NOKIA Bell Labs

VQEG

Deconstructing AR applications for 5G

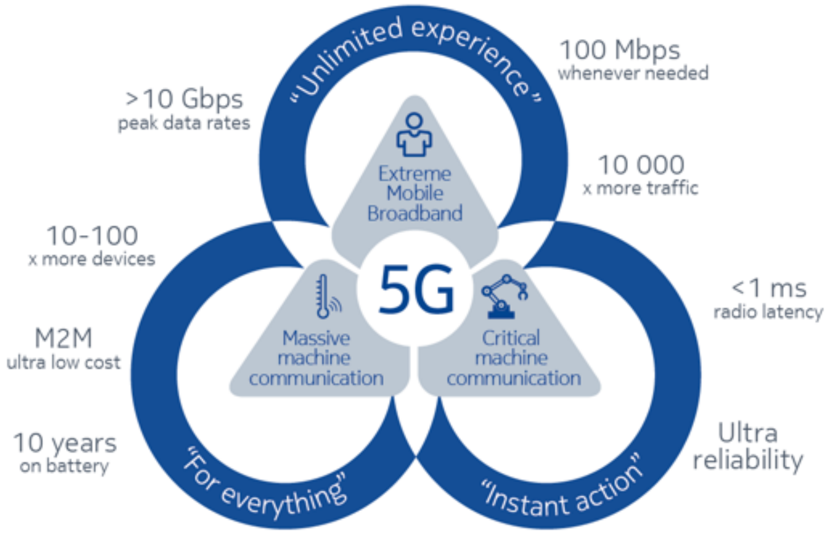
Diego González & Pablo Pérez

VQEG Plenary Meeting, Shenzhen, October 2019

What’s so new about 5G?

Throughput – Latency - Density

eMBB (Enhanced mobile broadband)



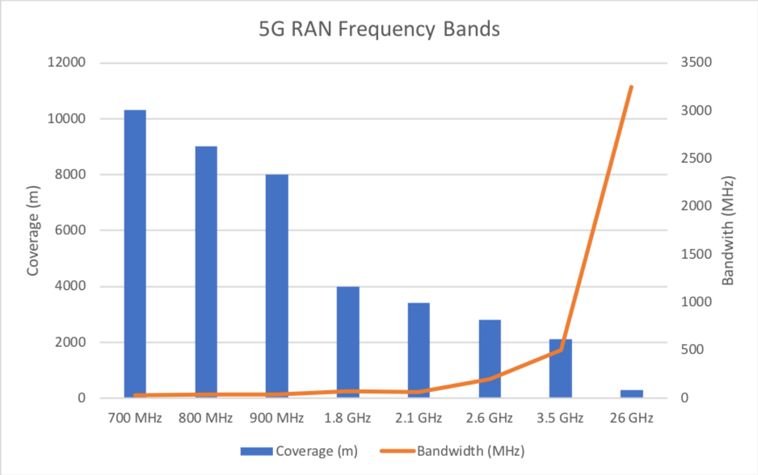
mMTC (Massive machine type communication)

URLLC (Ultra-reliable low latency communication)

