5G CONNI

Private 5G Networks for Connected Industries

Deliverable D3.2

Report on Network Performance Analysis, Maintenance and Monitoring

Co-funded by the Horizon 2020 programme
of the European Union in collaboration with Taiwan

Date of Delivery:	31/07/2022		
Project Start Date:	01.10.2019	**Duration:**	36 Months

Document Information

Project Number:	861459
Project Name:	Private 5G Networks for Connected Industries

Document Number:	D3.2
Document Title:	Report on Network Performance Analysis, Maintenance and Monitoring
Editor:	Sergio Barbarossa (SAP)
Authors:	Mathis Schmieder (HHI), Sven Wittig (HHI), Alper Schultze (HHI), Jack Shi-Jie Luo (ITRI), Frank Li-Fong Lin (ITRI), Henrik Klessig (BOSCH), Stefania Sardellitti (SAP), Jiun-Cheng Huang, Cheng-Yi Chien, Yueh-Feng Li, Ling-Chih Kao (CHT), CC Weng (ANI)
Dissemination Level:	Public
Contractual Date of Delivery:	31/07/2022
Work Package	WP3
File Name:	D3.2_Report_on_Network_Performance_Analysis_Maintenance_and_Monitoring _v4.0.docx

Revision History

Version	Date	Comment
0.1	20.05.2022	First Skeleton
2.0	30.06.2022	Consolidated version
3.2	08.07.2022	Consolidated version
3.3	13.07.2022	III updated Section 5.2 HHI updated Section 2 ANI updated Section 5.1 CHT updated Section 6. SAP integrated the novel contributions and updated the Executive summary and the conclusion.
4.0	18.07.2022	Consolidated version submitted to internal review
5.0	29.07.2022	Consolidated version after internal review

Executive Summary

This report covers the advancement of the research carried out in WP3 with respect to the results reported in the previous deliverable D3.1. The new results include the advancements in developing suitable channel models to be used in an industrial environment, with specific focus on angular modeling, channel modeling in the 300 GHz band, and indoor-to-outdoor propagation. The new channel models have been incorporated in the new algorithms developed to perform an optimal service placement of radio and computational resources in the edge cloud. New algorithms have also been developed and tested to extrapolate connectivity measures, like throughput, from sparse measurements of SNR values, highlighting the differences with respect to the algorithms reported in D3.1. Finally, new monitoring tools have been developed to monitor the network, including the radio access network, the edge cloud, and the core network.

List of Figures

List of Acronyms

5G CONNI	5G for Connected Industries
TSP	Topological Signal Processing
VUCA	Virtual uniform circular array antenna
IF	intermediate frequency
PDP	power delay profiles
APDP	average power delay profiles
FSPL	free space path loss
LOS	Line-of-sight
NLOS	Non-line-of-sight
OLOS	Obstructed line-of-sight
CIR	channel impulse response
CDF	cumulative distribution function
PLE	path loss exponent
REM	Radio Environment Map
GSP	Graph Signal Processing
EMF	Electro-magnetic field
MSE	Mean Square Error
CSI	channel state information
MEC	Mobile edge computing
ES	Edge server
UE	User equipment
SINR	Signal-to-Interference-plus-Noise Ratio
MCS	Modulation scheme and Coding Schemes
SS-RSRP	synchronization signal reference signal received power
ICO	Intelligent Cell Optimization
RAN	Radio access network
MIMO	Multiple Input Multiple Output
OLSM	Open-Loop Spatial Multiplexing
QAM	Quadrature Amplitude Modulation
BLER	Block Error Ratio
CQI	Channel Quality Indicator
SSS	Secondary Synchronization Signal
CSI-RS	Channel State Information Reference Signal
CSI-SINR	Channel State Information-Signal to Interference plus Noise Ratio
CNC	Computer Numerical Control
IMTC	Intelligent Machinery Technology Center
RF	Radio Frequency
RMS	Root mean square

Table of Contents

1 Introduction

The goal of this deliverable is to report the advancements achieved in Tasks T3.1 and T3.2 of WP3 with respect to the activity reported in the first deliverable D3.1. We recall that one of the main goals of T3.1 is the development of a realistic propagation model in an industrial environment, supported by an intensive measurement campaign. The first results on channel modeling based on the measurement campaign carried out at Bosch's premises were reported in D3.1. In this deliverable we add a spatial channel modeling, a model for the 300 GHz band and the measurement of the indoor-to-outdoor propagation. A second goal of T3.1 was the development of algorithms to create a connectivity map based on a limited number of measurements. In D3.1, we reported the results achieved by using a topological model exploiting simplicial complexes, as an extension of graph-based models. In this deliverable we present an alternative approach, based on the line-graph models and our goal is to extrapolate throughput data starting from a graphical model based on the received signal strength and a few sparse measurements of throughput data.

In parallel, one the objectives of task T3.2 is to set-up an application-centric design able to allocate resources in order to enable the production process to proceed smoothly. This requires remote control of the whole process, including the RAN, MEC and core network. In D3.1, we showed the results of an optimal service placement algorithm that optimized the association between mobile devices, radio access points and edge servers. In this deliverable, we expand that approach by including the allocation of network resources, such as bandwidth and CPU clock rates, and we show significant performance improvements. Furthermore, we incorporated the channel model tailored to the industrial environment, so as to provide a service placement algorithm suitable for an application to Industry 4.0. Finally, in this deliverable, we report the advancement in the development of tools to monitor the edge cloud, including monitoring of the core network, the edge cloud and the radio access network.

