Performance Analysis of Local 5G Operator Architectures for Industrial Internet

Yushan Siriwardhana, *Member, IEEE,* Pawani Porambage, *Member, IEEE,* Mika Ylianttila, *Senior Member, IEEE,* and Madhusanka Liyanage, *Senior Member, IEEE,*

***Abstract*—5G calls for a network architecture that ensures ultra-responsive and ultra-reliable communication links, in addition to the high degree of flexibility and customization required by different vertical sectors. The novel concept called local 5G networks enables a versatile set of stakeholders to operate 5G networks within their premises with guaranteed quality and reliability to complement Mobile Network Operators' (MNOs) offerings. In this paper, we propose a descriptive architecture for a local 5G operator which provides user specific and location specific services in a spatially confined environment i.e. industrial internet environment. In addition to that, we propose hybrid architecture options where both the local 5G operator and MNO collaboratively contribute to establishing the core network to cater to such communications. The architecture is discussed in terms of network functions and the operational units which entail the core and radio access networks in a smart factory environment which supports Industry 4.0 standards. Moreover, to realize the conceptual design, we provide simulation results for the latency measurements of the proposed architecture options with respect to an Augmented Reality (AR), massive wireless sensor networks and mobile robots use cases. Thereby we discuss the benefits of deploying core network functions locally to cater to specialized user requirements, rather than continuing with the conventional approach where only MNOs can deploy cellular networks.**

***Index Terms*—Local 5G Networks, Industrial Internet, Industry 4.0, 5G, Augmented Reality, Architecture**

I. Introduction

The landscape of mobile communication service requirements is rapidly changing with the proliferation of digitization technologies. Consequently, in the future, more emphasis needs to be placed on location specific services in different vertical sectors such as automotive, health, energy, industry and media [1]. Hospitals, shopping malls, smart cities, factories and universities are identified as some of the common locations, which are heavily benefited by these location specific services. Location specific requirements stipulate high demands on reliability, high data rates, low latency, privacy and security. The key focus of the future 5G wireless systems is to serve such case specific requirements along with the provisioning of the traditional mobile broadband services [2]. These case specific and localized requirements are expanding beyond the current capabilities of the traditional MNOs whose services are often designed to serve masses. To serve the location specific future communication requirements, the need for establishing local 5G networks is evident. In speeding up local service delivery with 5G networks, the present mobile communication market needs to be opened for local 5G networks deployed by different stakeholders such as recently proposed in the micro operator (uO) concept [3].

Unlike the traditional MNOs with wide area coverage, uOs are local operators who intend to offer case specific and location specific services through locally deployed 5G networks [4], [5] . Therefore, the system architecture for a local 5G operator (L5GO) should be carefully designed in such a way that it enables efficient and reliable local service deliveries. The local operator must deliver 5G services and the system architecture of a L5GO must contain the network functions defined by the 3rd Generation Partnership Project (3GPP). Since these L5GOs provide tailored services, the system architecture and its specific deployment may also depend on the use case.

A. Our Contribution

Besides the novelty of L5GO concept, very few efforts have been taken to define the system architecture for a L5GO. A network architecture for emerging 5G uO, which serves one use case (i.e. Augmented Reality) in a smart factory environment has been proposed [6]. This paper extends the work for two other industrial use cases called massive wireless sensor networks and mobile robots [7]. The proposed architecture comprises 5G network functions and the operational units which entail the core and the access networks to serve the communications of the use cases. Apart from defining a pure local architecture, this paper considers different variants of hybrid architectures where both L5GO and MNO collaboratively deploy network functions to establish the core network. To realize the conceptual design and present the simulation results, we compare three network deployment models for a factory, first being served solely by a L5GO, second being served by a traditional MNO and third being served by an operator having a hybrid architecture. In a hybrid architecture, certain core network functions are deployed locally while other core network functions are deployed by MNO. Based on the simulation results, we discuss the benefits of locating the core network functions closer to user locations to serve specialized user requirements, rather than continuing with the traditional MNO driven approach. Moreover, we conduct a testbed implementation to compare and analyze the real world

Yushan Siriwardhana, Pawani Porambage and Mika Ylinattila were with the Centre for Wireless Communications, University of Oulu, email: firstname.lastname@oulu.fi.

Madhusanka Liyanage was with the School of Computer Science, University Collage Dublin, Ireland and Centre for Wireless Communications, University of Oulu, Finland, email: madhusanka@ucd.ie, madhusanka.liyanage@oulu.fi.

behavior of a L5GO, hybrid operator and a MNO using the performance metrics latency and throughput.

B. Paper Organization

The remainder of the paper is organized as follows: Section II describes the related work on 5G architecture, L5GOs, industry 4.0 and network slicing. Section III describes the selected industrial use cases and Section IV explains the proposed architecture for L5GO which serves the industrial use cases. Section V illustrates the different architectural options and different algorithms used in deriving those architectural options. Section VI explains the approach to simulations and presents the key parameters. Section VII discusses the simulation results for a typical MNO setting and proposed local and hybrid architectures. Finally, Section VIII concludes the paper with the future research directions.

II. Background and Related Work

A. Future 5G Networks and 5G Network Architecture

5G will be a paradigm shift from present wireless communication technologies and it will also be highly integrative to provide universal high-rate coverage and a seamless user experience [8]. Key characteristics of 5G wireless systems are identified as extremely high data rates, ultra-reliability and low latency, and massive communication between devices [9]. Three main 5G service classes are as enhanced Mobile BroadBand (eMBB), Ultra Reliable and Low Latency Communication (URLLC) and massive Machine Type Communication (mMTC), [10], [11]. Based on the communication needs of different verticals, future 5G operators must possess the capability of providing case specific services in addition to the present generic communications services [1].

3GPP has released the specifications for 5G system architecture [12] and the network architecture of a L5GO should also include the same Network Functions (NF) defined in generic 5G architecture [3]. Instead of the network elements defined in Evolved Packet Core (EPC) in 4G systems, Software Defined Networking (SDN) [13] and Network Function Virtualization (NFV) are involved in creating Network Functions (NF) in 5G systems architecture. Network functions can be implemented on a dedicated hardware or as a software instance on a dedicated hardware or as a virtualized function instantiated on an appropriate platform such as a cloud. The concept of network functions has led operators to add flexibility over the functionality of the underlying physical infrastructure of the 5G network. 3GPP specifications represent the 5G architecture in two ways:

- **Service based representation:** Shows how NFs within the control plane enable other authorized NFs to access their services, as in Figure 1.
- **Reference point representation:** Shows the point-to-point interaction existing between two NFs, as depicted in Figure 2.

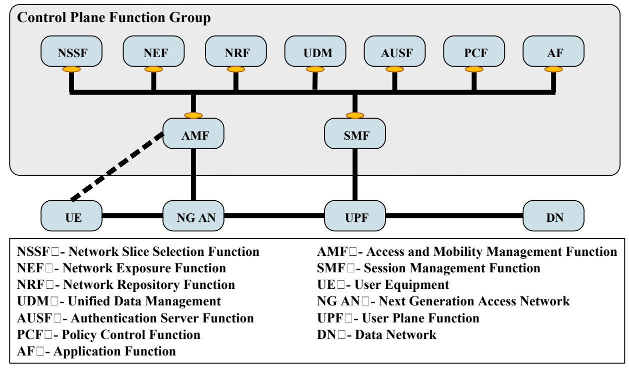


Fig. 1: Service based representation of 5G architecture [12]

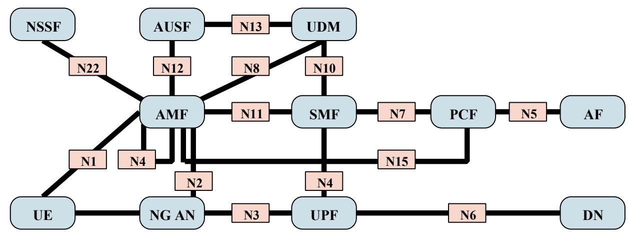


Fig. 2: Reference point representation of 5G architecture [12]

B. Local 5G Networks

Currently, local 5G networks are gaining increasing attention because the mobile communication market can be expanded to new stakeholders and allowing them to deploy local 5G networks to complement conventional MNOs. These L5GOs such as uOs are expected to provide tailored 5G services and fulfill case specific and versatile local wireless communication needs with extremely low latency [3]. A L5GO can respond rapidly to future communication requirements than a MNO as it provides case specific, location specific needs while MNO's main focus is serving to masses [14]. Moreover, it will be operationally difficult for a MNO to use existing macro cells to serve these requirements because most of the traffic will be originated from indoors. L5GOs have a flexible design in which they can operate a closed network to serve its own customers, an open network to offer its services to other MNO's customers, or a mix of both. Key regulatory elements and the techno economic aspects related to the uOs are discussed in [4]. Business model options for local 5G uOs and the different network deployment options are discussed in [14].

C. Industrial Internet

Industry 4.0 or smart factory environment is one particular location which will be highly benefited by the services of L5GOs. Industry 4.0 refers to the advancement of the present industries into the next generation [15], [16]. It aims to interconnect the devices inside the factories, make them smart by adding more intelligence into the device and ultimately resulting in improved adaptability, resource efficiency, and the supply and demand process between factories [17]. In industrial environments, Machine-to-Machine (M2M) communication plays a critical role, especially with the deployment of sensor networks and Automated Guided Vehicles (AGV).

Wireless Sensor Networks (WSN) in current industries are moving towards industrial wireless networks because of the low latency, high mobility and high capacity requirements in the future industries [18]. A study report has been released by 3GPP focusing on typical use cases in Industry 4.0 such as motion control, mobile robots, augmented reality, massive wireless sensor networks [7].

D. Network Slicing

Use cases of future 5G communications will demand diverse and sometimes extreme requirements [19]. Serving the needs of these diverse requirements using a monolithic network infrastructure will not be efficient, therefore the need for flexible and scalable networks to serve these requirements is a must. Also, having such flexible and scalable networks will make the introduction of new services much easier. This can be achieved using network slicing.

Network slicing sub-divides a network into logically isolated sub-networks. These logical networks enable different types of communications on a common infrastructure [20]. With network slicing, network usage can be optimized by serving different types of communication via different slices than serving them with just one network. 3GPP introduces three network slice management functions for creating and managing network slices in 5G networks called Communication Service Management Function (CSMF), Network Slice Management Function (NSMF) and Network Slice Subnet Management Function (NSSMF) [21], [22].

- **CSMF:** Responsible for translating communication service related requirements to network slice related requirements.
- **NSMF:** Derive network slice subnet related requirements from network slice related requirements and, responsible for management and orchestration of Network Slice Instances (NSI).
- **NSSMF:** Responsible for management and orchestration of Network Slice Subnet Instances (NSSI).

In addition to the slice management functions, 3GPP introduces Network Slice Selection Function (NSSF) to select the appropriate network slice for a given communication.

III. Use Cases

3GPP Study on communication for automation in vertical domains [7] describes a number of use cases which will appear in the factories of the future. To define the network architecture of the L5GO, we consider three use cases from the 3GPP study which will frequently be seen in future industrial environments. The use cases are selected in a way that they cover all three generic 5G services i.e. eMBB, URLLC and mMTC. The selected use cases are Augmented Reality (AR), massive wireless sensor networks and mobile robots as outlined in Table I.

As with any IoT device, power management of the deivces is a critical concern for these use cases, especially at the smart factory environments where massive number of such devices will be used. L5GO deployments will primarily be with small cell Ultra-Dense Networks (UDN), hence the user devices are

TABLE I: Industry 4.0 Use Cases

Use Case	eMBB	URLLC	mMTC
Augmented reality	✓	✓	×
Massive wireless sensor networks	×	×	✓
Mobile robots	×	✓	✓

in close proximity to the base station. This helps to reduce the transmit power of the devices such as AR devices, sensors, and mobile robots. Terminal devices such as wearable AR devices may not be able to perform complex processing within the device, thus requiring computational offloading. The improved coverage provided by UDN will support high throughput transmissions required by the use case for offloading so that the devices only perform mandatory processing, leading to extend the battery life of the devices.

