjustment of TTI in coherence with the service requirements per link. This allows optimization of spectral efficiency, latency and reliability for each link. The frame structure is based on in-resource physical layer control signaling that follows the corresponding data transmission for each user.

The basic concept is illustrated by the time-frequency grid shown in Fig 5. Each tile refers to the smallest allocation unit of time-duration Δt and frequency size Δf. The value of Δt determines the minimum TTI size for scheduling a user, as well as the resolution for other TTI scheduling options.

In coherence with such flexible design, enhancement of HARQ retransmissions entails flexible duration of the ACK/NACK duration and configurability per user.

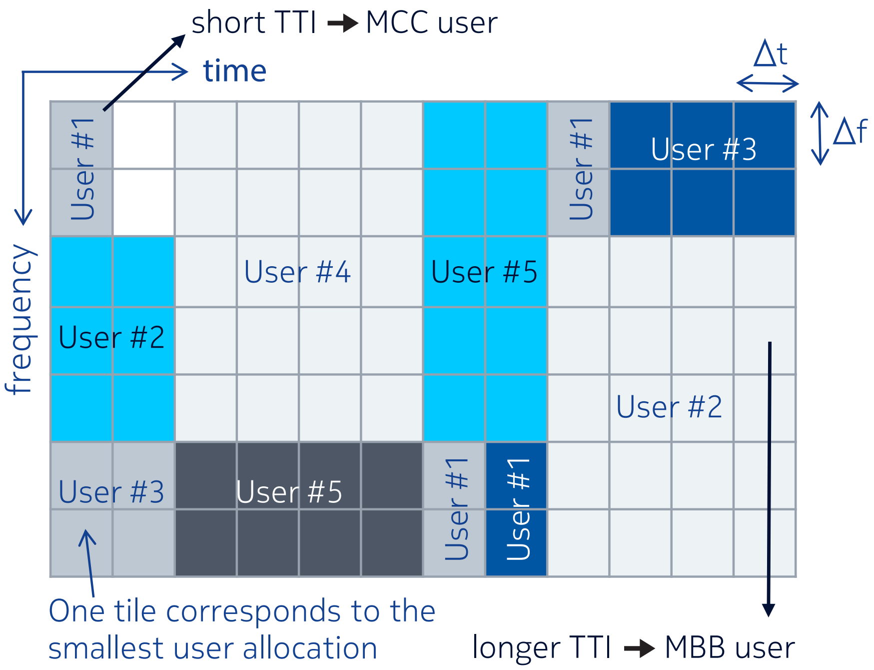

Figure 5: Optimized frame structure for low latency

3.2 Programmable 5G multi-service architecture

To address ultra-reliability and low latency we will need to build a resilient system dynamically managed that offers high-availability and brings content close to users, on demand and instantly. The key network architecture evolution comes from the following concepts.

3.2.1 Network slicing

Network architecture has been traditionally built around a specific use-case. For example, GSM was built primarily for voice and LTE for mobile data. In the future, this “one use case per one physical network” approach will be obsolete. The 5G network will be designed to be flexible enough for an operator to create an instance of an entire network virtually, that is, a customized network for each diverse use case. Different customized virtual networks will exist simultaneously and without interfering with each other. This is so-called Network Slicing. For example, a customized virtual network for ultra-low latency autonomous vehicle control can co-exist with a customized virtual network for 3D video /4K screen viewing, which requires extremely high throughput.

3.2.2 Programmable networks

A flexible network will be needed to adapt to various performance requirements. Software-defined functions create a programmable infrastructure, which means that the path of packets through the network is not restricted by a fixed architecture and can be programmed and optimized for latency. Software Defined Networks (SDN) in the mobile backhaul (MBH), aggregation and backbone network enable the use of traffic optimization, bandwidth allocation and Mobile-Edge Computing (MEC) to reduce latency.

Transport SON agent(s) can collect information such as delay (per class/per interface), loss (per class/per interface), throughput (total/per PHB/per GTP), queue length, active bearers or active devices and apply this to SDN control. Nokia already offers all-IP transport solutions, including fully integrated options, for high scalability, high capacity, low delay and close synchronization to meet the connectivity, backhaul and fronthaul requirements of a modern mobile broadband network.