2 Channel modeling based on field measurements carried out in an industrial premise

As described in the previous deliverable D3.1, an extensive channel measurement campaign was conducted by HHI at a BOSCH manufacturing facility. Based on the data gathered during these measurements, large scale parameters were estimated and presented in D3.1. Furthermore, the measurement data enabled HHI to derive the channel models that have been used as input to the algorithms developed by SAP for optimal service placement, as detailed in Section 4.

Besides estimation of large scale parameters, the measurements conducted allow also the evaluation of spatial information at the receiver both at 3.7 and 28 GHz. Furthermore, several measurements at 300 GHz have been conducted inside the BOSCH facility and the amount of radiation leaving the hall at 3.7 and 28 GHz have been evaluated.

In the following sub-sections, the evaluation of spatial information at 3.7 and 28 GHz, the evaluation of the 300 GHz measurements and the indoor-to-outdoor experiments will be presented. The results of the large scale parameter estimation as presented in deliverable D3.1 is summarized again in Table 1. Together with the angular results in this deliverable, a complete channel model of the characterized environment is given.

Table 1: Estimated Large Scale Parameters

Scenario	**Perimeter (Scenario 1)**				**Storage Area (Scenario 2)**				**Shop Floor (Scenario 3)**			
Frequency	**3.7 GHz**		**28 GHz**		**3.7 GHz**		**28 GHz**		**3.7 GHz**		**28 GHz**	
LOS/NLOS	**LOS**	**NLOS**	**LOS**	**NLOS**	**LOS**	**NLOS**	**LOS**	**NLOS**	**LOS**	**NLOS**	**LOS**	**NLOS**
Path Loss (FI)												
$PL_0(d_0)$ (dB)	42.44	50.77	59.99	63.06	44.90	60.21	59.92	63.71	45.53	48.56	62.77	64.63
n	1.98	1.53	1.94	1.93	1.66	0.73	2.03	2.06	1.38	1.89	1.73	2.44
σ (dB)	1.39	2.27	1.21	2.60	1.69	1.61	1.04	2.53	3.45	4.41	1.82	3.20
Path Loss (FR)												
$PL_0(d_0)$ (dB)	43.81	43.81	61.38	61.38	43.81	43.81	61.38	61.38	43.81	43.81	61.38	61.38
n	1.88	1.96	1.82	2.04	1.75	2.14	1.90	2.27	1.57	2.27	1.86	2.71
σ (dB)	1.43	2.41	1.28	2.62	1.70	2.46	1.10	2.54	3.49	4.46	1.85	3.26
K-factor												
Mean (dB)	4.28	-	7.22	-	4.53	-	5.78	-	4.16	-	-4.99	-
σ (dB)	3.92	-	7.68	-	4.06	-	4.98	-	3.19	-	5.87	-
Delay Spread												
Mean (ns)	26.27	56.04	19.88	30.82	24.12	34.54	20.93	24.89	19.62	38.11	21.53	25.65
Median (ns)	20.71	52.61	16.37	30.19	21.77	34.68	20.75	26.34	16.78	38.82	21.03	25.97
σ (ns)	17.28	21.37	10.99	13.47	9.09	7.02	8.07	3.72	6.31	8.15	7.57	8.41
95% Conf. (ns)	68.99	93.10	39.08	57.49	38.69	48.63	33.89	28.24	29.51	49.19	35.00	38.92

2.1 Evaluation of spatial information

Besides evaluation of large scale parameters, the channel sounder setup used to conduct the measurements also allows the extraction of spatial information at the receiver in the form of a temporally and spatially resolved list of discrete propagation paths and their powers. Using this information, angular power profiles (APP) and root mean square (RMS) of the Azimuth angle spread of arrival (ASA) have been estimated.

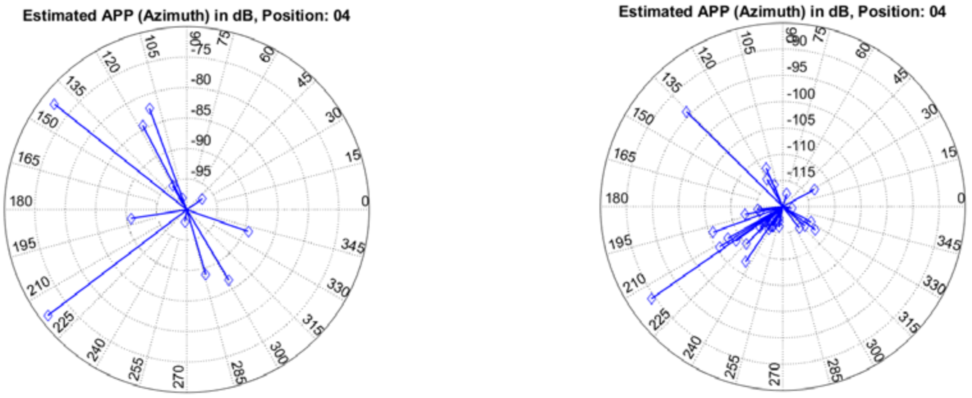


Figure 1: Angular power profiles at measurement point 4 in scenario 2, 3.7 and 28 GHz, LOS

In Figure 1, the angular power profiles in line of sight condition for two measurements at the same measurement point in scenario 2 at 3.7 (left) and 28 GHz (right) are shown. At both frequencies, the LOS component can be seen to impinge at the receiver from around 215° with a strong multi-path component at around 140°. Several more multi-path components are visible in the APPs with similar angles of arrival at both frequencies. Two strong components arriving at the receiver from 285° and 300° are only apparent at 3.7 GHz. The larger number of estimated multi-path components at 28 GHz can be explained by the fact that the temporal resolution was five times higher than at 3.7 GHz due to measurement bandwidth limitations. It should be noted that while most APPs shared a high similarity in terms of angles of arrival between 3.7 and 28 GHz, somewhere also quite different in terms of angles of multi-path components other than the line of sight component. This hints towards a frequency selectivity of the reflective elements in the scenario and is especially visible in non-line of sight conditions as displayed in Figure 2.