https://networks.nokia.com/5g/resources

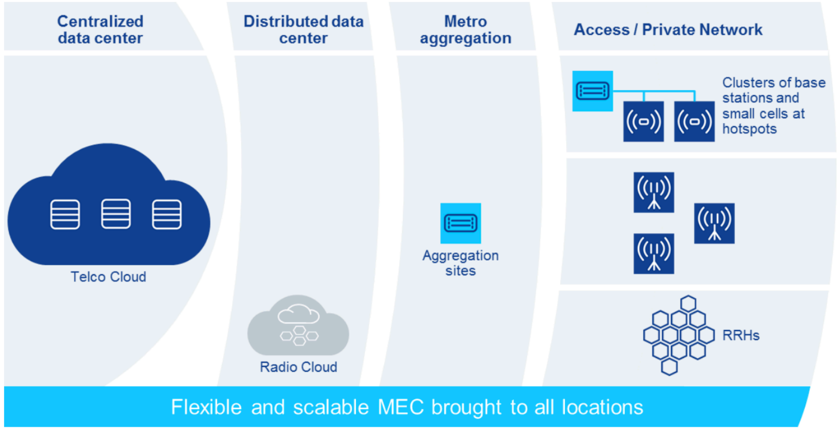
What’s so new about 5G?
Architecture

New Radio Access Network



- More frequencies (inc. mmW)
- Carrier aggregation, massive MIMO, …
- Network slicing

Multi-Access Edge Computing



- Edge cloud
- Virtualized apps
- Low latency

5G Ultra Dense Networks
Key Performance Indicators

Ultra Dense Networks research

- RAN PHY/Link/MAC enhancements
  - Technology: mMIMO, mmW, VLC
- Dynamic spectrum management
  - Spectrum sharing and carrier aggregation
- Energy efficiency

- **KPIs**
  - Bandwidth, throughput
  - Availability
  - Latency
  - Energy consumption
  - …

Video Challenges

- Distributed computing
- Low-latency coding
- Remote rendering
- Responsiveness to variable throughput and latency

- **KQIs** to be defined
  - Quality of the xR Experience
  - Breakdown into factors: interactivity, visual quality, segmentation quality, segmentation delay, etc..

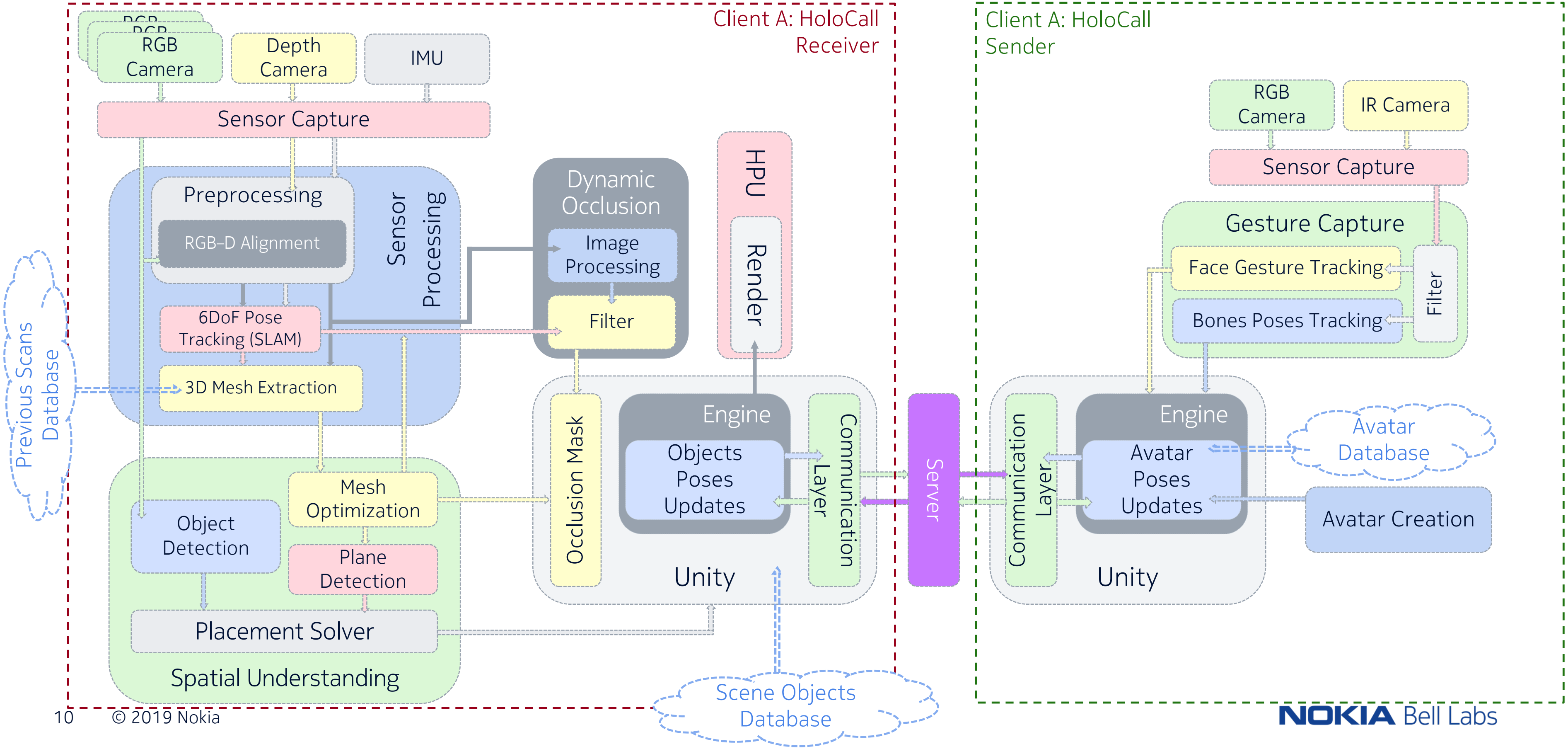
Deconstructing AR
From network/system KPIs into QoE