3.2.3 Network resiliency

Network elements must deliver high availability. This can be achieved by pooling a number of core elements and using load balancing to ensure no interruption in service should one or more core elements fail. The failed core element can be left to recover while the other core elements continue to function. Even should the backhaul become unavailable, service will continue almost unaffected by using a stand-alone mode of operation.

3.2.4 Mobile-Edge Computing (MEC)

Moving the gateway and application server closer to the radio can significantly reduce latency even further. Services are no longer tied to a single point-to-point IP connection, enabling the connectivity path to be freely chosen according to actual service demand. This any-to-any connectivity model, in which devices communicate directly through local switching at the RAN level avoids unnecessary data forwarding to centralized mobility anchors (gateways). This offers the shortest and best path for routing traffic that needs low latency while at the same time ensuring continuity and seamless mobility.

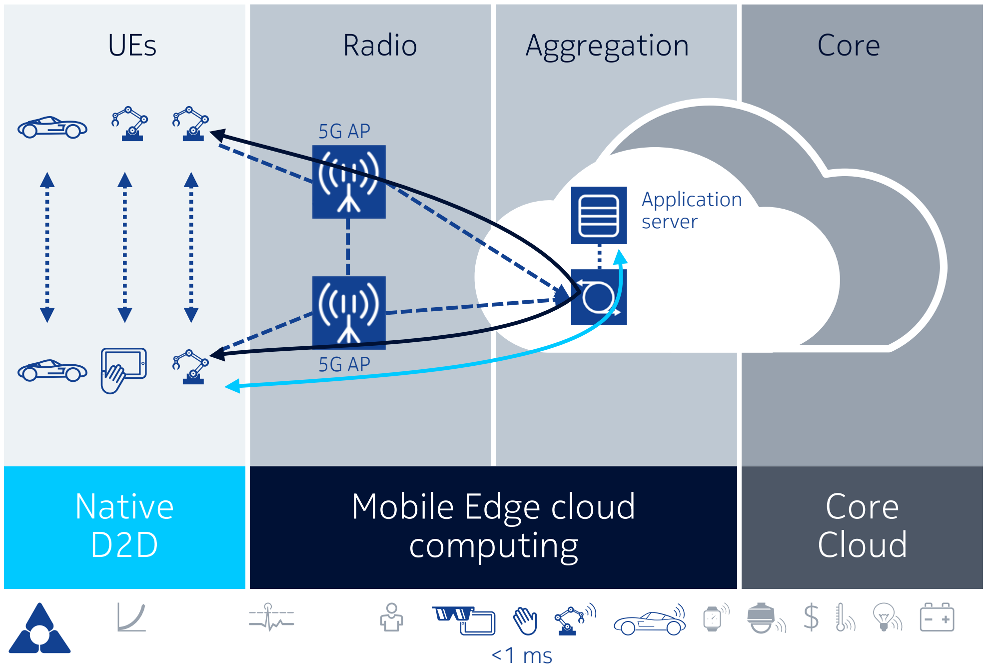

Figure 6: Mobile-Edge Computing

3.3 Device-to-Device communication

Device-to-Device (D2D) communication is direct communication between two devices without data traffic going through any infrastructure. D2D will be an important 5G communication method to address ultra-reliability and low latency requirements. Low latency is achieved through direct communication between devices over short distances with minimal propagation delay, no involvement of network elements for processing the traffic data, or any transport network that introduces delay[9]. Reliability is also improved because an additional D2D link provides a diversity path and extends network coverage. The D2D ad-hoc network can be used as a backup solution in case of a failure of the network infrastructure or if network infrastructure becomes unavailable.

Radio Resource Management (RRM) in D2D

Resource allocation for D2D communication can be centralized or distributed. In a distributed solution, devices can transmit data immediately without any association procedure or dedicated control channel. The centralized solution on the other hand effectively avoids collisions, but increases complexity and average delay.

Figure 7 shows an example of Vehicle to Vehicle (V2V) communication. Each vehicle is periodically broadcasting information about its location, speed, travelling direction and more. If there is high traffic density, the probability of collision increases. One distributed option is to choose a geo-location-based access using the location information (e.g. GPS coordinates) of a device as a unique variable (or locally unique). The road is divided into segments such that the orthogonal resource chosen depends on the location information.
As shown in the example in Figure 7, the road lane is divided into consecutive areas mapped into each access resource (AR).