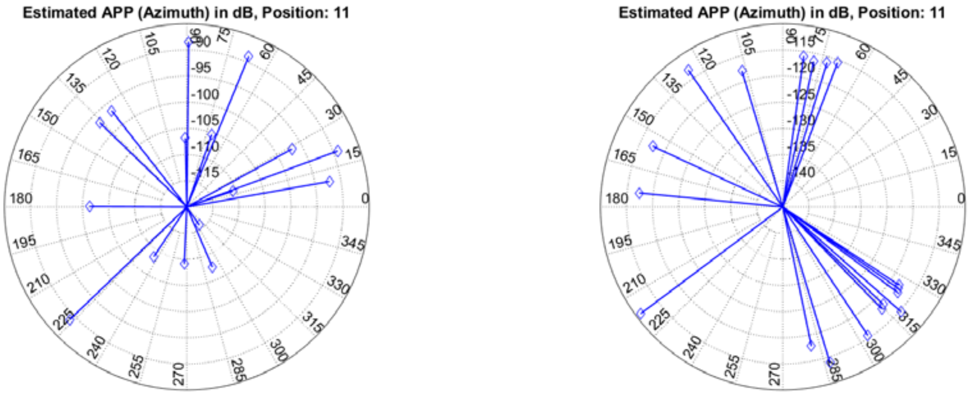


Figure 2: Angular power profiles at measurement point 11 in scenario 3, 3.7 (left) and 28 GHz (right), NLOS

While a strong component impinging at around 220° can be seen for both frequencies, the other angles of arrival do not share a strong similarity. At 3.7 GHz, clusters of multi-path components can be seen around 15° and 135° with several more components at 65, 90, 180 and 270°. On the other hand, at 28 GHz, clustered multi-path components are received around 75° and 300° together with other components at 105, 130, 155 and 175°.

Based on the estimated angular power profiles, statistical parameters of the direction of arrival information in the form of the RMS azimuth spread of arrival (ASA) were calculated. In Table 2, the mean μ_{ASA}, median m_{ASA}, standard deviation σ_{ASA} and 95%-quantile $Q_{ASA,95}$ for the RMS ASA are summarized.

Table 2: Statistical parameters of ASA

Scenario	Storage Area (Scenario 2)				Shop Floor (Scenario 3)			
Frequency	3.7 GHz		28 GHz		3.7 GHz		28 GHz	
Condition	LOS	NLOS	LOS	NLOS	LOS	NLOS	LOS	NLOS
$\boldsymbol{\mu_{ASA}}(^\circ)$	39.12	59.37	30.43	45.37	36.78	66.39	47.02	58.34
$\boldsymbol{lg}(\boldsymbol{\mu_{ASA}})$	1.59	1.78	1.48	1.66	1.56	1.82	1.67	1.88
$\boldsymbol{m_{ASA}}(^\circ)$	37.00	61.28	30.04	48.42	35.35	66.34	31.97	48.25
$\boldsymbol{\sigma_{ASA}}(^\circ)$	32.55	23.60	11.92	17.18	3.77	20.01	40.52	31.29
$\boldsymbol{Q_{ASA,95}}(^\circ)$	82.34	87.87	49.59	67.95	42.36	100.23	114.76	126.31

As expected, it can be seen that the mean angular spread is higher in NLOS condition than in LOS condition. It has to be noted though that in scenario 2, very few measurement points were in NLOS condition, and in scenario 3, very few were in LOS condition. The 3GPP TR 38.901 *Indoor Factory* channel model lists the mean RMS ASA in LOS condition as

$$\lg(\mu_{ASA}) = -0.18\log_{10}(1+f_c)+1.78$$

where f_c is the carrier frequency in GHz, which calculates to 1.66 at 3.7 GHz and 1.52 at 28 GHz. In NLOS condition, the RMS ASA is not frequency depended and reported with a value of 1.78. Comparison with the measurement results of scenario 2 in LOS reveal a slightly lower RMS ASA than in the 3GPP model which can be explained by the relatively low density of the environment in this scenario. For scenario 3 in NLOS condition, the RMS ASA evaluated for the measurements is slightly higher than the 3GPP model, highlighting the high number of reflective scatterers in the environment.

2.2 Evaluation of measurements at 300 GHz

Besides the measurements at 3.7 and 28 GHz, most of the measurement points in scenario 2 (Storage Area) and some in scenario 3 (Shop Floor) were also characterized at 300 GHz. Instead of using a Virtual uniform circular array antenna (VUCA) with omni-directional antenna element, spatial information was captured by rotating the receiver in 15° steps from -180 to 180°. This step size was chosen to coincide with the half-power beam width of the sectoral E-plane horn antenna. For each angle, 250 sequences of the periodic 2 GHz wide channel sounding signal were captured for a total measurement duration of 25 ms per angle. The transmitter with its directional open waveguide antenna was aligned with each measurement point. The sequences per measurement angle are regarded as a sequence set. Assuming that the radio channel is static over the measurement time of 25 ms, a phase drift correction and subsequent averaging can be performed for each set. Afterwards, the averaged phase-corrected set is correlated with a pre-recorded back-to-back calibration, generating a CIR snapshot per measurement point and angle. The path gains of all estimated paths are summed-up angle wise and binned in 15° steps from -180 to 180° corresponding to the measurement angles. Figure 3 shows the measurement setup on the shop floor in scenario 3.