A. Augmented Reality

In AR use case, factory workers are supported by AR devices and these devices are used to identify production flaws, obtain step by step guidance to carry out pre-defined tasks, obtain support from the supervisors. In this context, AR devices should be highly energy efficient and lightweight. This requires AR devices to carry out minimal processing and more intensive tasks to be offloaded to a separate image processing server located inside the factory. Typical communication between AR device and the image processing server of an Industry 4.0 AR network is depicted in Figure 3.

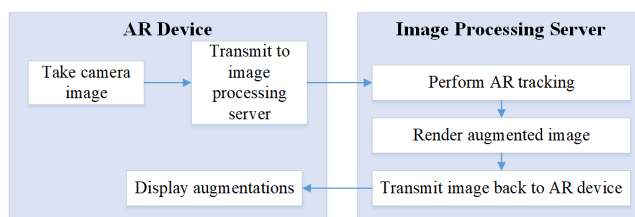


Fig. 3: AR system model with offloaded processing [7]

AR device takes the images and transmits them to the image processing server. Server then performs the processing of the images and sends the augmentations back to AR device to display. The video transfer from the AR device to the image processing server needs the 5G system to support higher bandwidth. Moreover, 5G system supporting this communication should be able to provide an End-to-End (E2E) latency of less than 10 ms for one-way communication with a 99.9% success of frame delivery [7]. A typical AR network along with the core and access networks of the served operator is depicted in Figure 4.

B. Massive wireless sensor networks

Massive wireless sensor networks will be used for monitoring the working environment in future industrial environments. Sensors can monitor various types of parameters such as pressure, humidity, temperature, CO_2 and sound. The main purposes of having a sensor network is to monitor the environment, detect malfunctions in the surrounding, take appropriate

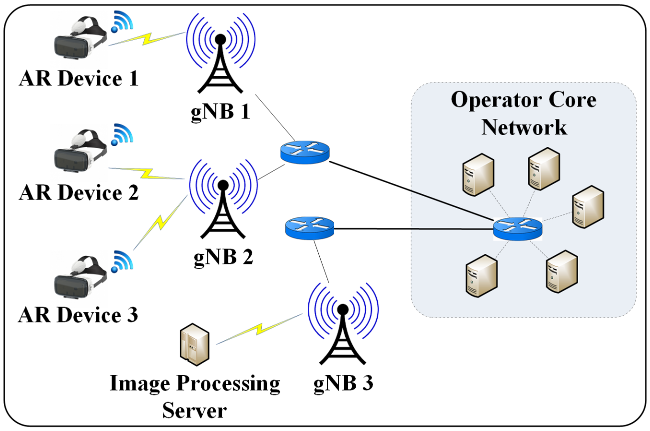


Fig. 4: An AR network with terminal devices and the AR server

actions by a decision-making entity so that the effect from the malfunction is mitigated. As an example, when an anomaly of the room temperature in detected, a machine can be triggered to its emergency stop. Placement of the monitoring function is a key design aspect of sensor networks which directly affects the latency. In case of less computationally complex sensor nodes, this functionality can be placed in a central cloud server. At the same time, the functionality can be placed inside the factory environment to support low latency requirements. A typical industrial wireless sensor network along with the core and access networks of the served operator is depicted in Figure 5.

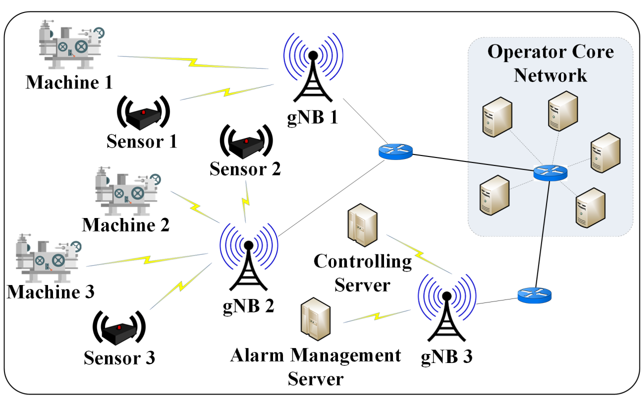


Fig. 5: Wireless sensor network with management and controlling servers

In our study, we consider that the sensors are transmitting the sensed data to a centralized management server, which decides any controlling action based on the sensed values. Management server then instructs the controlling entity and the controlling entity takes the appropriate mitigating action. The latency required for event based condition monitoring such as a detecting a high temperature inside the factory premises is from 50 ms to 1 s. Interval based condition monitoring such as humidity measurements taken every 1 hour also requires latency varying from 50 ms to 1 s. The highest priority is given for condition monitoring for safety, needs the action to be taken within 5 - 10 ms [7].

C. Mobile robots

In future industrial environments, mobile robots such as Automated Guided Vehicles (AGV) will be used in numerous applications and will play an extremely important role. These robots can be programmed to execute multiple operations fulfilling number of tasks such as transporting goods and providing assistance to workers (collaborative robots or cobots). These robots can sense and react with their environment, therefore they operate more intelligently than the traditional machines programmed to travel along the pre-defined paths.

For the proper operation of mobile robots in an industrial environment, they should be monitored and controlled by a guidance control system. This will avoid collisions between robots, assign driving jobs and manage the traffic of mobile robots. In a smart factory environment, robots can guide themselves using their own sensors such as cameras and lasers. In our study, we consider the robots and the guidance controller are connected to each other using 5G. Figure 6 depicts the connectivity between the robots and the guidance control server.

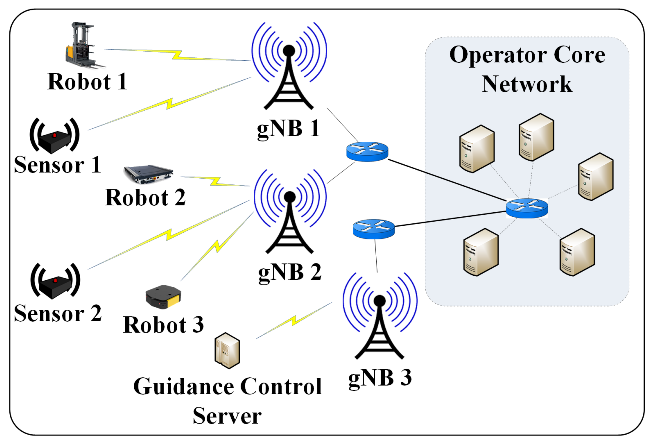


Fig. 6: Mobile robots controlled by a guidance control server

Communications of mobile robots can be categorized into three main cases based on the fact that who is communicating to whom [7]. They are, communication between mobile robots and guidance control system, communication between mobile robots, communication between mobile robots and peripheral facilities. Very stringent latency is a key requirement for mobile robots operating in an industrial environment. A cycle time of 1 ms to 10 ms is required for machine control, 10 ms to 50 ms for cooperative driving of robots, 10 ms to 100 ms for video operated remote control and 40 ms to 500 ms for standard mobile robot operation [7].

IV. Proposed Architecture

A local 5G network covering the factory could be used to address the needs of the three use cases. This local 5G network should comprise architectural components inherited from generic 5G systems architecture. The concept of locally deploying the network provides flexibility over the selection of architectural components and the location where the core network is hosted. For a low latency requirement, the desirable implementation is to have the core network within the factory premises itself, but not mandatory.

In our study, we define the architectural components needed in the core network to serve each use case separately and then derive the final architecture which can simultaneously serve all three use cases.

A. Network Architecture for AR communication

Generally, AR use case requires the 5G system facilitate the following three steps of communications.

- Registering the AR devices into the network
- Establishing data session between the AR device and image processing server
- Data transfer between AR device and image processing server

Architectural components needed for completing above steps can be identified based on the message transfer between each element in 5G system including AR device, Next Generation NodeB (gNB) and core network functions.

a) Device registration: We define the registration procedure for AR device based on 3GPP specifications [23]. Figure 7 illustrates the message sequence between the entities in the architecture. AR device initiates the registration process by sending registration request to gNB. gNB forwards the request to AMF. After that, AMF and AR device exchange the identity request and response messages. In the next step, AMF contacts AUSF for the device authentication. AUSF facilitated the authentication after contacting UDM and retrieving the authentication data. Once the authentication data is received from UDM, AUSF sends the authentication response to AUSF. Identity request/response messages are transmitted between AMF and the AR device again. After the identity verification, AMF then works with PCF for the policy association for the AR device. Once the policy association is successful, AMF sends an update to SMF informing the session context. AMF also sends the registration accept message to the AR device and the device then sends the registration complete message to AMF concluding the registration process.

b) Session establishment: After completing the registration process, AR device has to establish a data session with the image processing server to enable continuous data transfer. We define the Protocol Data Unit (PDU) session establishment procedure between AR device and the image processing server based on 3GPP specifications [23]. Figure 8 illustrates the message sequence required for PDU session establishment process.

Here, AR device initiates the process by sending PDU session establishment request to AMF via gNB. AMF then sends a request for a new session creation to SMF. In the next step, SMF registers with UDM, subsequently UDM stores data related to the session. After that SMF sends the Response to AMF. Then, PDU session authentication/authorization process occurs by exchanging messages between AR device, gNB, AMF, SMF, UPF and Server. Once this step is completed, SMF works with PCF for policy association for the session. Then UPF and SMF exchange the session establishment/modification request and the respective response. Message transfer from SMF to AMF allows AMF to know which

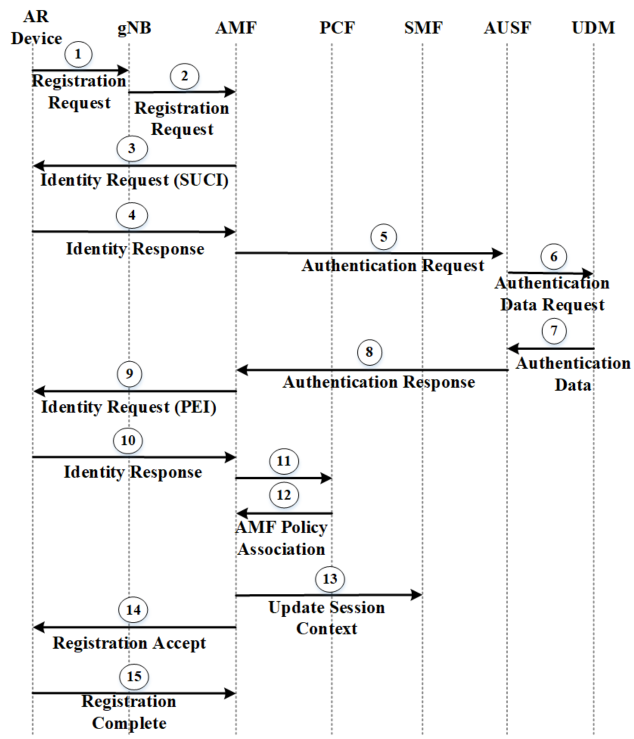


Fig. 7: Message sequence chart for AR device registration procedure

access towards the AR device to use. AMF then sends the PDU session ID information to gNB so that gNB can work with AR device for the gNB specific resource setup. After that gNB sends the acknowledgement for the PDU session request to AMF. Based on that, AMF sends request regarding PDU session update to SMF. SMF then requests UPF for session modification. Once SMF received the response from UPF, SMF finally sends the response for PDU Session update to AMF completing the PDU session establishment process.

c) Data transfer: After successful completion of above steps, AR device can send a continuous data stream to the server and retrieve the augmentations sent by the server, so that the device can overlay them on the camera view. Entities involved in this data transfer process are AR device, gNB, UPF and the server. Based on the above steps, 5G network functions needed to serve the AR use case can be identified to derive uO architecture depicted in Figure 9.

Network functions that are not used in the architecture are Network Exposure Function (NEF) which handles the masking of network and user sensitive information to external Application Function's (AF) according to the network policy, Network Repository Function (NRF) and AF.