Network KPIs	Application KPIs	Application KQIs	QoE Model
RAN throughput	TCP Goodput	Visual Quality	
User concurrency	AL-RTT	Responsiveness	QoE = f(KQIs)
MEC CPU/GPU resources	UDP loss rate	Localization accuracy	
RAN packet loss rate	…	…	

Augmented Reality Telepresence

The first breakdown

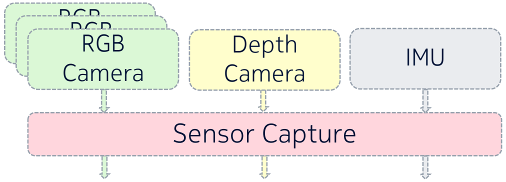
Diagram



NOKIA Bell Labs

The first breakdown

AR Device - Sensor Capture



Hololens 1 – Sensors

- 1 IMU
- 4 Grayscale cameras
- 1 120deg Depth camera
- 1 HD Video Camera
- 4 Channel microphone

Hololens 2 – Sensors

- 1 IMU
- 2 IR Cameras for Eye Tracking
- 4 RGB cameras
- 1-MP ToF Depth sensor
- 1 1080p30 Video Camera
- 5 Channel microphone

Samsung Galaxy S10 5G

- 1 12MP Camera
- 1 12MP Wide Camera
- 1 16MP Ultra Wide Camera
- Depth Camera (72deg)
- 1 Microphone
- 1 IMU

NOKIA Bell Labs

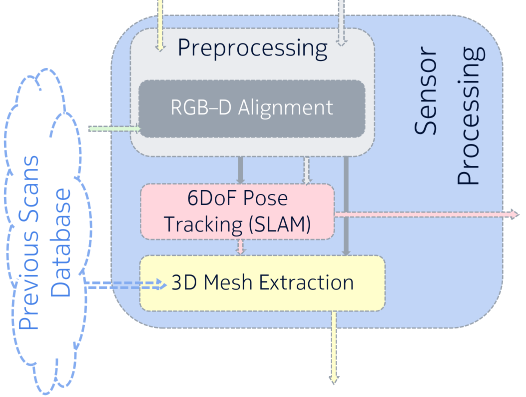
The first breakdown

AR Device - Sensor Processing

The goal of this block is to accurately localize the AR device and extract 3D understanding from the real environment

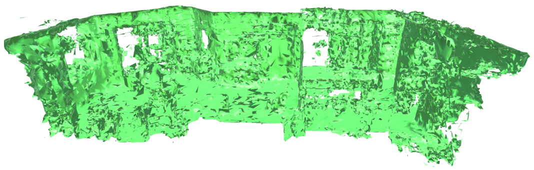
A. <u>Input – Output Analysis</u>

The simplified input-output diagram is:



Input	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
RGB Feed	30 - 60	1920 * 1080 * 3 * 8 bits	15 – 30
Depth Feed	30 - 60	1920 * 1080 * 8 bits	5 - 10
Cloud Mesh	0 (1 Time)	(45 + 13*triangles + 10*vertex)*8 bits	-

Output	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
6 DoF Pose	30 - 60	3*float + (4 or 12)*float = (208 – 480)bits	0.006 – 0.028
3D Mesh	0.5	(45 + 13*triangles + 10*vertex)*8 bits	-
RGB-D	30 - 60	1920 * 1080 * 4 * 8 bits	20 - 40



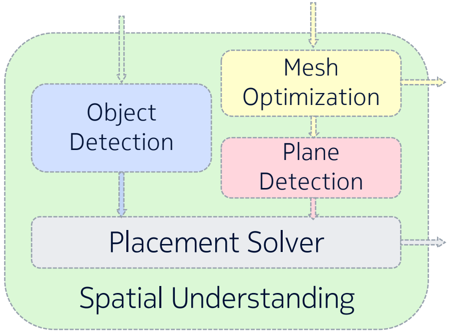
Example Room:

- Dimensions: 12.5x3.2x9.2 m
- Vertex Num: 100286
- Triangle Num: 438711
- Total Serialized Size: 13.72 Mbit
- Data Rate: 6.86 Mbits (if all the mesh is updated)

  [1] Mesh Serializer Unity: https://wiki.unity3d.com/index.php/MeshSerializer2

The first breakdown

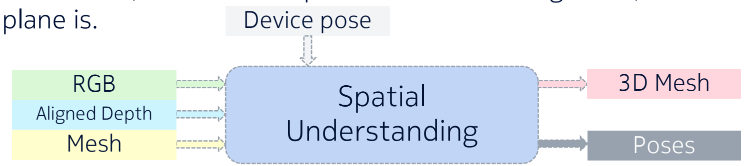
AR Device – Spatial Mapping



The goal of this block is to extract semantics from the 3D scanned environment and place the virtual content in the real world according to such semantics. For instance, if we want to place an avatar sitting down, we need to identify what is a chair, and where the sitting plane is.

A. <u>Input – Output Analysis</u>

The simplified input-output diagram is:



Input	**Frequency (Hz)**	**Raw Data Rate Per Frame**	**Data Rate (Mbps)**
RGB Feed	5 - 20	1920 * 1080 * 3 * 8 bits	15 – 30
Aligned Depth Feed	5 – 20	1920 * 1080 * 8 bits	5 - 10
Device Pose	5 - 20	208 – 408 bits	<< 1 Mbps
Mesh	0.5 – 1 Hz	(45 + 13*triangles + 10*vertex)*8 bits*	-

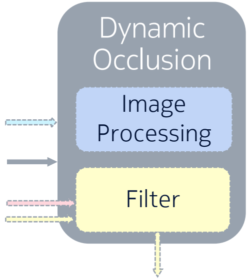
Output	**Frequency (Hz)**	**Raw Data Rate Per Frame**	**Data Rate (Mbps)**
6 DoF Pose Virtual Objects	1 Time per object	3*float + (4 or 12)*float = (208 – 480) per object	~0
3D Mesh	0.5 – 1 Hz	(45 + 13*triangles + 10*vertex)*8 bits*	-

* In this case, only the updates in the mesh(es) are received, so the data rate should be much smaller than in the cloud mesh case

NOKIA Bell Labs

The first breakdown

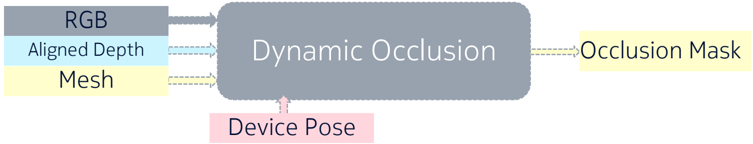
AR Device – Dynamic Occlusion



The goal of this block is to handle the dynamic of occlusion of virtual object. It is a key block in almost every AR application and it is still a problem that has not been solved in the state of the art. It requires high computation power, and extremely low latency to satisfy the very demanding real-time constraints.