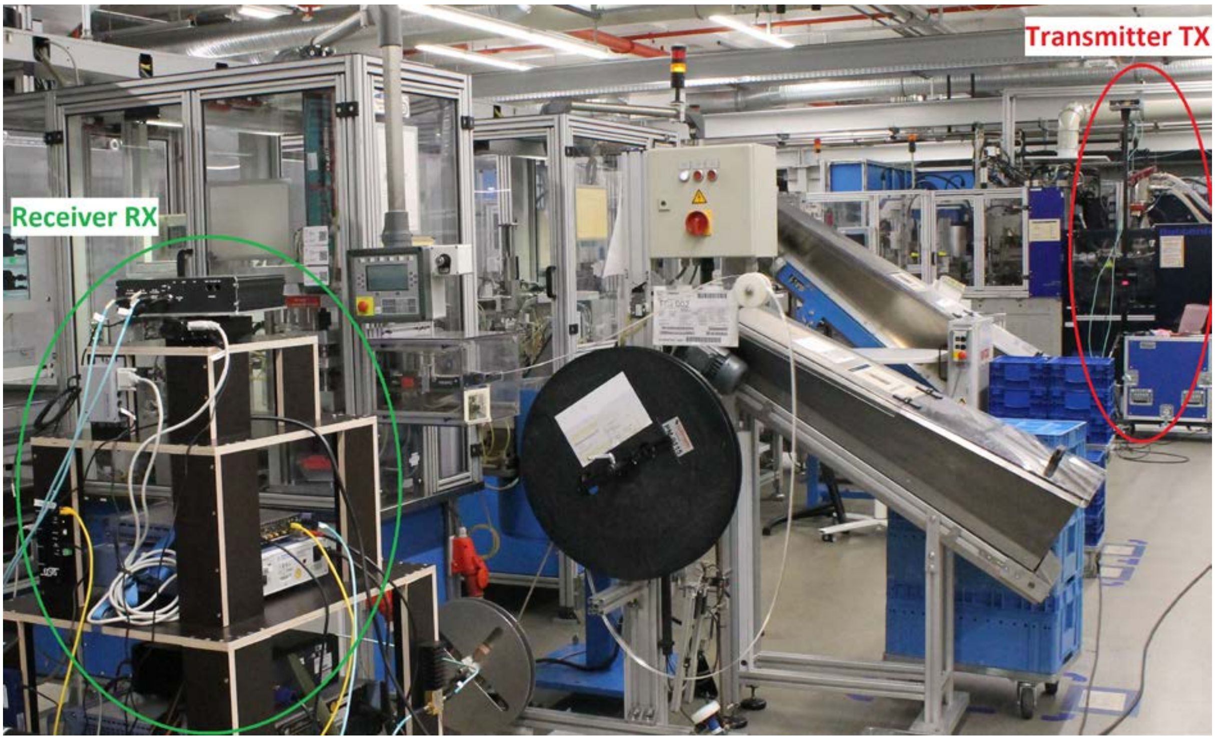


Figure 3: 300 GHz measurement setup in scenario 3

In order to estimate large scale parameters, instantaneous power delay profiles (IPDPs) are derived from the CIR snapshots. Paths within each measurement point's compendium of IPDPs are estimated by identifying local maxima within the compendium through a two dimensional peak search algorithm. A synthesized omni-directional IPDP for each measurement is obtained by superimposing the estimated paths within each compendium of IPDPs. Similar to the evaluation at 3.7 and 28 GHz, an absolute threshold of 10 dB above the noise floor is applied.

Figure 4 and Figure 6 illustrate the OMNI-IPDPs for all measurement points in scenarios 2 and 3. All figures report a multitude of detected paths for each measurement. It can be noted that, for all measurements that meet LOS or OLOS (Obstructed line of sight, the path between transmitter and receiver is partially blocked, but not completely) conditions, the strongest detected path corresponds to the LOS path. Measurement point 2 of the storage and prototyping area scenario and measurement point 2 of the production shop floor scenario represent the only NLOS condition measurements and both reveal stronger multipath components than LOS. With regards to the number of detected paths, the metal hallway scenario's results show the highest amount of multipath components following from the scenario's composition. All measurements denote an exceptionally high dynamic range that finds its peak at about 60 dB.