B. Network architecture for massive wireless sensor networks

The following steps of communications should be supported by 5G network operator to serve the massive wireless sensor networks use case.

- Registering the sensors, actuators and servers in to the 5G network

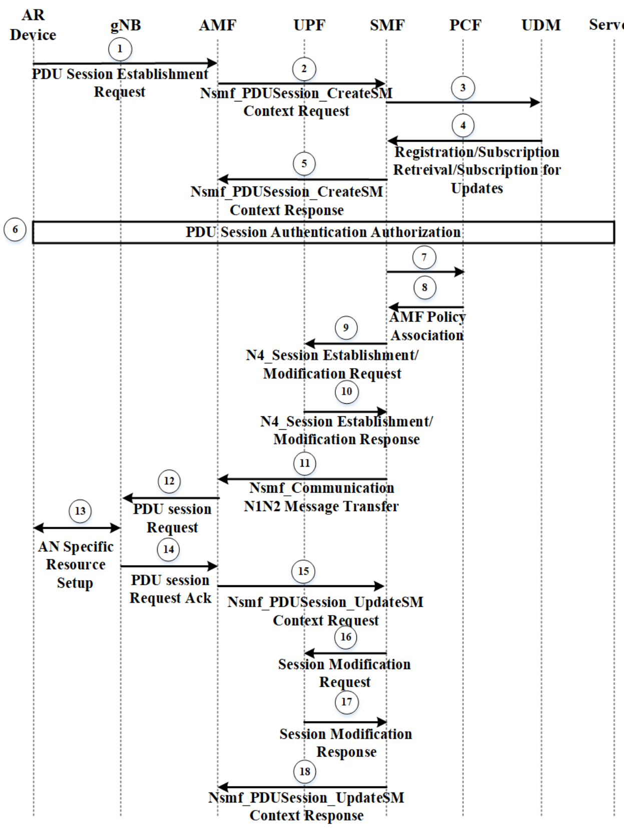


Fig. 8: Message sequence chart for session establishment between AR device and server

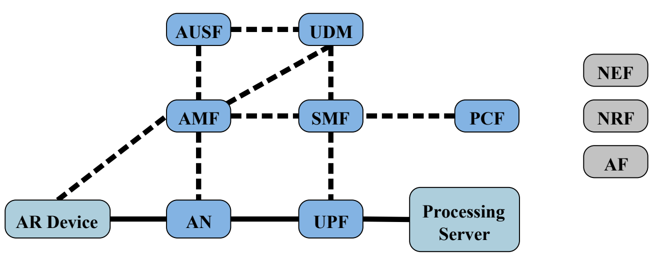


Fig. 9: Architectural components of L5GO to serve the AR use case

- In case of anomaly, establishing data connectivity from the sensor devices to the alarm management server
- Transfer the data related to anomaly from sensor to the management server
- Session establishment between the management server and the controlling instance
- Data transfer from management server to the controlling server
- Session establishment between the controlling server and the actuator
- Taking appropriate action on the actuator by transferring action related data to the actuator

We define the message sequence for the sensor network communication scenario using 3GPP specifications [23]. Figure 10 illustrates the message sequence diagram for a typical sensor network communication. Registration process uses the comprehensive message sequence illustrated in Figure 7, where the terminal devices in this case being the sensors, machines or servers. Then the sensor detects and abnormal behaviour and communicates it to the alarm management server. For this to happen, the sensor requires a PDU session to be established with the alarm management server. This happens according to the message sequence illustrated in Figure 8, while sensor being the initiator in this case. Once the PDU session is established, sensor transfers the data to the alarm management server. The server then analyses the severity of the anomaly and establishes a PDU connection with the controlling server, again using a similar message sequence illustrated in Figure 8. Once this is done, alarm management server successfully transfers the data to the controlling server. Controlling server decides which action to take on which machine by processing the received data. Then the controlling server establishes a PDU session with the relevant machine. Next, controlling server transfers the data to the relevant machine completing the communication process. Relevant action will be executed on the machine after this process. The action could be an emergency stop of a machine, decreasing or increasing temperature of an air conditioner, sound an alarm.

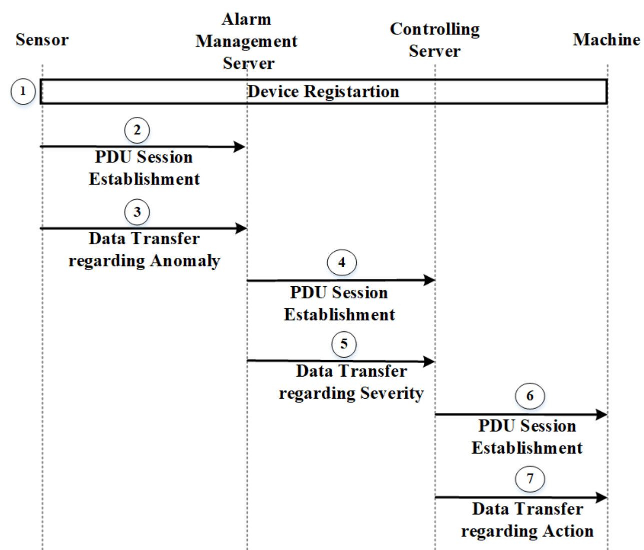


Fig. 10: Message sequence chart for emergency action based on sensor detection

Using the message sequence depicted in Figure 10, we designed the architecture for the L5GO catering to the needs of sensor network communication. For the device registration process AN, AMF, PCF, AUSF and UDM network functions are needed. For the PDU session establishment process AN, AMF, UPF, SMF, PCF and UDM network functions are needed. For the data transfer process AN and UPF network functions are used. Therefore, the architecture should comprise all the above mentioned network functions to support the communication. Hence, we can use the network architecture depicted in Figure 11 to support the sensor network communication, which is the similar to the one we used for AR communication.

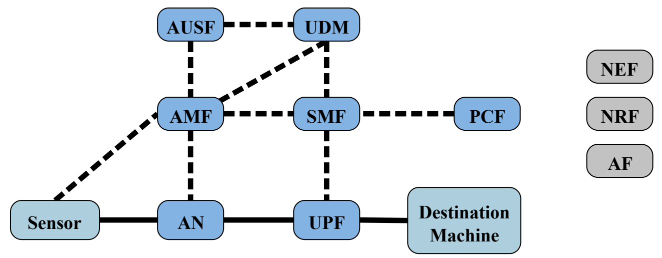


Fig. 11: Architectural components of L5GO to serve the massive wireless sensor networks use case

C. Network architecture for Mobile robots

With the increased usage of mobile robots in future factories, the robots themselves should have the ability to take decisions on their own by analyzing the data they receive from the sensors in the environment, the data they receive from the other robots and the data they receive from the guidance control system. Following communication steps will be seen in Industry 4.0 mobile robots use case.

- Registering the environment sensors, mobile robots and server to the 5G network
- Establishing data connectivity between elements such as mobile robots, guidance control server and sensors
- Robots' regular operation which includes data transfer between different elements
- Perform handovers due to the mobility of robots

Regular operations of mobile robots require registration in the 5G network and become the network elements of 5G, establishing PDU session between the relevant devices, data transfer between the devices and seamless mobility. We already identified the network functions needed for registration and session establishment under previous two use cases in Figure 7 and 8. Hence, in this case also, it is mandatory to have AN, AMF, PCF, AUSF, UDM, UPF and SMF.

We define the message sequence for seamless mobility based on the handover process in 3GPP specifications [23] as depicted in Figure 12. The handover process execution is initiated by the source gNB forwarding the data to the target gNB. Target gNB then sends path switch request to AMF. Once AMF receives the path switch request, AMF sends a session update request to SMF and then SMF sends the session modification request to UPF. UPF then sends the session modification response to back to SMF. In the meantime, UPF sends the end marker packets to source gNB and it also sends the downlink packets to mobile robot via target gNB. After that SMF sends the PDU session update response to AMF, and AMF then acknowledges the path switch request to target gNB. Then the target gNB sends release resources message to source gNB completing the handover process. Therefore the mandatory NFs to cater to handover process are AN, AMF, SMF and UPF. Hence, we can use the network architecture depicted in Figure 13 to support the communication of mobile robots.

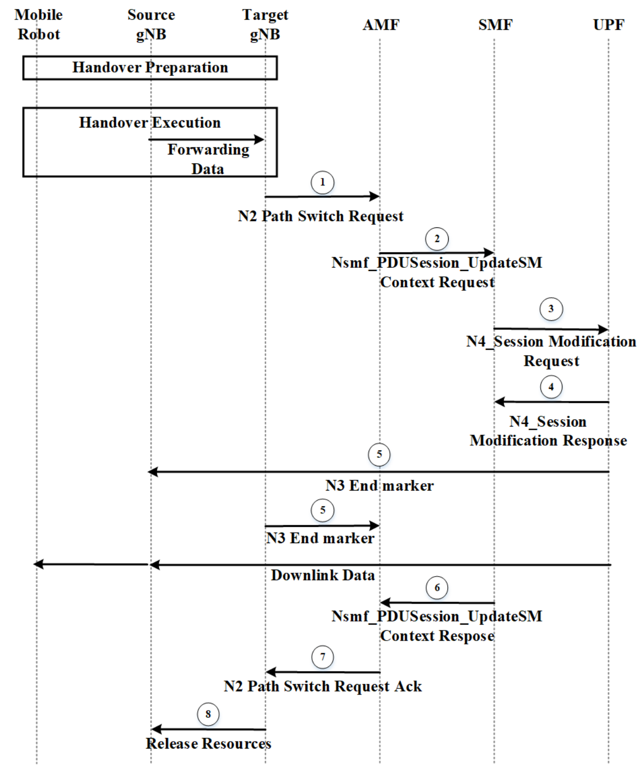


Fig. 12: Message sequence chart for handover procedure of robots

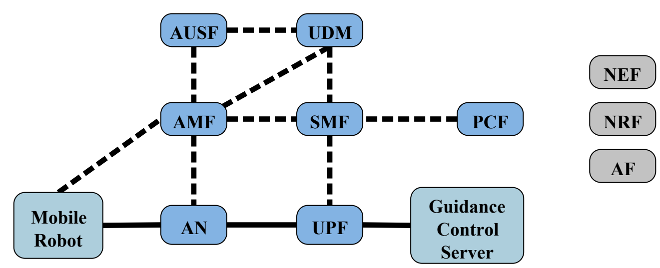


Fig. 13: Architectural components of L5GO to serve the mobile robots use case

D. Consolidated Architecture

When L5GO uses three network slices to cater to the three use cases, it must create network slices before any actual communication happens over a selected slice. To simultaneously cater to AR, massive wireless sensor network and mobile robots use cases, the L5GO needs to create NSIs before the communication begins. To create the network slices, three network slice management functions CSMF, NSMF and NSSMF should also be there in the architecture of L5GO. These three network functions allow the communication service requirements to be translated to network slice requirements, and then those network slice requirements to be translated to network slice subnet requirements and ultimately the network slice subnets can be created. These network slice subnets are then used to create the network slice instances.

Apart from that, the best fitting slice must be selected before the communication begins. This is done by NSSF, which is an obligatory element in the architecture of L5GO

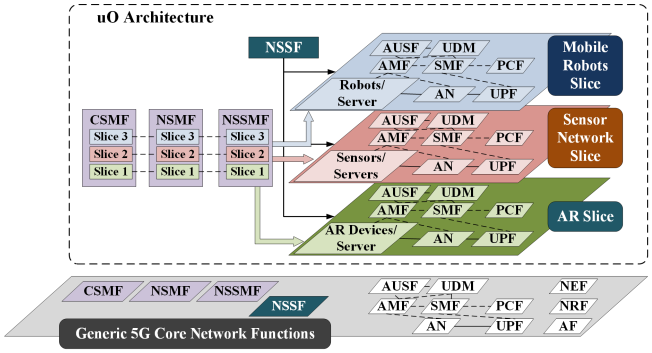


Fig. 14: Proposed architecture of L5GO with three network slices

which supports communication over multiple network slices. Combining these concepts, final architectural components to support AR, massive wireless sensor network and mobile robots use cases can be derived. Figure 14 depicts the derived architectural components for L5GO.

