# Insights into deploying Private 5G and outlook to making 6G happen

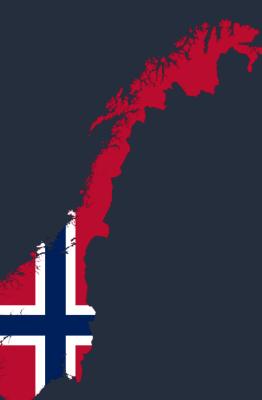
Dr. Dirk Trossen

Chief Scientist & VP R&D

**Norwaves Networks AS** 









# **Norwaves Strategic Direction**





Keep advancing through 6G (ex:EU Research projects)



**Private 5G – All About "Critical Services"** 





#### **Oslo Kommune**

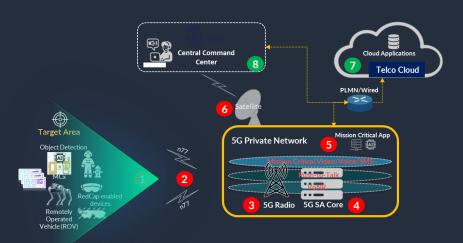
- **Outdoor and indoor trials of Mission Critical Services over a Private 5G Network for Oslo** Municipality, conducted on 17th May as part of the 'Krisekommunikasjon 5G - KRIKOM5G' project.
- **5G Use Cases:**

Rapidly deployable standalone 5G Resilient 5G and interreference free network

**Mission Critical** Voice/Video/group

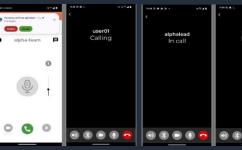
**Outdoor & Indoor Scenarios** 

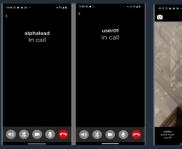
**Delivered using a standalone Private 5G** network with dedicated spectrum acquired from NKOM on behalf of Oslo Municipality.



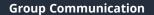
#### **MCX Voice Communication**

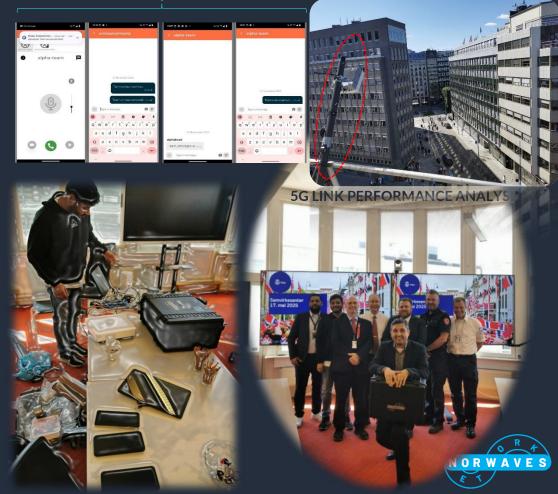
**MCX Video** 











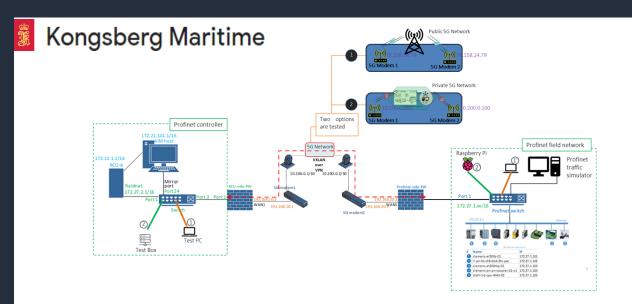
- **Deployed Private 5G Network as part of** the European Commission Project "FIDAL" at Kongsberg Maritime Headquarters, **Kongsberg**, Since May.2024
- **5G Use Cases:**

**Ship Onboard** communications (uRLLC/mMTC)

**Connecting Layer** 2 devices natively to 5G network

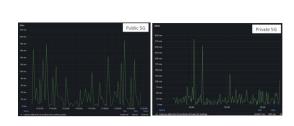
UE-UE communication **Data aggregation** UseCase