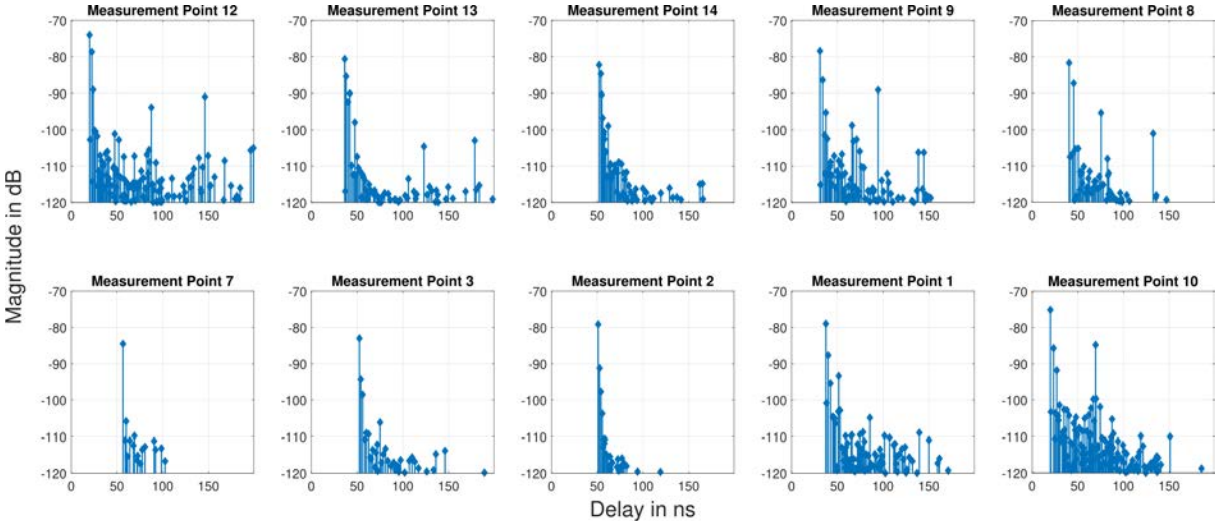


Figure 4: Synthesized omni-directional IPDPs for all measurements in scenario 2

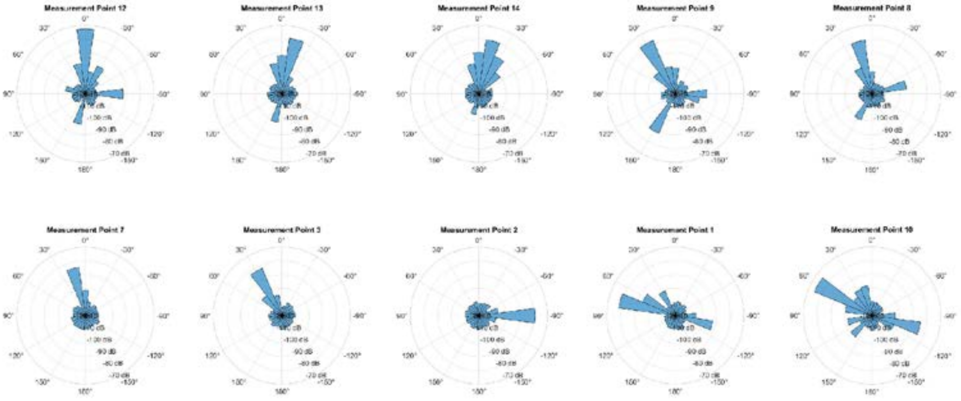


Figure 5: Direction of received power for all measurements in scenario 2

Figure 5 and Figure 7 visualize the direction of received power for all measurements. Therefore, the path gains of all estimated paths are summed up angle-wise and displayed in polar coordinates with θ representing the approached angles (-180 to 180° in 15° steps) and the radius representing the corresponding power. The figures clearly point out the angle-of-arrival (AOA) of the strongest paths. For measurements that meet LOS or OLOS condition, the strongest path denotes the LOS path. For measurements that meet NLOS condition, the direction of the strongest multipath components can be identified. All measurements validate the assumptions on received path directions based on scenario composition.

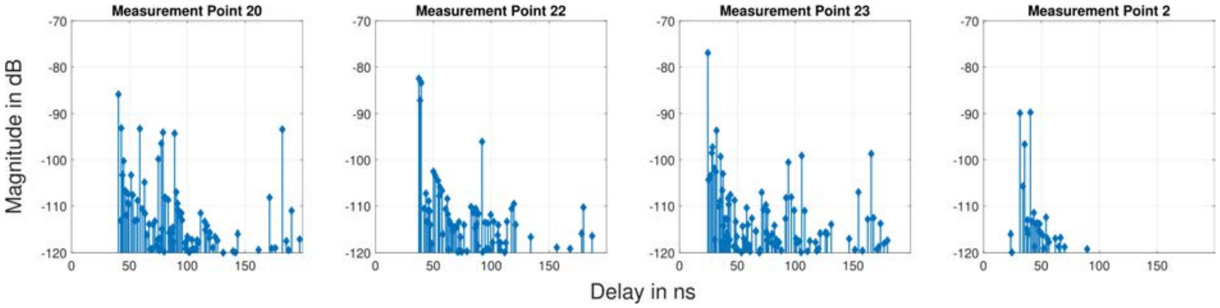


Figure 6: Synthesized omni-directional IPDPs for all measurements in scenario 3

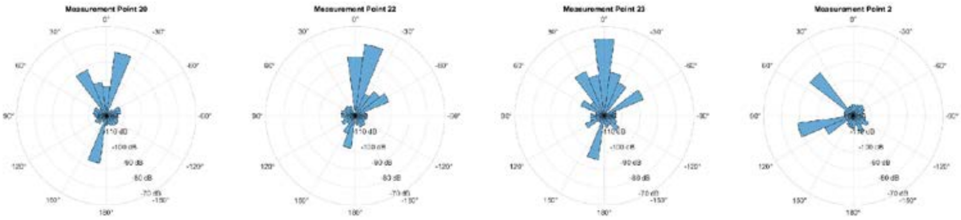


Figure 7: Direction of received power for all measurements in scenario 3

Depending on measurement scenario and measurement position, the presence of multipath components varies from low to high. To quantify both LOS and multipath components' proportion of all measurements within a scenario, the percentage of the first three strongest paths and the residual power are listed respecting LOS-condition in Table 3. All LOS measurements report that about 70% of the whole accumulated power is in the first strongest received path. Regarding OLOS or NLOS measurements, the values vary for the scenarios. Whereas, for the storage and prototyping area, the bulk of the accumulated power is in the first path, for the production shop floor scenario the power is shared quasi equally with around 40% between the first and second strongest path.

Table 3: Percentage division of path gain for all measurements within a scenario considering LOS condition. Subdivision of path gain in strongest path (First Path (FP), Second Path (SP), Third Path (TP)) and Residual Power (RP)

Scenario	LOS Cond.	FP (%)	SP (%)	TP (%)	RP (%)
Scenario 2	LOS	72	17	5	6
	OLOS	87	7	3	4
	NLOS	92	6	1	1
Scenario 3	LOS	69	6	5	22
	OLOS	43	36	15	6
	NLOS	44	43	9	4

With regards to further channel parameters, such as total path loss (PL), root-mean-square (RMS) delay spread (DS) and angular spread (AS) Table 4 and Table 5 summarize the channel parameters for each measurement within a scenario. For DS and AS estimation, an additional evaluation threshold of 30 dB below the strongest component has been applied.