A. Input – Output Analysis

The simplified input-output diagram is:



Input	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
RGB Feed	30 - 60 (120?)*	1920 * 1080 * 3 * 8 bits	15 – 30
Aligned Depth Feed	30 - 60	1920 * 1080 * 8 bits	5 - 10
Mesh	0.5 – 1 Hz **	(45 + 13*triangles + 10*vertex)*8 bits*	-
Device Pose	5 – 20	208 – 408 bits	<< 1 Mbps

* In see-through devices it might be necessary to increase the update frequency of the color/depth feeds to ensure a proper occlusion quality.

** After the room has been scanned and the final optimization si done, is not necessary to keep updating the mesh.

Output	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
Occlusion Mask	30 – 60 (120?)	1920 * 1080 * (8 – 12)	5 - 15

The first breakdown

Calling Device - Sensor Capture

RGB Camera

IR Camera

Sensor Capture

Kinect 1

- Color: 640x480x32bpp @ 30fps
- Depth: 320x240x16bpp @ 30fps
- Audio: 16bit @ 16kHz
- 20 joints per user

Kinect 2

- Color: 1920x1080x16bpp @ 30fps
- Depth: 512x424x16bpp @ 30fps
- IR: 512x424x11bpp @ 30fps
- Latency: 60 ms with processing
- Audio: 4-mic array with 48kHz
- 26 joints per user

Realsense D435

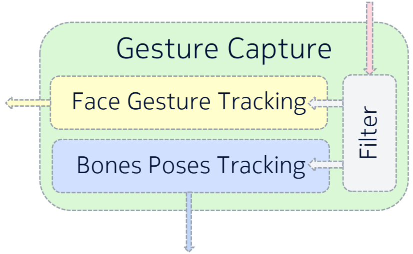
- Color: 1920x1080 @ 30fps
- Depth: 1280x720 @ 30-90 fps
- Global Shutter

Realsense D415

- Color: 1920x1080 @ 30fps
- Depth: 1280x720 @ 30-90 fps
- Rolling Shutter

NOKIA Bell Labs

The first breakdown



Calling Device - Gesture Capture

The goal of this block is to track the caller's joints' positions and rotations along with his/her face gestures in real time. The input is the RGB and depth feeds. The output is the real-time update poses of the main face anchors, and the body joints.

A. Input – Output Analysis

The simplified input-output diagram is:

Input	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
RGB Feed	30 - 60 (120?)*	1920 * 1080 * 3 * 8 bits	15 – 30
Aligned Depth Feed	30 - 60	1920 * 1080 * 8 bits	5 - 10

Output	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
Joints Poses	30 – 60	(204-480)*(20 or 26) = (4080 – 12480)bits	< 1Mbps
Face Anchors – RGB-D *	30 - 60	1920 * 1080 * 4 * 8 bits	20 - 40
Face Anchors – Only Anchors **	30 - 60	(30 to 100 points)*96 bits = 2880- 9600 bits	< 1Mbps