With D2D, diversity can be further exploited. One example is the simultaneous direct D2D link and cellular link operation, with different ways for selecting the optimal link for packet transmission/retransmission and potentially combining the received signal over multiple links.

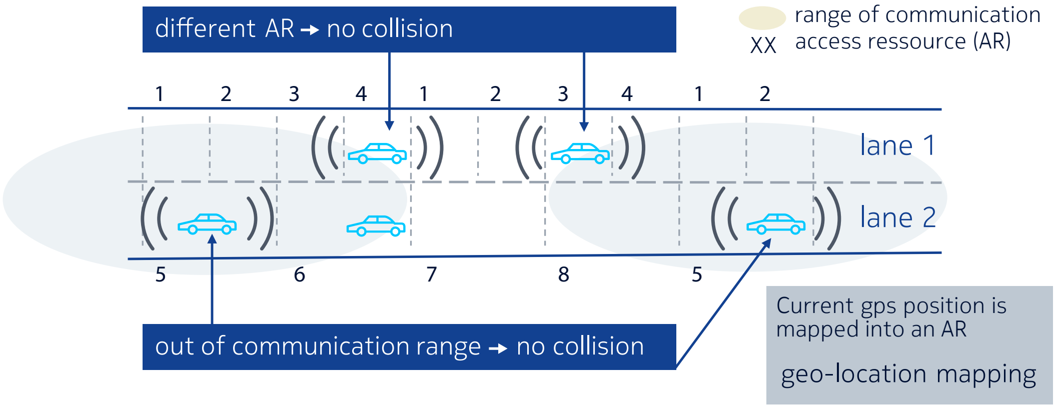

Figure 7: Geo-location-based access for D2D

D2D communications can benefit from Full Duplex (FD), such that transmission and reception take place simultaneously and in the same frequency band. Therefore, the problem of pre-configuring the transmission direction - downlink or uplink - does not apply. With a minimal control overhead, latencies are reduced. Full Duplex operation does however require a good level of attenuation of self-interference, that is the interference of own transmitted signal to own receiving antenna.

4. Conclusion and Next steps

Nokia foresees a phased approach for the deployment of radio access and programmable 5G multi-service architecture[10].

5G radio access with Diversity, Multi-Connectivity, Interference Management and Flexible Frame Structure is likely to be introduced and integrated with the LTE core network in the first phase. At the same time, programmable 5G multi-service architecture will introduce Network slicing, Programmable Networks, Network Resiliency and Mobile-Edge Computing (MEC).

In a second phase, 5G radio access will be integrated with a 5G core network without the need for an LTE anchor. In this phase, the end-to-end 5G system will meet the requirements of mission critical communication by supporting lowest latency with full mobility and highest reliability.

Nokia has already showcased various 5G technologies and is working to make the 5G ultra-reliability, low latency system a reality:

- Nokia's new AirFrame Data Center Solution helps operators bring their data centers into the cloud and drives 5G network architecture evolution.
- Nokia unveiled its programmable 5G multi-service architecture in September 2015. The architecture will help operators to offer network functions to any kind of industry under a Network-as-a-Service business model11.
- During Mobile World Congress 2015 Nokia presented use cases for mission critical communication and key technology enablers to fulfil ultra-reliability and low latency requirements.
- In the European 5G Public Private Partnership (5G-PPP) Nokia takes a leading role in Europe's largest 5G projects. Nokia has one of the leads in the 5G NORMA as the project coordinator. In this project a novel radio multiservice adaptive network architecture will be developed. Furthermore, Nokia takes the technical lead in the METIS-II project for the overall 5G RAN design and the spectrum work in METIS-II.
- Nokia and Deutsche Telekom have demonstrated real-time vehicle communication. The project involved upgrading Deutsche Telekom's existing LTE network at sections of the A9 motorway test bed with Nokia MEC technology, enabling information to directly route within cells, instead of transporting data through the mobile network via the cloud12.