**Delivered using a Standalone Private 5G Network with Kongsberg Maritime's dedicated Private 5G Spectrum** 

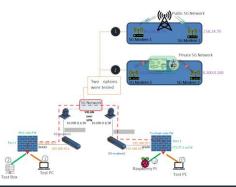


#### **5G PERFORMANCE IN PROFINET FIELD NETWORK**

#### **PUBLIC 5G VS PRIVATE 5G TESTS**

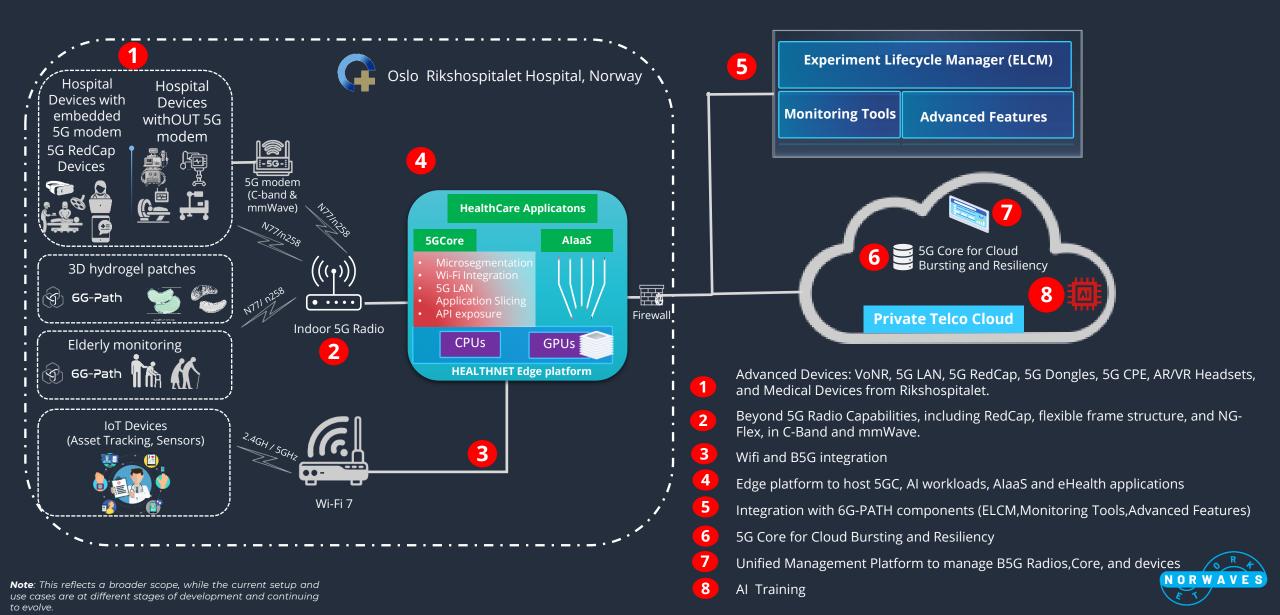


#### **5G LINK PERFORMANCE ANALYSES**





#### **Public Health**



# Capabilities



# Atomic Capabilities to Customize for Verticals and Scale...





Roadmap



**Breakout** 8

**Flexible** Uplink/Dow nlink Speeds 2

9 5G LAN,

**Ethernet** 

PDU,TSN

**Mission** Critical **Services** (MCX)

(10)

Native 5G **Voice & SMS** (VoNR, SMS)

3

Slicing /

**Quality Of** 

**Service** 

(QOS)

(11)

**Positioning** 

Al for **Industry** (Al Apps)

5G & Wi-Fi Integration

5

Telco Cloud/ Edge Computing

6

**Industrial** AR/VR **Applications** 

12

RedCap/ eRedCap 13

**Ambient IoT** (Passive-IoT)

TN-NTN Integration (Satellite)

14



#### **Enable Various Industries**





**Use Cases** 

- · AR-assisted surgeries
- Al-driven medical imaging

IoT-based ward monitoring





**Use Cases** 

- Ship/platform-to-shore Com
- Ship/platform-to-ship Com





**Use Cases** 

Command Control System Multicast-Broadcast

**Media Broadcasting** 









**Use Cases** 

- Remote 4K Video broadcast
- 5G Network-in-Box with backhaul



#### **Embodied in Norwaves 5G Network in a Box (NIB)**

Norwaves 5G Network in a Box (NIB) – a plug-and-play, all-in-one solution for rapid 5G deployment. Packed into a compact briefcase, it comes pre-integrated with a fully functional, high-performance 5G Radio Access Network (RAN), 5G Standalone (SA) Core, and 5G SIM-equipped devices, enabling swift, seamless, and hassle-free 5G network setup.

