V. Architectural Options

In our architecture, we consider that the factory owns all the end devices such as AR devices, sensors, robots and servers. All of them are located within factory premises. It is assumed that the servers are located in a different cell site within the factory premises away from other end devices. We consider several architectural options for the 5G network operator(s) who serve the communication needs of the use cases.

A. Pure local architecture

First one is the pure local architecture where all NFs are deployed by the L5GO inside the factory premises. This includes both the core network which includes all NFs and the access network. This is direct use of the proposed architecture in Section IV. The architecture comprising the core network of L5GO, access network, backhaul connectivity and the terminal devices is depicted in Figure 15a.

B. MNO architecture

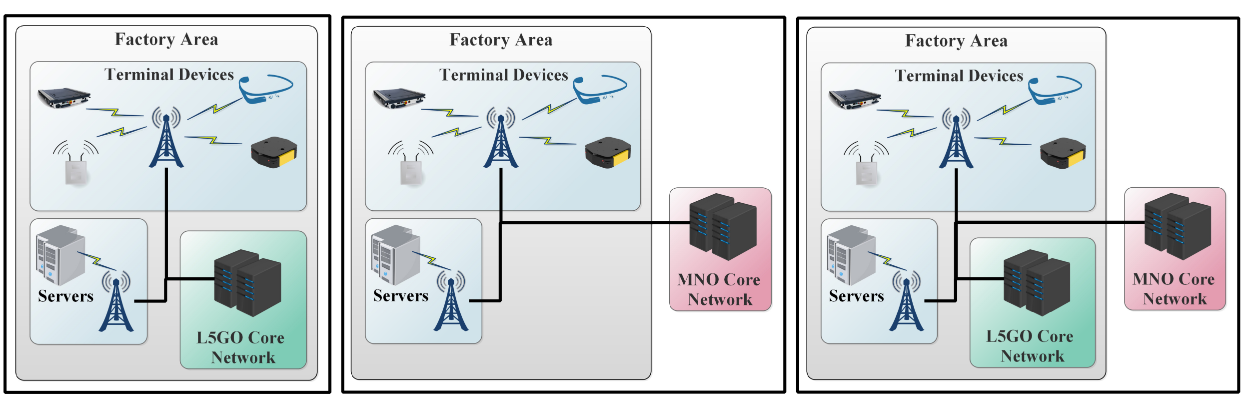
Secondly, we consider MNO architecture where all core NFs are deployed by MNO and they are located outside the factory premises at a given distance. The backhaul connects the factory location and the core network location of the MNO. This is depicted in Figure 15b.

C. Hybrid architecture

Finally, we consider hybrid architecture options where the 5G communication network is collaboratively deployed by both L5GO and MNO as depicted in Figure 15c. Both L5GO and MNO serve the factory communications using their core networks. A set of NFs are provided by L5GO and the remaining NFs belong to MNO. We consider that L5GO is limited in resources and can have only a limited number of NFs deployed at a given time instant. Assuming L5GO is the operator with the highest priority, the first M NFs needed in the communication are deployed by L5GO which is located on factory premises and all the remaining NFs are deployed by MNO. Here, M is the number of NFs deployed at L5GO. Therefore, the communication may also need the involvement of a NFs deployed by MNO based on the value of M. It is evident that the hybrid architecture approaches MNO architecture when M = 0 and it approaches pure local architecture when $M = M_{max}$. If 0 < M < M_{max}, then it becomes a hybrid architecture and in that case, messages between core network functions may also utilize the backbone network to travel from L5GO to MNO and vice versa, incurring a higher latency.

We consider different design approaches for these hybrid architecture deployments based on the idea of which network functions are provided by which operator. One of our architecture design approaches are based on NF placement algorithms proposed in [24], which is First Come First Served (FCFS) based allocation by L5GO. Secondly, we consider Most common NFs at L5GO (MCNF) and our third approach is Operator Policy Based Placement (OPBP) of NFs. Finally, we consider a Predictive Placement (PP) Algorithm similar to the one proposed in [25].

1) First Come First Served algorithm (FCFS): In this design approach, outlined in Algorithm 1, NFs are deployed



(a) Pure L5GO (b) MNO (c) Hybrid

Fig. 15: Architectural Options

at L5GO using the basis of when they are first used. If a certain communication happens first, then the NFs needed for the communication are deployed first. Based on the value of M, number of NFs deployed at L5GO is decided and the remaining NFs are deployed at MNO. The communication procedures involve the device registration procedure (reg), session establishment procedure (ses), data transfer (dat) and handover procedure (hnd). Each procedure contains multiple message transfers and each message has to be considered in identifying the NFs needed for communications. As an example, the device registration usually happens first, therefore NFs needed for message transfers in registration procedure will be deployed first.

Algorithm 1: FCFS NF Placement

```
Result: Order of NFs to be deployed at L5GO
Procedure List = {reg, ses, dat, hnd};
for Proc in Procedure List do
    for each Message in Proc do
        Identify NF needed for the Message transfer;
        if NF is already identified then
            Discard NF and move to next Message;
        end
    end
end
```

2) Most Common NFs at L5GO (MCNF): In the second design approach outlined in Algorithm 2 we place the most common NFs to all use cases in L5GO core network while placing the remaining NFs in MNO core network, irrespective of the amount traffic flow. This is decided after considering the message sequence charts for all use cases. As an example for M = 1, the most common NF is SMF and it is deployed by L5GO and other NFs are deployed by MNO. When the number M increases, more NF will be deployed by L5GO.

Algorithm 2: MCNF Placement

```
Result: Order of NFs to be deployed at L5GO
Procedure List = {reg, ses, dat, hnd};
for Proc in Procedure List do
    for each Message in Proc do
        Calculate Instance Count against each NF;
    end
end
Arrange Instance Count in Descending order;
Define Order of NFs;
```

3) Operator Policy Based Placement (OPBP): In the third algorithm, we consider the placement of NFs based solely on network operator policies. This means that the decision on where to deploy each network function is taken by the operator based on the external facts. For example, MNO may decide that certain core network functions should not be deployed outside MNO premises due to security reasons. Another example could be to deploy most heavily loaded NFs locally to reduce congestion at the infrastructure links and the MNO core network. Deployment cost of NFs at different places would also be another factor. In the policy based architecture, we take that the first NF to be deployed at L5GO is UPF because it is the most used NF while AMF and SMF are deployed as the last option because MNO prefers for a centralized control functions. Then we take UDM, PCF and AUSF based on their usage. Hence in our policy based algorithm the order of NFs would be UPF, UDM, PCF, AUSF, AMF and SMF.

4) Predictive Placement algorithm (PP): In the predictive placement depicted in Algorithm 3, we consider a given period and for that period we identify the past usage of the NFs. The period could be hours, days or even a longer duration. Based on the history data, allocate NFs to L5GO in such a way that minimum latency is achieved with the minimum number of NFs at L5GO. This algorithm considers the usage of a particular NF as well as the frequent interactions between NFs to come up with an optimized result. This way, most of the traffic towards the core network is handled by the L5GO ensuring less traffic transfer to the core network of MNO. The allocation is re-defined after every period and the present usage records can be used in the algorithm to take decisions for future days.

Algorithm 3: Predictive Placement of NFs

```
Result: Order of NFs to be deployed at L5GO
NF List = {UPF, AMF, SMF, UDM, PCF, AUSF};
M = 1, M_max = NF count in NF List;
for M < M_max do
    for each NF in NF List do
        Calculate avg. Latency using past Data;
        Find NF with Latency_min;
    end
    Identify M^th NF to Deploy using Latency_min;
    Remove NF with Latency_min from NF List;
end
```

VI. Numerical Analysis

We consider three deployment models in our simulations which uses the architectures explained in Section V. In the first model, the factory is served by a local 5G network having pure local architecture. Communication is facilitated by a 5G network deployed inside factory premises including the core network. This setup is depicted in Figure 16.

In the second model, we assume that the entire factory is covered by a 5G network deployed by an MNO. We consider that the MNO is simultaneously serving a total of N such factories having similar use cases. Each factory having similar network setup and similar requirements as seen in Figure 17. Core network of MNO is located outside the factory. Figure 18 illustrates the MNO based model, serving for the use cases of a given factory.

Without loss of generality, we have assumed that L5GO's processing power is 1/N of the possessing power of an MNO. It validates the fact that MNO's resources are equally divided among N L5GOs serving each factory.

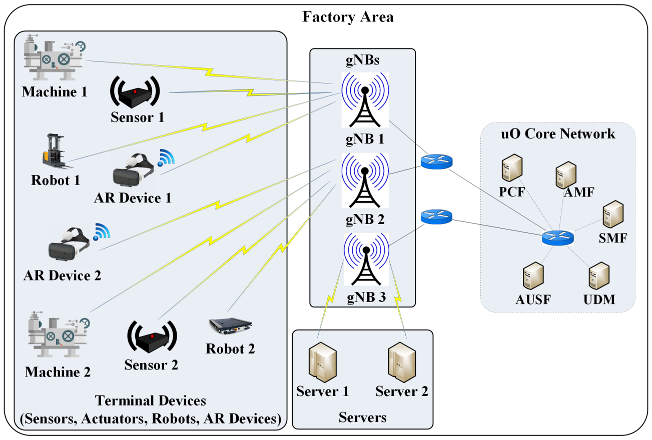


Fig. 16: Deployment model for pure local operator serving the factory use cases

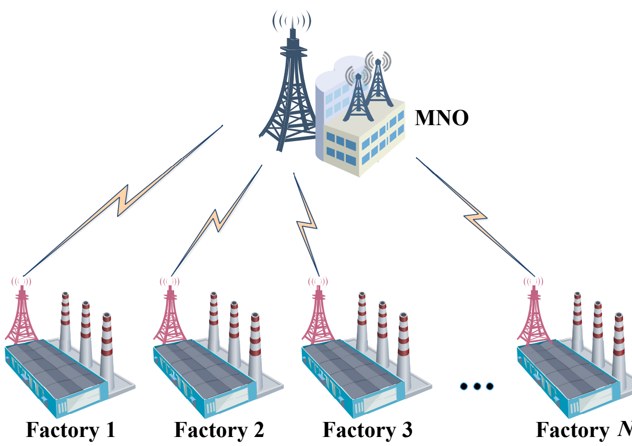


Fig. 17: MNO's service for N factories having the three use case

The third model represents the hybrid architecture where both L5GO and MNO simultaneously contribute to form the architecture, which is depicted in Figure 19.

For all pure local, MNO and hybrid models, we simulate the entire factory operation comprising of all three use cases AR, massive wireless sensor networks and mobile robots. We assume all three use cases are simultaneously occurring. All the simulations are carried out using MATLAB [26] and the latency results are averaged over 100 iterations. Simulations are carried out for a single day period of factory operation. To model a real factory operation, we assumed the frequency of occurrence of each use case during a simulation period. If the use cases are served by L5GO, then there is no external traffic from other factories on the core NFs but in the case of MNO, the NFs are loaded with the traffic from other $N-1$ factories. In the hybrid architecture, only the NFs located at MNO are subjected to external traffic. Since we have a slice based architecture for each use case, the traffic of one use case is not interfered by the traffic of other two use cases.