The determined threshold for DS and AS evaluation is well above the considered noise threshold. The AS is calculated according to 3GPP TR 138 901.

Table 4: Summary of channel parameters for each measurement in scenario 2

Measurement Point	LOS Condition	Path Loss (dB)	Delay Spread (ns)	Angular Spread (°)
12	LOS	95.5	16.0	16.0
13	LOS	101.7	9.9	10.2
14	LOS	102.2	2.3	7.3
9	OLOS	100.2	17.0	29.3
8	LOS	103.3	10.6	18.6
7	OLOS	107.4	2.9	1.8
3	LOS	105.5	1.8	1.1
2	NLOS	101.8	0.5	-
1	LOS	101.1	4.7	21.3
10	LOS	97.1	14.6	36.7

Table 5: Summary of channel parameters for each measurement in scenario 3

Measurement Point	LOS Condition	Path Loss (dB)	Delay Spread (ns)	Angular Spread (°)
20	LOS	105.4	39.7	63.3
22	OLOS	101.9	9.5	19.6
23	LOS	99.6	13.4	13.4
2	NLOS	109.2	5.0	31.2

For the storage and prototyping scenario, the DS varies from 0.5 ns to 17 ns and the AS from 1.1° to 36.7°. DS and AS spread values proceed equally, meaning that for increasing or decreasing DS values, the AS spread behaves likewise. At measurement point 2, just a single path impinging from a single direction was evaluated. For the production shop floor scenario, the DS varies from 5.0 ns to 39.7 ns and the AS from 13.4° to 63.3°. Also in this scenario, the DS and AS spread proceed equally. In comparison to the prior discussed scenario, the variations of DS and AS are significantly higher.

Considering path loss values throughout the scenarios, an average PL of 101.6 dB was obtained for the storage and prototyping area scenario and an average PL of 104.0 dB for the production shop floor scenario.

2.3 Evaluation of indoor-to-outdoor measurements

Both from a regulatory point of view and for security concerns, it is desirable that as little RF energy as possible is radiated outside of the targeted industrial environment. Due to the limited available spectrum for private 5G networks, especially in areas with a large number of neighboring potential network operators, it is vital to keep inter-site interference to a minimum. Additionally, due to the potentially confidential nature of the transmitted data and the safety of the machines that are controlled through the private 5G network, the operator might prefer that none of the signals are receivable outside of the intended area.

In order to quantify the indoor to outdoor radiation at 3.7 and 28 GHz, the transmitter was placed at Tx position 1 as indicated in Figure 8. The receiver was then placed at three positions on a road just north of the production hall. Measurement position one was in line with the transmitter position and positions two and three were 5 meters to the east and west.