* Processing done on the receiver side

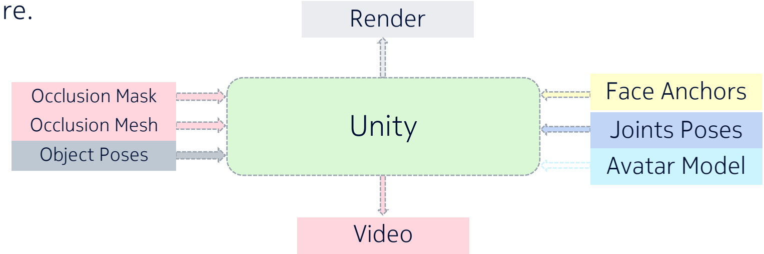
** Processing done on the sender side

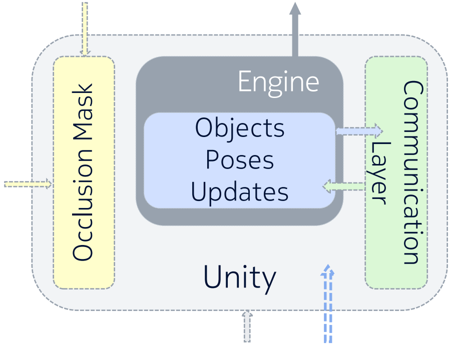
The first breakdown

Unity Client Receiver

This is the main app on the receiver side. The processing and memory requirements analysis will be done in the future.

A. Input-Output Analisys:

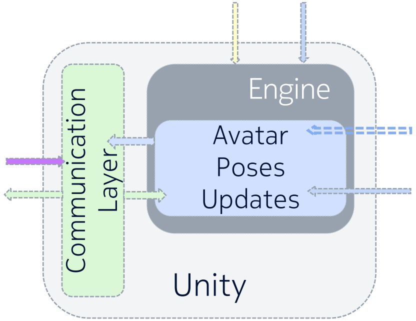




Input	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
Occlusion Mask	30 – 60 (120?)	1920 * 1080 * (8 – 12)	5 - 15
Optimized Mesh	1 Time*	(45 + 13*triangles + 10*vertex)*8 bits	-
Objects poses	1 Time *	208 – 408 bits per object	-
Face Anchors – Only Anchors	30 - 60	(30 to 100 points)*96 bits = 2880- 9600 bits	< 1Mbps
Joints Poses	30 – 60	(204-480)*(20 or 26) = (4080 – 12480)bits	< 1Mbps
Avatar Model	1 Time	10-100MB = 80-800 Mbits	-

Output	Frequency (Hz)	Data Rate (Mbps)
Rendered Frame	30 – 60	~50 Mbps
Conference Video (360?)	30-60	15– 30 Mbps

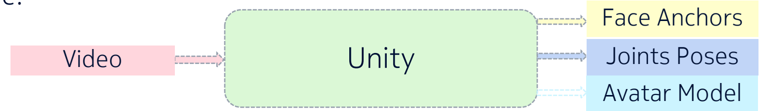
The first breakdown



Unity Client Receiver

This is the main app on the sender side. The processing and memory requirements analysis will be done in the future.

A. Input-Output Analisys:



Input	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
Conference Video (360?)	30 - 60	30-60	15– 30 Mbps

Output	Frequency (Hz)	Raw Data Rate Per Frame	Data Rate (Mbps)
Joints Poses	30 – 60	(204-480)*(20 or 26) = (4080 – 12480)bits	< 1Mbps
Face Anchors – RGB-D *	30 - 60	1920 * 1080 * 4 * 8 bits	20 - 40
Face Anchors – Only Anchors **	30 - 60	(30 to 100 points)*96 bits = 2880- 9600 bits	< 1Mbps
Avatar Model	1 Time	10-100MB = 80-800 Mbits	-

AR Holocall
Next steps

- Finish breakdown analysis
- Build a (simplified) prototype
- Do some measurements
- Create a first KQI/QoE model

NOKIA Bell Labs

5GKPI
What’s next?

- Contribution to ITU-T standardization?
  - Competition vs collaboration?
  - Interest in Q13? Others?
- Explore generation of open/reference datasets?
  - Leverage existing and future 5G assets of participating members
  - With what purpose?

NOKIA Bell Labs

TeamUp5G

New RAN TEchniques for 5G UltrA-dense Mobile networks

NOKIA

Copyright and confidentiality