The average latency of the system comprising all use cases can be expressed as follows.

$$L_{avg} = \frac{L_{tot}}{P_{tot}} \tag{1}$$

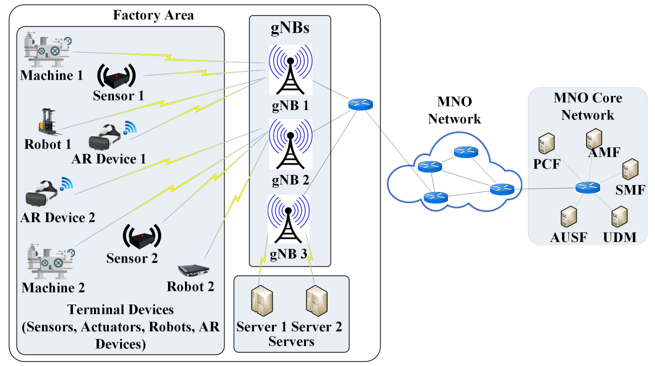


Fig. 18: Deployment model for MNO serving the factory use cases

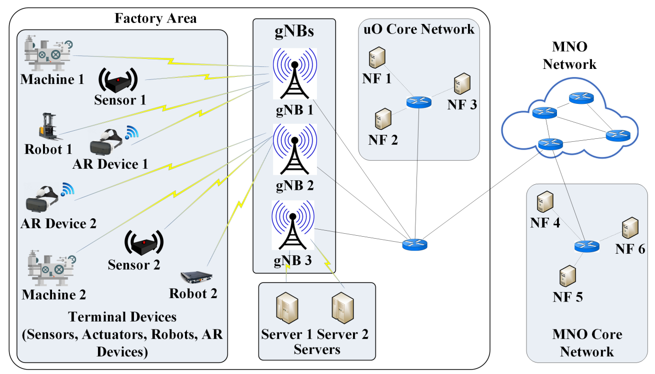


Fig. 19: Hybrid deployment model to serve factory use cases

where L_{tot} is the aggregated latency of all packets and P_{tot} is the count of all data packets transferred during a day. For a factory having AR, massive wireless sensor networks and mobile robots use cases, L_{tot} can be expressed as

$$L_{tot} = L_{AR} + L_S + L_R \tag{2}$$

where L_{AR} is the aggregated latency of all AR packets, L_S is the aggregated latency of all sensor network packets and L_R is the aggregated latency of all packets from robots during a day. P_{tot} can be expressed as

$$P_{tot} = P_{AR} + P_S + P_R \tag{3}$$

where P_{AR} is AR data packet count, P_S is sensor data packet count and P_R is data packet count from robots, during a day.

Simulations for each use case are described below along with the respective formulas for latency. For the AR use case, we derive all the formulas and extend those formulas to other two use cases.

1) Augmented Reality: For AR use case, we assume the properties of AR device deciding the parameters of the video stream. We take the resolution of AR device screen as 400 x 240 with a frame rate of 20 frames per second. We take that the AR device captures 8-bit color video. We assume that the video transmission is uncompressed due to the processing power limitations at the AR device. We take the average duration of video is 10 s for the simulation purpose, even though it may be longer in reality. The device registers with the network as soon as it turned on each day and we assume this happens only once a day. Whenever there is a need to use the AR device by a factory worker, it establishes a data connection with the server and then transfers data. Number of such data transfers are decided based on the frequency of AR sessions per day.

Mathematical expression for the aggregated AR latency L_{AR} will be

$$L_{AR} = L_{reg} + N_{ses} .(L_{ses} + L_{dat}) \tag{4}$$

where L_{reg} is the aggregated latency of registration process, L_{ses} is the aggregated latency of one session establishment and L_{dat} is the aggregated latency of one data transfer process. N_{ses} is the average number of AR sessions per day. L_{reg}, L_{ses} and L_{dat} are latency components belong to AR communications. Those latency components include the latency of 5G access network, gNB, access router, backhaul, core router and NF processing delay. Therefore L_{reg} for L5GO can be expressed as follows.

$$\begin{aligned} L_{reg\_lo} = &k_1 .T_{acc} + k_2 .T_{gnb} + k_3 .T_{router\_acc} + \\ &k_4 .T_{back\_lo} + k_5 .T_{router\_lo} + k_6 .T_{NF\_lo} \end{aligned} \tag{5}$$

where T_{acc} is the mean delay from AR device to AN, T_{gnb} is the mean delay at gNB due to processing and queuing, T_{router_acc} is the mean queuing delay at the access router, T_{back_lo} is the delay of the backhaul network from access router to core router, T_{router_lo}, is the mean delay at the L5GO's core router and T_{NF_lo} is the mean NF processing delay of L5GO. k_1, k_2, k_3, k_4, k_5, k_6 are the the number of times where the registration message passes through each respective element.

Average queuing delay of elements such as routers can be expressed with their packet arrival rate and service rate. We model the router queues as M/M/1 queues and the average delay at any router T_{router_xx} can be expressed as

$$T_{router\_xx} = \frac{1}{\mu - \lambda} \tag{6}$$

where λ being the arrival rate while μ being the service rate.

Since we compare the performance of the pure local, MNO and hybrid architectures, first three terms of equation (5) are common and can be considered as a constant k_{reg}. Therefore equation (5) can be re-written as

$$L_{reg\_lo} = k_{reg} + k_4 .T_{back\_lo} + k_5 .T_{router\_lo} + k_6 .T_{NF\_lo} \tag{7}$$

Similarly, we can write the expressions for L_{ses_lo} and L_{dat_lo} and finally, the aggregated latency L_{AR} for L5GO can be expressed as

$$L_{AR\_lo} = a_{AR} + a_1 .T_{back\_lo} + a_2 .T_{router\_lo} + a_3 .T_{NF\_lo} \tag{8}$$

where a_{AR}, a_1, a_2, a_3 are constants applicable for M = 0. These constants are taken to represent the entire AR communication in one day, therefore N_{ses} is also integrated into these constants.

For MNO, equation (8) can be used with minor modifications as in (9).

$$\begin{aligned} L_{AR\_mno} = &a_{AR} + a_1 .T_{back\_mno} + \\ &a_2 .T_{router\_mno} + a_3 .T_{NF\_mno} \end{aligned} \tag{9}$$

For hybrid architecture, both L5GO and MNO cores are involved and the latency of AR use case can be expressed as

$$\begin{aligned} L_{AR\_hybrid} = &a_{AR} + a_{11} .T_{back\_lo} + a_{12} .T_{back\_mno} + \\ &a_{21} .T_{router\_lo} + a_{22} .T_{router\_mno} + \\ &a_{31} .T_{NF\_lo} + a_{32} .T_{NF\_mno} \end{aligned} \tag{10}$$

where a_{11}, a_{12}, a_{21}, a_{22}, a_{31}, a_{32} are being the number of times where different messages are processed at respective element. When M = 0, equation (10) simplifies to equation (9) and when M = M_{max} equation (10) simplifies to equation (8).

a_{AR} is constant for AR use case for all architectures while a_{ij} values depend two factors. 1. the number of NFs hosted at L5GO (M) and 2. which NFs are hosted for that M. Therefore, deriving a direct relationship between L_{AR_hybrid} and M is not possible and the latency results should be taken using the simulations for each M.

We can express T_{back_mno} using the distance to core network D_{back_mno} as follows.

$$T_{back\_mno} = a .D_{back\_mno} \tag{11}$$

where a is a constant represents per km latency over fiber backhaul channel.

To achieve the given E2E AR latency defined in Section III-A, for a known M, we can identify D_{back_mno} in a hybrid architecture using equation (10) and (11).

2) Massive wireless sensor networks: For sensor network communications, we assume that the sensors are working 24 x 7 and device registration to the network is already completed. Whenever there is a need to transfer sensor data and take appropriate action, the message flow occurs according to Figure 10. It includes three session establishments and three data transfers between different entities.

Similar to the AR use case, we can derive expressions for L_{S_lo}, L_{S_mno} L_{S_hybrid} as follows.

$$L_{S\_lo} = b_S + b_1 .T_{back\_lo} + b_2 .T_{router\_lo} + b_3 .T_{NF\_lo} \tag{12}$$

$$\begin{aligned} L_{S\_mno} = &b_S + b_1 .T_{back\_mno} + b_2 .T_{router\_mno} + \\ &b_3 .T_{NF\_mno} \end{aligned} \tag{13}$$

$$\begin{aligned} L_{S\_hybrid} = &b_S + b_{11} .T_{back\_lo} + b_{12} .T_{back\_mno} + \\ &b_{21} .T_{router\_lo} + b_{22} .T_{router\_mno} + \\ &b_{31} .T_{NF\_lo} + b_{32} .T_{NF\_mno} \end{aligned} \tag{14}$$

where b_S is constant for massive wireless sensor network use case for all architectures while b_{ij} values have the similar interpretation as in AR use case. For a given M, to identify D_{back_mno} satisfying a given latency in any event of sensor networks use case, equation (11) can be used with equation (14), similar to the way it is used in AR use case.

3) Mobile robots: For the mobile robot simulations, we consider the devices are switched on at the beginning of each day making device registration process mandatory for each robot. Session establishments are data transfer processes occur based on the frequency of robot usage. Handovers due to the mobility of robots are also simulated with a given frequency.

Similar to the AR and sensor network use cases, we can express L_{R_lo}, L_{R_mno} L_{R_hybrid} for mobile robots use case as follows. Here, latency due to handover in also included in these formulas.

$$L_{R\_lo} = c_R + c_1 .T_{back\_lo} + c_2 .T_{router\_lo} + c_3 .T_{NF\_lo} \quad (15)$$

$$\begin{aligned} L_{R\_mno} =& c_R + c_1 .T_{back\_mno} + c_2 .T_{router\_mno}+ \\ & c_3 .T_{NF\_mno} \end{aligned} \quad (16)$$

$$\begin{aligned} L_{R\_hybrid} =& c_R + c_{11} .T_{back\_lo} + c_{12} .T_{back\_mno}+ \\ & c_{21} .T_{router\_lo} + c_{22} .T_{router\_mno}+ \\ & c_{31} .T_{NF\_lo} + c_{32} .T_{NF\_mno} \end{aligned} \quad (17)$$

where c_R is constant for mobile robots use case for all architectures while c_{ij} values have the similar interpretation as in previous use cases.

Table II outlines the general simulation parameters. Latency of Access Network (AN) is based on the 3GPP study on next generation access technologies [27] and we assume that access networks of both MNO and L5GO have similar properties. We take backhaul as a fiber connection and the latency parameters are selected based on a study of 5G backhaul challenges [28]. We assume that the gNB supports packet forwarding at 1 Gbps which provides a service time of 0.5 s. Similarly, we consider the access router has 1 Gbps interfaces while core routers having 10 Gpbs interfaces, providing 0.05 ms of service time. This approach is motivated by [29]. For the service time of network functions, we assume that the operations require a certain number of CPU cycles [30] and the computing capacity at each network function is designed to handle the loads proportionally. Therefore, we take them as 1 ms for L5GO and 0.1 ms for MNO. The same concept is used for taking the processing delay of servers.For each experiment, we measure the E2E latency of the communications.

TABLE II: General simulation parameters

Parameter	Value
Parameters common to all use cases	
Number of factories served by MNO (N)	10
Simulation duration for factory (per day)	8 working hours
Distance to L5GO core network	500 m
Distance to MNO core network	250 km
Latency between terminal devices and AN	0.5 ms [27]
Latency between AN and core network	0.05 ms/km [28]
Mean service time of gNB	0.5 ms [29]
Mean service time of access router	0.5 ms [29]
Mean service time of L5GO core router	0.05 ms [29]
Mean service time of MNO core router	0.05 ms [29]
Mean service time of NF at L5GO	1 ms [30]
Mean service time of NF at MNO	0.1 ms [30]
Use case 1 - AR	
Mean video duration	10 seconds
Size of UDP packet from AR device	64 kB
Frequency of AR sessions	every 30 minutes
AR server processing delay	1 ms [30]
Number of AR devices operating in parallel	3 [7]
Use case 2 - Sensor networks	
Data packets needed for sensor data	10
Frequency of sensor data transfer	every 10 minutes
Alarm server processing delay	1 ms [30]
Action server processing delay	1 ms [30]
Processing delay of the machine	1 ms [30]
Number of sensors operating in parallel	20 [7]
Use case 3 - Mobile robots	
Data packets needed for robot data	16
Frequency of robot data transfer	every 5 minutes
Frequency of robot handover	every 30 minutes
Processing delay of the robot	1 ms [30]
Number of robots operating in parallel	10 [7]

VII. Simulation Results Analysis

In this section, we present the simulation results for pure local deployment, MNO deployment and hybrid architecture deployments. First we consider each use case separately and analyze the latency. This is possible because each use case is served by a different network slice. Under the hybrid architecture, the performance of four algorithms are discussed while varying M from 0 to M_{max}. We keep the distance to core network constant at 250 km and identify the best algorithm to be used to deploy NFs. We also consider an average latency for a factory comprising all three use cases. Secondly, we use the best algorithm selected at the first step, vary the distance to core network and identify at which core network distance each event of the use cases can be satisfied. Finally, we formulate a relationship between the number of NFs needed to deploy locally with respect to the MNO core network distance so that the required latencies are satisfied.