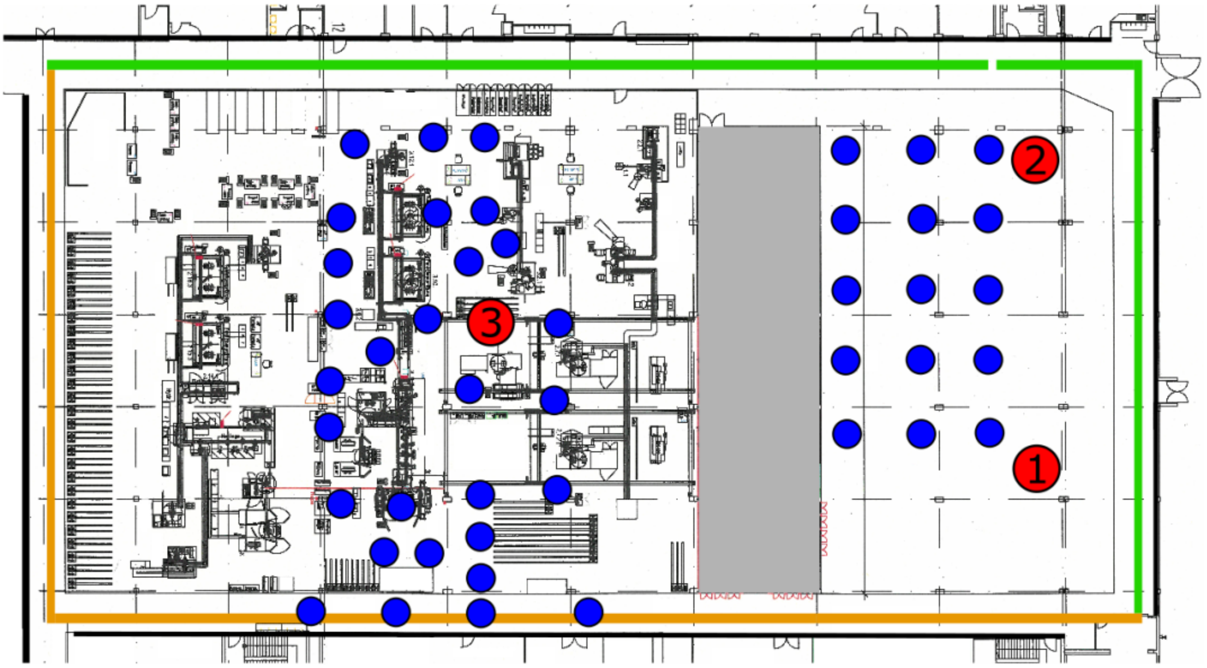


Figure 8: Floor plan of the measurement scenario

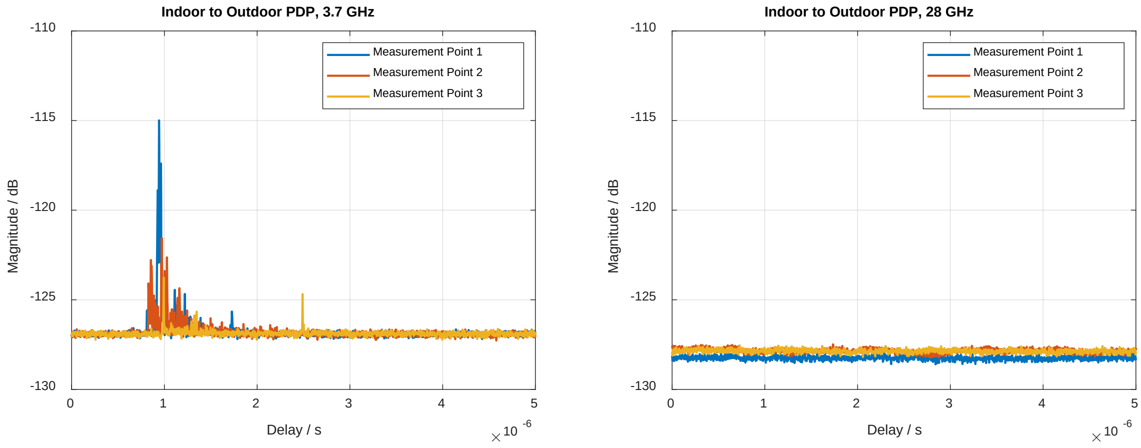


Figure 9: Indoor-to-Outdoor power delay profiles at 3.7 and 28 GHz

Figure 9 shows the power delay profiles of the indoor to outdoor measurements at 3.7 and 28 GHz. At 3.7 GHz, some reception is possible, with a maximum SNR of 12 dB at measurement point 1, 5 dB at measurement point 2 and 3 dB at measurement point 3. On the other hand, at 28 GHz, no reception of the sounding signal was possible at all measurement positions.

These results show a considerable attenuation of the signal during indoor-to-outdoor transmission in an industrial environment. While some signal was received at 3.7 GHz with a maximum SNR of 12 dB, the RF levels are not relevant in terms of interference. With sophisticated measurement equipment and signal processing, it might be possible to recover some of the signaling and/or data. This is not possible at 28 GHz where, even with the high processing gain of the channel sounder setup, no signal was received outside of the hall.

3 Extrapolation of link parameters from connectivity map

One of the activities reported in deliverable D3.1 focused on the learning of the connectivity map of some key performance indicators such as the received signal strength (RSS). In D3.1, we proposed a strategy exploiting higher order graph structures, such as simplicial complexes, to extrapolate link parameters from sparse measurements. The method illustrated in D3.1 was based on the inference of a simplicial complex of order two, i.e., containing vertices, edges and triangles, from a training set containing the parameters of interest. Then, based on the inferred simplicial complex, the method was able to reconstruct the connectivity map over a region of interest, based on a limited number of measurements.

In this deliverable, we propose two advancements with respect to what we have reported in D3.1. First of all, we developed a new method to learn the connectivity map based on the *line graph*. In this graph the nodes correspond to the communication links between transmitter/receiver (Tx-Rx) pairs and the adjacencies between nodes catch the correlation between pairs of Tx-Rx performance measures. The method starts inferring the line graph from the observations of the RSRP between pairs of transmitters/receivers. The overall connectivity map is then reconstructed from the observations of a small set of points using sampling theory on graphs [TSI16]. Being based on graphs, rather than simplicial complexes, the proposed method is simpler to implement than the method presented in D3.1.

The second advancement refers to the new algorithm developed to infer link parameters such as throughput, exploiting the connectivity map inferred from RSRP measurements. In such a case, the line graph has the same structure as before, but the goal is to reconstruct a map of throughput, instead of RSRP, from sparse measurements. The activities reported in this section are the result of a collaboration between ITRI, who provided the data sets, and SAP, who developed the algorithms. More specifically, two data sets were produced, namely received signal strength and throughput, using the RANPLAN tool running on the 3D model of the IMTC plant in Taiwan.

3.1 Line graph based method

Our first goal, as in D3.1, is to reconstruct the overall EM field (RSRP or throughput) from a small subset of samples. The proposed method consists of two main steps: 1) learn the line-graph from a training set; and 2) recover the overall performance indicator of interest from the line graph and from sparse measurements. The difference with respect to what presented in D3.1 are that: 1) to simplify the method, we infer a line graph instead of a simplicial complex; 2) we extrapolate a link parameter, such as throughput, from a line graph inferred from RSRP measurements and from sparse throughput measurements. The method is composed of the following two tasks:

1) Line-graph learning

As a first step, we need to associate a graph with the observed field to capture the correlation properties of the measurements and to enable the use of graph signal sampling tools [TSI16]. As an application example, let us consider the RSRP field measurements in the scenario depicted in Figure 10, composed of N_{tx}=5 transmitters, placed in the bottom-right side of the figure, and 986 (potential) receivers located in the colored circles. The colour of each node encodes the RSRP field received at that point, when illuminated from transmitter Tx_1 (red point). We focused without loss of generality on the recovery of the portion of the field in Figure 10 bounded by the red rectangle and composed of N=265 points.