Gaps – to be Covered by 6G?

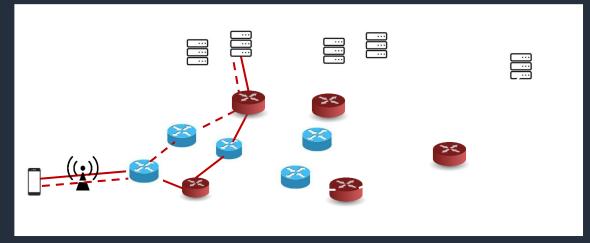


## **Routing on Service Addresses (ROSA)**

- DNS maps service addresses, e.g., URIs onto IP addresses
  - Even local DNS incurs about 10 to 15ms <u>latency</u> for every lookup
  - <u>Changes</u> in name->address assignment can only change longer-term
    - -> counter the idea of edge computing and several virtualised service instances
    - -> fallback to edge orchestration, which comes with complexity & latency challenges itself
  - Increasing <u>dependency</u> on non-EU players (Google, CloudFlare, Amazon, ...)

#### Enter ROSA

- Piggyback name resolution on first IP data packet exchange (usually a TLS handshake)
- Route first packet over ROSA overlay, which forwards based on <u>name entries</u>
  - Make pathfinding decision based on ingress scheduling or overlay routing
- Route subsequent packets directly over IPv6 after first packet was delivered to target IP
- Redo assignment at every flow start (e.g., TCP/QUIC connection)





# **Routing on Service Addresses (ROSA)**

#### **Capabilities**

- No resolution latency (overlay routing can be almost as flat as direct IPv6)
- MUCH faster path finding than name resolution (since core overlay router can handle >100k packets per second)
- Largely backwards compatible requires using user space library at endpoints (similar to QUIC)

#### **Scenarios**

- Smoother video streaming: replicate video at N instances, schedule each chunk request to instance in weighted fair scheduling (weight is proportional to instance processing power)
  - -> 30% reduced jitter, about 40% more users per server being served below latency boundary (IFIP Networking 2022)
- Faster web browing: place ingress overlay router at CDN ingress, place short tail of web objects into ROSA name space, rest via DNS
  - -> up to 60% faster page loading times, depending on ROSA object proportion
- Carbon-negative AI: place edge AI servers at renewable energy source, activated ONLY on surplus energy, route to edge servers only when active
  - -> target is carbon-negative AI processing (still to be shown)



#### Flexible core federation

- Private 5G is not just about localised processing and communication
- Private 5G is not just about interconnection via Internet
- Next Generation of P5G will be about building flexible federations of localised entities
  - Single patient network, consisting of hospital P5G, ambulance Network-on-Wheel, and patient home
  - Single industrial network, consisting of factory network(s), frontline workers, sales personnel, and customer equipment
  - Entertainment federation, consisting of entertainment venue (gaming hub), players home, real-life deployments in the field
  - •
- Need flexible interconnection at runtime that carries all service-critical functions across
  - Slicing, bandwidth allocations, service routes, ...

All this needs alignment with compliance, sovereignty, & privacy concerns



# The Elephant in the Room



# **Artificial Intelligence (AI)**

6G is all about Al-native, is it not?
-> Al for 6G

...but is it about supporting AI services natively?