A. Analysis for a fixed MNO distance

1) Augmented Reality: Simulation result for the average latency of pure local architecture is 4.77 ms and for MNO based architecture, it is 29.29 ms. This means that the pure uO architecture can satisfy the required one way E2E latency of 10 ms [7]. For the hybrid architectures, the average latency will vary based on which NFs are deployed at the L5GO. Different algorithms will provide different set of NFs to be deployed at L5GO, therefore the average latency of the architecture derived from different algorithms may vary even for the same M.

a) FCFS algorithm: To identify the NF which is needed first, it is fair to assume that the device registrations occur before the session establishments and data transfers in daily factory operations. Therefore, under this algorithm, NFs needed for device registration will be deployed at the L5GO's core network first. Table III outlines the NFs which will be implemented at L5GO under this algorithm for each M.

The average latency of AR packets under FCFS algorithm is depicted in Figure 20 with 95% confidence intervals. Latency of pure local architecture and MNO architecture is also depicted in Figure 20 for reference. It is clearly observed that when M = 0, latency of hybrid architecture approaches MNO

TABLE III: NFs deployed at L5GO under FCFS algorithm

Number of NFs (M)	NFs
0	–
1	AMF
2	AMF, AUSF
3	AMF, AUSF, UDM
4	AMF, AUSF, UDM, PCF
5	AMF, AUSF, UDM, PCF, SMF
6	AMF, AUSF, UDM, PCF, SMF, UPF

TABLE IV: NFs deployed at L5GO under MCNF algorithm

Number of NFs (M)	NFs
0	–
1	SMF
2	SMF, AMF
3	SMF, AMF, UPF
4	SMF, AMF, UPF, UDM
5	SMF, AMF, UPF, UDM, PCF
6	SMF, AMF, UPF, UDM, PCF, AUSF

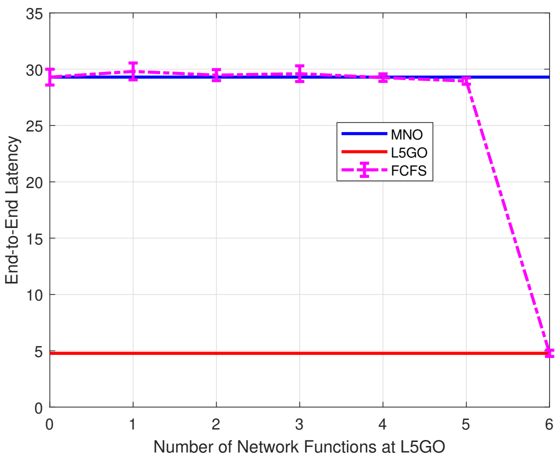


Fig. 20: Average AR packet latency under FCFS algorithm

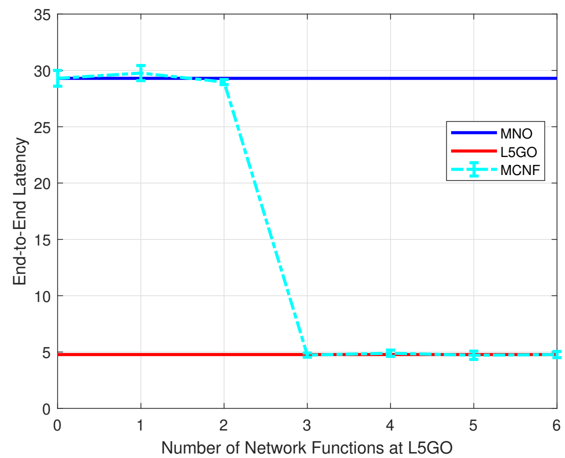


Fig. 21: Average AR packet latency under MCNF algorithm

latency while $M = M_{max}$, hybrid architecture takes the latency of pure local architecture.

The key observation is, for certain M values, the latency is even higher than MNO latency. Reason for this can be explained as follows. If two NFs have number of message transfers between them, then locating one of those functions at L5GO and the other function at MNO will increase the latency because every time when these NFs wants to communicate, the backhaul connection is used. In the message sequence diagrams, we observed that there are number of interactions between AMF and SMF. Since AMF is the first NF to be deployed at L5GO, until SMF is deployed at $M = 5$, latency is higher than MNO latency. Second observation is that there is no significant latency reduction until UPF is deployed at L5GO. This is due to the fact that, UPF is the most utilized NF in AR communications which is used for data transfers (user plane data) and it has comparatively less interactions with other NFs during the data transfer. Overall, the architecture derived from FCFS algorithm does not provide better latency results until $M = 6$.

b) MCNF algorithm: In the factory communications, certain NFs are used more frequently than the others. MCNF algorithm utilizes this fact and deploys most common NFs at L5GO. Most common NFs are decided after considering all three use case communications. Table IV outlines the order which the NFs are assigned under each M.

The average latency of AR packets is depicted in Figure 21 along with the MNO and L5GO latency. Significant reduction in latency is observed when UPF is deployed at L5GO for $M = 3$, the same observation with UPF under FCFS algorithm.

c) Operator Policy Based Placement: Allocation of NFs for operator policy based placement under each M is listed in Table V. The basis for this was described in Section V-C3.

Average AR latency under OPBP is depicted in Figure 22. As expected, a significant improvement in latency is achieved at $M = 1$ by deploying UPF by L5GO. After that, latency has no specific pattern because the main focus of this algorithm is not to achieve minimum latency, but to follow the policy.

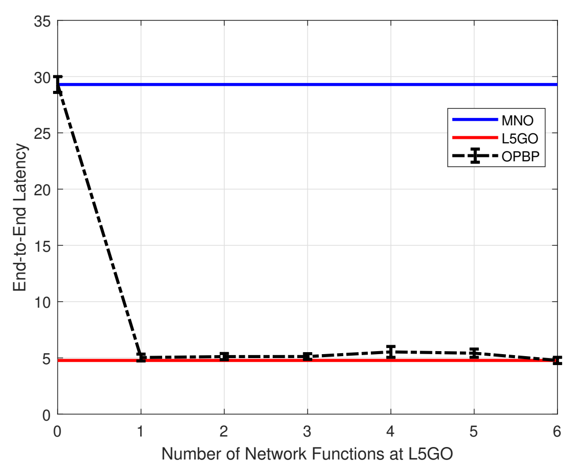


Fig. 22: Average AR packet latency under OPBP algorithm

d) Predictive Placement algorithm: Predictive placement algorithm utilizes available information on NF usage of the factory use cases and tries to minimize the latency with minimum M. Simulation results of FCFS, MCNF and OPBP algorithms showed that deploying UPF at L5GO contributes in significant reductions of latency. Next two NFs should be AMF and SMF because they are the next most used NFs in these communications. The rest of the functions will be UDM, PCF and AUSF respectively. Table VI shows the order of NFs

TABLE V: NFs deployed at L5GO under OPBP

Number of NFs (M)	NFs
0	–
1	UPF
2	UPF, UDM
3	UPF, UDM, PCF
4	UPF, UDM, PCF, AUSF
5	UPF, UDM, PCF, AUSF, AMF
6	UPF, UDM, PCF, AUSF, AMF, SMF

deployed at L5GO with respect to M. Generic algorithm used in obtaining the order of NFs is explained in Section V-C4.

TABLE VI: NFs deployed at L5GO under PP algorithm

Number of NFs (M)	NFs
0	–
1	UPF
2	UPF, AMF
3	UPF, AMF, SMF
4	UPF, AMF, SMF, UDM
5	UPF, AMF, SMF, UDM, PCF
6	UPF, AMF, SMF, UDM, PCF, AUSF

Average latency results of AR use case under predictive algorithm is depicted in Figure 23. As expected, a significant improvement in latency is achieved at $M = 1$ by deploying UPF at L5GO. The next improvement is observed when both AMF and SMF are at L5GO when $M = 3$. Finally, it approaches pure local architecture latency when $M = M_{max}$.

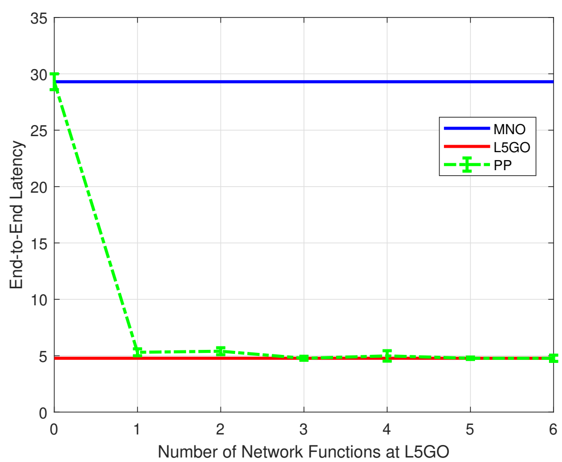


Fig. 23: Average AR packet latency under PP algorithm

Latency performance of hybrid architectures obtained using all the algorithms for AR use case is shown in Figure 24. Since AR use case has significant data transfers in user plane, latency is heavily reduced once UPF is deployed locally. It is clear that the predictive placement algorithm derives the architecture which provides the best performance for AR use case with respect to latency.

2) Massive Wireless Sensor Networks: Simulation results for the massive wireless sensor networks use case under all algorithms are depicted in Figure 25 with 95% confidence intervals. As we can see, FCFS algorithm provides higher latency than MNO until $M = 4$. In the sensor networks use case, there are multiple session establishments and SMF is heavily involved in these session establishments with terminal

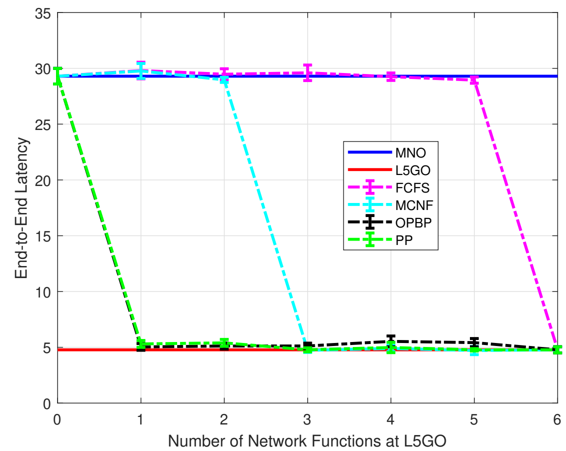


Fig. 24: Average packet latency of AR use case

devices. Under FCFS algorithm, SMF is deployed locally at $M = 5$, therefore there is a significant deduction in latency is observed for $M = 5$. Under MCNF algorithm, even though SMF gets deployed locally first, it makes AMF and SMF being deployed in two locations. This results of an increase in latency due to the communications between AMF and SMF. Therefore latency is higher than MNO with $M = 1$. Since the most used NFs (SMF, AMF and UPF) are deployed locally at $M = 3$ onwards, latency gets significantly lower. In general,predictive algorithm provides better latency with minimum M in this use case also, because deploying UPF locally at $M = 1$ will help low latency data transfers, similar to the behavior of MCNF from $M = 3$ onwards.