Nokia has been demonstrating various 5G ultra-reliability, low latency use cases through several proofs of concept including:

- Autonomous vehicles communicating and steering in a city featuring car platooning and driving without the need for traffic lights to control crossings. The proof of concept demonstrates the need for ultra-reliability, low latency communication between cars and the network to prevent accidents and better road utilization.

- Collaboration in a Virtual Reality environment focused on a training/education task. A student learns about astronomy inside the VR world conducted by a remotely located teacher. As the teacher points to constellations, the student can follow and turn them into life. An ultra-low latency and high throughout proof of concept 5G system is used to demonstrate the fully immersive user experience.
- Industry 4.0 featuring fast and synchronized collaboration of robots to balance a ball on a moving platform. As a user moves a ball in any direction, the robots coordinate and react to move the ball back to the optimal location. An ultra-low latency 5G proof of concept system is used as a communication platform among robots in order to fulfill the task.
- Reliable high performance low latency multicast provides new viewing experiences for stadium visitors. Multiple live video channels from multiple cameras around a stadium are made available for all visitors in the stadium to view simultaneously. Visitors can switch between the real-time video channels and can recommend a channel to another device. The user of the receiving device can accept the suggestion and watch the channel instantly.

This is just the beginning of a new era of mobile communication, and in this paper we can only briefly describe a few examples of the possibilities for mission critical communications. The opportunity is no longer limited to the Internet and the telecommunication industry, but will expand to other industries such as automotive, healthcare, manufacturing and logistics, even to the government/public sector. Nokia along with leading operators and partners will continue to drive the changes and lead in the development of 5G technologies.

Abbreviations

3GPP	3rd Generation Partnership Project
ACK/NACK	Acknowledged/Not Acknowledged
AR	Access Resource
AR/VR	Augmented Reality / Virtual Reality
BLER	Block Error Rate
D2D	Device-to-Device
FD	Full Duplex
FoF	Factory of the Future
GPS	Global Positioning System
HARQ	Hybrid automatic repeat request
ICIC/eICIC	Inter-Cell Interference Coordination/Enhanced
IEEE	Institute of Electrical and Electronics Engineers
MBB	Mobile Broadband
MBH	Mobile Backhaul
MCC	Mission Critical Communication
MEC	Mobile-Edge Computing
MIMO	Multiple Input Multiple Output
RAN	Radio Access Network
RAT	Radio Access Technologies
RRM	Radio Resource Management
RTT	Round Trip Time
SDN	Software-Defined Network
SINR	Signal to Interference and Noise Ratio
SON	Self-Organized Network
TTI	Transmission Time Interval
TX/RX	Transmitter/Receiver
V2I/V2V	Vehicle-to-Infrastructure/ Vehicle-to-Vehicle

NOKIA

References

Ref 1: METIS D1.1: Scenarios, requirements and KPIs for 5G mobile and wireless system

Ref 2: IEEE: A. Frotzscher et al., Requirements and current solutions of wireless communication in industrial automation, ICC 2014

Ref3: Driverless car market watch

Ref4: Gartner Says Augmented Reality Will Become an Important Workplace Tool

Ref5: 5G-PPP White paper: 5G and the factories for the future

Ref6: IEEE: Interference coordination for dense wireless networks

Ref7: Signal Quality Outage Analysis for Ultra-Reliable Communications in Cellular Networks

Ref8: IEEE: A Flexible Frame Structure for 5G Wide Area

Ref 9: Chapter 5 D2D Communication in A. Osseiran, J. Monserrat, and P. Marsch, 5G Mobile and Wireless Communications Technology, Cambridge University Press, ISBN: 9781107130098, 2016

Ref10: Nokia: 5G system of System white paper

Ref11: Nokia Networks unveils its programmable 5G multi-service architecture

Ref12: Continental, Deutsche Telekom, Fraunhofer ESK, and Nokia Networks Showcase First Safety Applications at “Digital A9 Motorway Test Bed”

Nokia is a registered trademark of Nokia Corporation. Other product and company names mentioned herein may be trademarks or trade names of their respective owners.

Nokia
Nokia Solutions and Networks Oy
P.O. Box 1
FI-02022
Finland

Visiting address:
Karaportti 3,
ESPOO,
Finland
Switchboard +358 71 400 4000

Product code C401-011946-WP-201601-1-EN

© Nokia 2016