## The Rise of the New Killer Application

#### A Deja-Vu

- The 90s saw a significant rise of video content delivery over the still young Internet with negative user experience in E2E delivery
- Content Delivery Networks (CDNs) were designed, implemented, and deployed at scale to bring video content to end users

Modern OTTs like Amazon, Netflix, YouTube and many others would not be possible without the existence of CDNs

- Al is the next killer application, like video in the 90s, but at MUCH larger scale
  - ChatGPT, launched just three years ago in 2022, and other LLM-based GPT services are following suit, with ChatGPT alone experiencing about 4.5 billion visits in March 2025

Context Network (transfer context) Compute (process contex (a) Fetching the text of context: Less data to send but higher computation delay KV cache Network (transfer KV) Compute (process prompt (b) Fetching KV cache of the context: Low computation delay but much more data to send Network (transfer compressed KV) Compute (decompress KV & (c) Our work: Load compressed KV (or text) to save both network delay and computation delay

[1] https://dl.acm.org/doi/pdf/10.1145/3651890.3672274

- Al inferencing can be accelerated through use of <u>KV caches</u> (SIGCOMM2024 [1]), enabling local
  inferencing instead of relying on the central LLM data centre
- Distributing KV caches is like streaming a video

6G and the Future Internet needs *Inferencing Delivery Networks* (IDN) to make AI a working reality in future networks

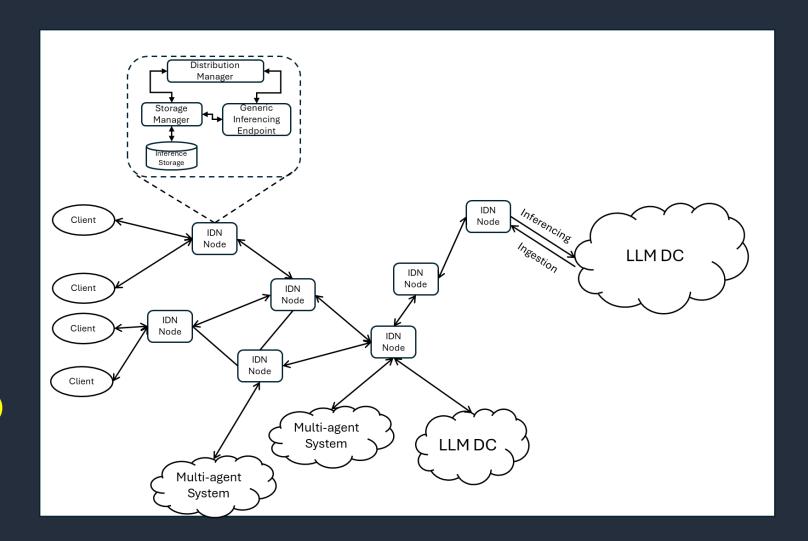
# **Challenges in Supporting AI Natively**

- Scale: Al is more than ultra-large GPTs, it is highly local, private, and distributed
  - -> Scale exceeds that of centralised content production and distribution of CDNs
- <u>Universal access</u>: Access to Al is a key capability of 6G (not just its use in building the system)
  - -> Must ensure that EVERYBODY has equal access and performance to it
  - -> Do not leave it to good traffic engineering of local networks!
- Deep 6G integration: Not just at the DN level but at UPF level
  - -> Solve the service and breakout challenges and you are mainly there!
- Contextuality & Locality: Al is deeply contextual as it is alive with new data & is deeply local in usage, not just in delivery but also relevance of data
  - -> Context (including but not limited to locality) plays a vital role in delivering but also refreshing Al
- <u>Freshness</u>: There is no global knowledge for Al but it lives from local data to be adjusted and refreshed
  - -> Native AI support needs to include refreshment of globally distributed knowledge
- **Privacy**: Al is deeply private in MANY parts, e.g., private industry
  - -> Native Al must support security through data perimeters, encryption, etc



# A Strawman for an Inference Delivery Network

- Structured like a CDN
  - Distributed
  - Decentralised
  - Regional
- Push/pull of global knowledge
  - Driven by localised context
- Localised operations
  - Processing (not just retrieval)
  - Contextualisation
  - Freshness





## Conclusions

5G has brought us a long way, enabling highly customised private networks

6G has to fill remaining gaps to make P5G an even more pervasive reality

BUT we must tackle the elephant in the room – AI – to not fall into the same trap as with video in the 90s



# THANKS FOR YOUR TIME

Get in Touch!