There is a significant difference in latency curves between AR use case and sensor networks use case. In AR use case, UPF is heavily used because of the high data transfers. Therefore deploying UPF locally contributes to a drastic decrease in latency as the data stream does not use the backhaul network to MNO. However, for sensor networks, a data transfer happens only after a session establishment process. There is no specific NF which has a significantly higher usage than the other. Therefore, drastic reductions are not observed.

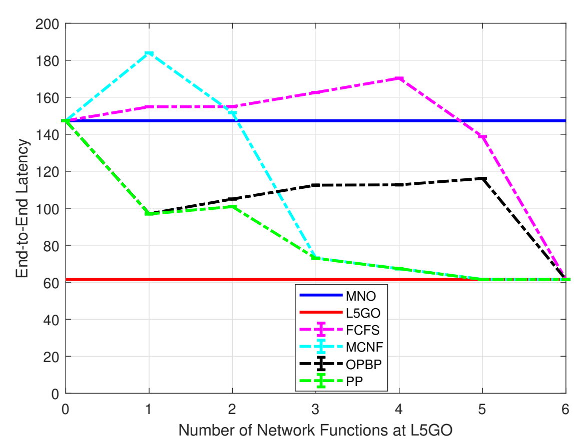


Fig. 25: Average packet latency of sensor networks use case

3) Mobile Robots: For the mobile robots use case, simulation results for the architectures derived from all the algorithms are depicted in Figure 26 with 95% confidence intervals. Results show that the FCFS algorithm provides the worst performance and policy based placement gives slightly better performance than FCFS. Mobile robot communications include a number of handovers. Because of this additional handover procedure, AMF and SMF are utilized heavily than the other use cases. Under FCFS, AMF is deployed locally at M = 5, therefore latency starts to reduce at M = 5. under policy based algorithm, both AMF and AMF are deployed last therefore it does not provide low latency performance until M = 6. MCNF algorithm provides comparatively better performance for M = 3 and beyond. Under MCNF algorithm AMF, SMF and UPF are deployed when M = 3 and most of the message transfers happens between L5GO's core network and the factory. A slightly better performance is seen when predictive algorithm is used. In general, mobile robots use case needs AMF, SMF and UPF to be deployed locally to achieve better latency performance.

Behavior of latency for mobile robots use case differs from both AR and sensor networks use cases because of the way each NF is used for the communication. A heavy usage of UPF is not seen in mobile robots use case like in AR use case. Also, deploying one NF locally does not result in better performance as seen in sensor networks use case. However, when the number of NFs deployed locally is three, far better performance is seen under MCNF and PP algorithms.

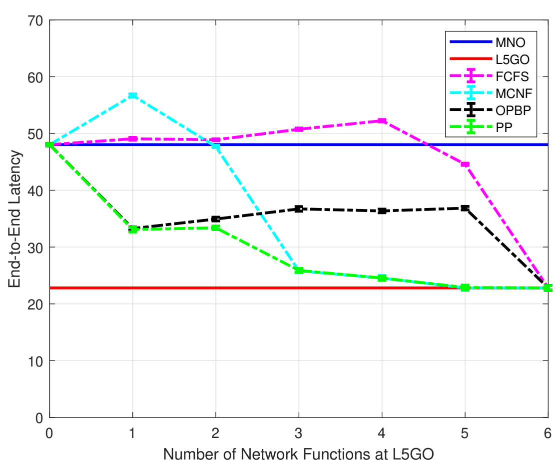


Fig. 26: Average packet latency of mobile robots use case

4) Average for all three use cases: Considering a factory environment having all three use cases, the average packet latency can be calculated using equation (1) and the results are illustrated in Figure 27 with 95% confidence intervals.

Considering the latency results of each use case and the average of all three use cases depicted in Figure 27, in general we can take the architecture derived from predictive algorithm is the best architecture to serve the factory. For the remaining analysis, we consider only the architectures derived using predictive placement algorithm.

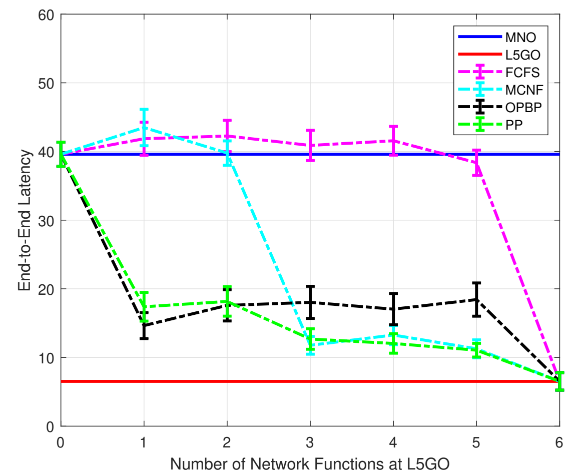


Fig. 27: Average packet latency combining all three use cases

B. Latency with respect to MNO distance

1) Augmented Reality: So, far we have considered a fixed 250 km as the distance to MNO core network. However in reality, the core network distance (d) could differ. Therefore, we take several values for d and observe how the deployment of NFs in a hybrid architecture affects the latency performance of these industry 4.0 use cases. We keep d at 50 km, 100 km, 250 km and 500 km, and observe how many network functions that a hybrid operator should deploy locally to cater to the requirements. We consider only the predictive placement algorithm for this analysis as it is the most efficient algorithm to achieve low latency.

Latency requirements of each use case were highlighted in Section III and we use those values for the analysis. As an example E2E latency of AR packets should be 10 ms. Results of the analysis for AR use case is depicted in Figure 28 alongside with the required AR latency.

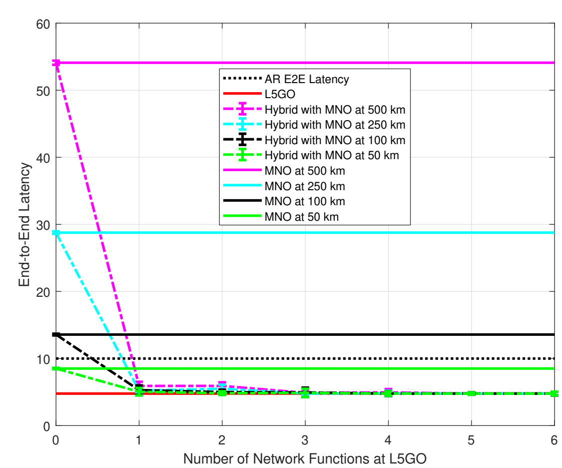


Fig. 28: Average packet latency of AR use case under PP algorithm

We can see that if the core network of MNO is at 50 km, MNO itself can satisfy the required AR E2E latency (10 ms).

However, if the core network is at 100 km, MNO cannot satisfy the requirement but a hybrid operator with UPF deployed locally would be able to serve. Since AR communication utilizes UPF heavily than other NFs due to its high data transfer needs, deploying UPF locally would be enough to satisfy the required AR E2E latency as seen in d = 250 km and d = 500 km curves.

2) Massive wireless sensor networks: The same analysis was carried out for massive wireless sensor networks use case and the results are depicted in Figure 29. Latency requirement of sensor network use case was highlighted in Section III-B. For the event based monitoring the requirement is between 50 ms and 1 s, therefore we consider the mean 475 ms. For condition monitoring for safety, the requirement is between 5 ms and 10 ms therefore we take it as 7.5 ms, again using the mean.

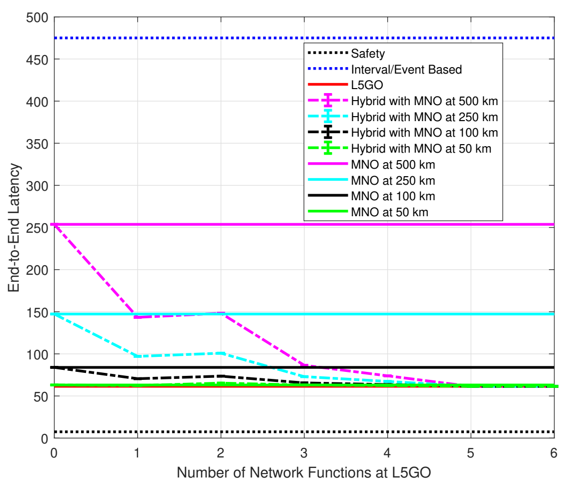


Fig. 29: Average packet latency of massive wireless sensor networks use case under PP algorithm

It is seen from the results that for interval and event based requirements can be catered by MNO whereas the latency requirement of condition monitoring for safety cannot be catered even by L5GO. This is because of the number of session establishments and data transfers involved in the sensor networks communications. One solution for further reduce the latency is to keep the sessions established at all times between the terminal devices such as sensors and servers. However, additional concerns like energy consumption will arise with continuous session establishments. Therefore, Keeping the sessions established only for critical sensors is a possible solution.

3) Mobile robots: We take the required latencies for each event in mobile robots based on Section III-C, 30 ms for cooperative driving of robots, 55 ms for video operated remote control.

Figure 30 depicts the latency performance of each event under mobile robots use case. For cooperative driving, when d is 100 km or less, MNO can satisfy the latency requirement, but when the core network is at 150 km it needs at least 1 NF deployed locally to meet the latency requirements. When d is 200 km and beyond, it needs at least 3 NFs. Similar analysis

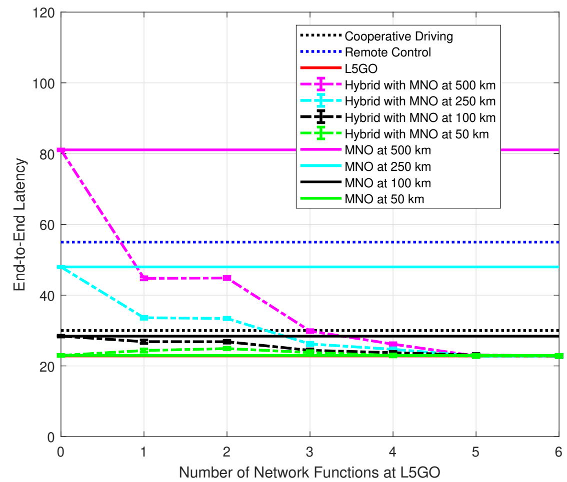


Fig. 30: Average packet latency of mobile robots use case under PP algorithm

is possible for remote control of robots. Results reveal that when d is high, it needs more NFs deployed locally to satisfy the required latency.

Figure 31 summarizes the results of the analysis by representing how many NFs are needed locally for each MNO distance so that it can cater to each event of the use cases. In this case, we vary d from 50 km to 500 km in 50 km intervals to obtain better results. In general, when the distance is increasing more NFs needs to be deployed locally and it depends on the nature of the use case communications.

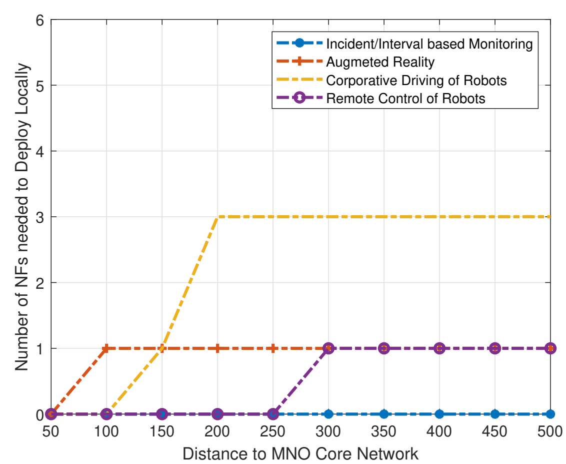


Fig. 31: Number of NFs needed with respect to d

VIII. Conclusions

The novel concept of local 5G operator enables a versatile set of stakeholders to operate 5G networks within their premises with a guaranteed quality and reliability to complement traditional Mobile Network Operator (MNO) offerings. In this paper, we analyzed the feasibility and performance advantage of using local 5G operators instead of a traditional MNO in a smart factory environment which supports industry 4.0 standards.

We proposed two architecture options called pure local architecture and hybrid architecture based on 5G standards which is customized for factory environment. The architectures were discussed in terms of 5G core network functions and the number of core network functions that are deployed locally instead of deploying them at MNO core network. To realize the conceptual design, we conducted several experiments for an Augmented Reality (AR), massive wireless sensor networks and mobile robots use cases which will be heavily used in future factories.

The experiments revealed that a local 5G network established within the factory premises can provide low end-to-end latency for each use case compared to an MNO provided 5G network, where the core network is located outside the factory premises. The end-to-end latency of the communication exhibits a significant increase over the distance between the core network and the factory. In a hybrid architecture where both local 5G operator and MNO collaboratively establish the core network, the decision on which network functions to be deployed locally depends on the nature of the use case(s). The most efficient way to achieve low latency with minimal network functions deployed locally is to analyze the past usage data and decide the network functions, as done in the predictive placement algorithm. When the collaborating MNO is located close to the factory environment, a given latency can be achieved with few network functions. When the distance to MNO core network is increasing, more network functions needs to be deployed locally to achieve the same latency. In a pure local 5G operator served factory, the data stream stays within the factory premises because the core network is located inside a confined environment. This ensures improved secure communication between terminal devices and the servers.

In future, we consider more network function placement algorithms to derive better hybrid architecture options. Moreover, we consider the analysis of power profile and management of IoT devices under the use cases served by different architectures.

Acknowledgement

This work is supported by Business Finland in uO5G and MOSSAF projects, Academy of Finland in 6Genesis Flagship (grant no. 318927) and 5GEAR (Grant No. 319669) projects, and European Union in RESPONSE 5G (Grant No: 789658) and INSPIRE-5Gplus (Grant No: 871808) projects.

References

[1] S. E. Elayoubi, M. Fallgren, P. Spapis, G. Zimmermann, D. Martín-Sacristán, C. Yang, S. Jeux, P. Agyapong, L. Campoy, Y. Qi *et al.*, “5G Service Requirements and Operational Use Cases: Analysis and METIS II Vision,” in *2016 European Conference on Networks and Communications (EuCNC)*. IEEE, 2016, pp. 158–162.

[2] European Commission, “Strategic Spectrum Roadmap towards 5G for Europe: RSPG Second Opinion on 5G Networks,” *Radio Spectrum Policy Group (RSPG)*, RSPG18-005,2018.

[3] M. Matinmikko, M. Latva-Aho, P. Ahokangas, S. Yrjölä, and T. Koivumäki, “Micro Operators to boost Local Service Delivery in 5G,” *Wireless Personal Communications*, vol. 95, no. 1, pp. 69–82, 2017.

[4] M. Matinmikko-Blue and M. Latva-aho, “Micro Operators Accelerating 5G Deployment,” in *Industrial and Information Systems (ICIIS), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–5.

[5] A. Prasad, Z. Li, S. Holtmanns, and M. A. Uusitalo, “5G Micro-Operator Networks-A key enabler for New Verticals and Markets,” in *2017 25th Telecommunication Forum (TELFOR)*. IEEE, 2017, pp. 1–4.

[6] Y. Siriwardhana, P. Porambage, M. Liyanage, J. S. Walia, M. Matinmikko-Blue, and M. Ylianttila, “Micro-operator driven Local 5G Network Architecture for Industrial Internet,” in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–8.

[7] 3GPP, “Study on Communication for Automation in Vertical Domains (CAV),” Technical Report, Dec 2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3187

[8] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5G be?” *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[9] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, “5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[10] M. Series, “IMT Vision–Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond,” *Recommendation ITU*, pp. 2083–0, 2015.

[11] M. Liyanage, I. Ahmad, A. B. Abro, A. Gurtov, and M. Ylianttila, *A Comprehensive Guide to 5G Security*. John Wiley & Sons, 2018.

[12] 3GPP, “System Architecture for the 5G System,” Technical Specification, June 2019. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[13] S. Sezer, S. Scott-Hayward, P. K. Chouhan, B. Fraser, D. Lake, J. Finnegan, N. Viljoen, M. Miller, and N. Rao, “Are we Ready for SDN? Implementation Challenges for Software-Defined Networks,” *IEEE Communications Magazine*, vol. 51, no. 7, pp. 36–43, 2013.

[14] P. Ahokangas, M. Matinmikko-Blue, S. Yrjölä, V. Seppänen, H. Hämmäinen, R. Jurva, and M. Latva-aho, “Business Models for Local 5G Micro Operators,” *IEEE International Symposium on Dynamic Spectrum Access Networks*, 22-25 October 2018, Soul, South Korea.

[15] M. Wollschlaeger, T. Sauter, and J. Jasperneite, “The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0,” *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 17–27, 2017.

[16] H. Lasi, P. Fettke, H. Kemper, T. Feld, and M. Hoffmann, “Industry 4.0.” *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014.

[17] A. Varghese and D. Tandur, “Wireless Requirements and Challenges in Industry 4.0,” in *Contemporary Computing and Informatics (IC3I), 2014 International Conference on*. IEEE, 2014, pp. 634–638.

[18] X. Li, D. Li, J. Wan, A. V. Vasilakos, C.-F. Lai, and S. Wang, “A Review of Industrial Wireless Networks in the Context of Industry 4.0,” *Wireless Networks*, vol. 23, no. 1, pp. 23–41, 2017.

[19] NGMN Alliance, “Description of Network Slicing Concept,” *NGMN 5G P*, vol. 1, 2016.

[20] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, “Network Slicing based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges,” *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[21] 3GPP, “Study on Management and Orchestration of Network Slicing for Next Generation Network,” Technical Report, June 2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3091

[22] C.-Y. Chang, N. Nikaein, O. Arouk, K. Katsalis, A. Ksentini, T. Turletti, and K. Samdanis, “Slice Orchestration for Multi-Service Disaggregated Ultra-Dense RANs,” *IEEE Communications Magazine*, vol. 56, no. 8, pp. 70–77, 2018.

[23] 3GPP, “Procedures for the 5G System,” Technical Specification, June 2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3145

[24] S. Clayman, E. Maini, A. Galis, A. Manzalini, and N. Mazzocca, “The Dynamic Placement of Virtual Network Functions,” in *2014 IEEE network operations and management symposium (NOMS)*. IEEE, 2014, pp. 1–9.

[25] B. Malet and P. Pietzuch, “Resource Allocation across Multiple Cloud Data Centres,” in *Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science*. ACM, 2010, p. 5.

[26] “MATLAB.” [Online]. Available: https://www.mathworks.com/

[27] 3GPP, “Study on Scenarios and Requirements for Next Generation Access Technologies,” Technical Report, September

2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2996

[28] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G Backhaul Challenges and Emerging Research Directions: A Survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[29] R. Ramaswamy, N. Weng, and T. Wolf, "Characterizing Network Processing Delay," in *IEEE Global Telecommunications Conference, 2004. GLOBECOM'04.*, vol. 3. IEEE, 2004, pp. 1629–1634.

[30] J. Ren, Y. He, G. Huang, G. Yu, Y. Cai, and Z. Zhang, "An Edge-computing based Architecture for Mobile Augmented Reality," *IEEE Network*, vol. 33, no. 4, pp. 162–169, 2019.

Yushan Siriwardhana (Member, IEEE) is currently a Doctoral degree student in the Centre for Wireless Communications, University of Oulu, Finland. He received the bachelor's degree in electronics and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2009 and the master's degree in wireless communication engineering from the University of Oulu, Finland, in 2019. He has over 7 years of experience in telecommunication industry. His research interests include 5G, Local 5G and future networks, Security in IoT, MEC.

Pawani Porambage (Member, IEEE) is a post-doctoral researcher at the Centre for Wireless Communications, University of Oulu, Finland. She received her doctoral degree on Communications Engineering from University of Oulu, Finland (2018). She obtained MSc and BSc degrees respectively from University of Nice Sophia-Anipolis, France (2012) and University of Moratuwa, Sri Lanka (2010). Her main research interests include lightweight security protocols, blockchain, security and privacy on IoT and MEC, Network Slicing and Wireless Sensor Networks.

Mika Ylianttila (Senior Member, IEEE) is a full-time associate professor (tenure track) at the Centre for Wireless Communications (CWC), at the Faculty of Information Technology and Electrical Engineering (ITEE), University of Oulu, Finland. He is leading a research team and is the director of communications engineering doctoral degree program. Previously he was the director of Center for Internet Excellence (2012-2015), vice director of MediaTeam Oulu research group (2009-2011), and professor (pro tem) in computer science and engineering, and director of information networks study programme (2005-2010). He received his doctoral degree on Communications Engineering at the University of Oulu in 2005. He has coauthored more than 150 international peer-reviewed articles. His research interests include edge computing, network security, network virtualization and software-defined networking. He is a Senior Member of IEEE, and Editor in *Wireless Networks* journal.

Madhusanka Liyanage (Senior Member, IEEE) is currently an Ad Astra Fellow/Assistant Professor at School of Computer Science, University College Dublin, Ireland. He is also an adjunct professor at the University of Oulu, Finland. He received his B.Sc. degree (First Class Honours) in electronics and telecommunication engineering from the University of Moratuwa, Moratuwa, Sri Lanka, in 2009, the M.Eng. degree from the Asian Institute of Technology, Bangkok, Thailand, in 2011, the M.Sc. degree from the University of Nice Sophia Antipolis, Nice, France, in 2011, and the Ph.D. degree in communication engineering from the University of Oulu, Oulu, Finland, in 2016. He was also a recipient of prestigious Marie Skłodowska-Curie Actions Individual Fellowship during 2018-2020. From 2011 to 2012, he worked a Research Scientist at the I3S Laboratory and Inria, Sophia Antipolis, France. He has been a Visiting Research Fellow at the Department of Computer Science, University of Oxford, Data61, CSIRO, Sydney, Australia, the Infolabs21, Lancaster University, U.K., and Computer Science and Engineering, The University of New South Wales during 2015-2018.

He has co-authored over 90 publications including two edited books with Wiley and one patent. He is the demo co-chair of WCNC 2019, publicity chair of ISWCS 2019 and poster chair of 6G Summit 2020. He served as a Technical program Committee Members at EAI M3Apps 2016, 5GU 2017, EUCNC 2017, EUCNC 2018, 5GWF 2018, MASS 2018, MCWN 2018, WCNC 2019, EUCNC 2019, EUCNC 2020, MASS 2020. ICBC 2021 conferences and Technical program co-chair in SecureEdge workshop at IEEE CIT2017 conference and Blockchain for IoT workshop at IEEE Globecom 2018, IEEE ICC 2020 and IEEE 5GWF 2020. He has also served as the session chair in a number of other conferences including IEEE WCNC 2013, CROWNCOM 2014, 5GU 2014, IEEE CIT 2017, IEEE PIMRC 2017, 5GWF 2018, Bobynet 2018, Globecom 2018, WCNC 2019, ICC 2020. Moreover, He has received two best Paper Awards in the areas of SDMN security (at NGMAST 2015) and 5G Security (at IEEE CSCN 2017). Additionally, he has been awarded three research grants and 22 other prestigious awards/scholarships during his research career.

Dr. Liyanage has worked for more than twelve EU, international and national projects in ICT domain. He held responsibilities as a leader of work packages in several national and EU projects. Currently, he is the Finnish national coordinator for EU COST Action CA15127 on resilient communication services. In addition, he is/was serving as a management committee member for four other EU COST action projects namely EU COST Action IC1301, IC1303, CA15107 and CA16226. Liyanage has over three years' experience in research project management, research group leadership, research project proposal preparation, project progress documentation and graduate student co-supervision/mentoring, skills. In 2015, 2016 and 2017, he won the Best Researcher Award at the Centre for Wireless Communications, University of Oulu for his excellent contribution in project management and dissemination activities. Additionally, two of the research projects (MEVICO and SIGMONA projects) received the CELTIC Excellence Award in 2013 and 2017 respectively.

Dr. Liyanage's research interests are 5G, SDN, IoT, Blockchain, MEC, mobile and virtual network security. http://madhusanka.com.