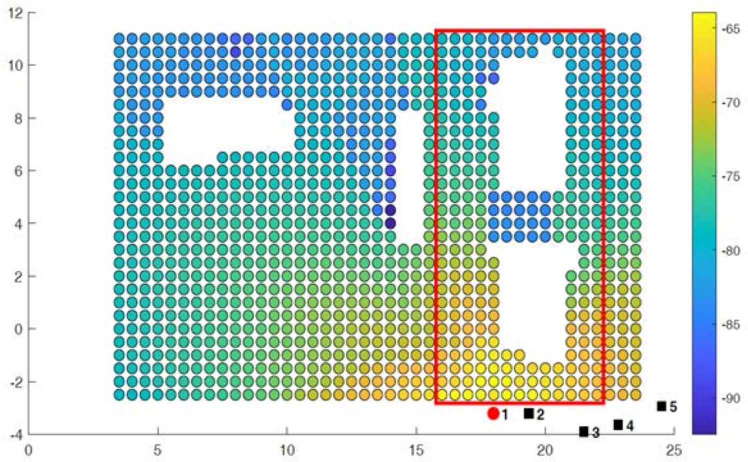


Figure 10: True RSRP

To capture the correlation between the RSRP associated to nearby points, we used a training dataset to infer the structure of a line graph. More specifically, we associate to each edge corresponding to a given pair of transmitters/receivers (Tx,Rx) a node of another graph, called the line-graph, that is composed of a total number of N_t=N·N_{tx} nodes. Two nodes in the line graph are then connected by an edge of the line graph if the RSRP signals observed over the corresponding two edges are similar according to a prescribed rule. More specifically, to connect the nodes of the line graph, we used the following criteria: 1) the pairs of nodes (Tx_i,Rx_j) and (Tx_i,Rx_n) are connected if the area of the triangle with vertices Tx_i,Rx_j,Rx_n is lower than a given threshold α and the RSRP observed at the receivers Rx_j, and Rx_n are quite similar, i.e. the absolute value of their difference is lower than a small coefficient β; 2) the pairs of nodes (Tx_i,Rx_j) and (Tx_m,Rx_n) in the line graph are connected if the area of the square with vertices Tx_i,Rx_j,Tx_m,Rx_n is lower than a threshold α and the absolute value of the difference of the RSRP observed at the receivers Rx_j,Rx_n is lower than a given coefficient β.

The output of the inference step is an adjacency matrix **A,** which encodes which nodes of the line graph are neighbors of which other nodes. Given the inferred adjacency matrix **A,** we build the Laplacian matrix **L** of the line graph, as **L**= **D** - **A**, where **D** is a diagonal matrix with entries $d(i)$ the number of neighbours of each node, for $i = 1, \ldots,$ N_t,. The Laplacian is a symmetric, semidefinite matrix that admits the following eigendecomposition $\mathbf{L} = \mathbf{U}\,\mathbf{\Lambda}\mathbf{U}^T$ where $\mathbf{U}$ is the N_t x N_t matrix with columns the eigenvectors of **L** and $\mathbf{\Lambda}$ is the diagonal matrix containing the associated eigenvalues. Using Graph Signal Processing (GSP) tools [SHU13], the eigenvectors of **L** form a suitable basis to represent signals observed over the nodes of the line graph.

2) Recovery of the EM field

The second step of our strategy uses sampling theory over graphs to recover the signals associated with the EM field (RSRP or throughput) from the observation of a small subset S of N_S samples. We use the Max-Det greedy sampling algorithm [TSI16] to select the sampled nodes. Denote with $\mathbf{x}_S = \mathbf{D}_S\mathbf{x}$ the observed graph signal where $\mathbf{D}_S$ is the selecting N_t x N_t diagonal matrix whose *i*-th diagonal entry is 1 if $i \in S$ and 0 otherwise. If the signal $\mathbf{x} \in$

$\mathbf{R}^{\mathrm{N_t}}$ is bandlimited with bandwidth B, possible ways to retrieve the overall signal from its samples [TSI16] are

$$\mathbf{x} = \mathbf{U}_{\boldsymbol{B}}\,(\mathbf{U}_{\boldsymbol{B}}^{\boldsymbol{T}}\,\mathbf{D}_{S}\,\mathbf{U}_{\boldsymbol{B}})^{-\mathbf{1}}\mathbf{U}_{\boldsymbol{B}}^{\boldsymbol{T}}\mathbf{x}_{S}$$

or, alternatively,

$$\mathbf{x} = \;(\mathbf{I} - (\mathbf{I} - \mathbf{D}_{S})\,\mathbf{U}_{\boldsymbol{B}}\mathbf{U}_{\boldsymbol{B}}^{\boldsymbol{T}})^{-\mathbf{1}}\mathbf{x}_{S}$$

where $\mathbf{U}_{\boldsymbol{B}}$ is a matrix whose columns are the B eigenvectors of $\mathbf{L}$ corresponding to the frequency indices associated with the bandwidth of $\mathbf{x}$.

As an example of application, we recover the RSRP associated to the transmitter placed in the red node of Figure 10 from the observation of the RSRP associated to the transmission from the four transmitters represented by the black squares in Figure 10. Using sampling on graph theory, we reconstruct the RSRP map from the observations of only N_S=6 samples, assuming α=15 and β=2. We identify with different colours the 6 Tx-Rx pairs of sampled RSRP. Specifically, the position of each measurement is represented by a square: the contour of the square around each sampled receiver (node) has the same colour of the associated transmitter. Adopting our proposed sampling selection method, the following Tx-Rx pairs have been identified: (2,134), (2,167), (3,117), (4,128), (4,134), (5,265). Then, in Figure 11, we report the field of RSRP values reconstructed within the red rectangle, when the region is illuminated by the first transmitter. By comparing Figure 10 and Figure 11, we can assess the goodness of the proposed method in recovering the true field by using only 6 samples. Finally, in Figure 12 we report the normalized sum of squared errors (NSE) in the recovery of the RSRP value, versus the number of samples. We can observe that the error remains very small also when using a limited number of samples.

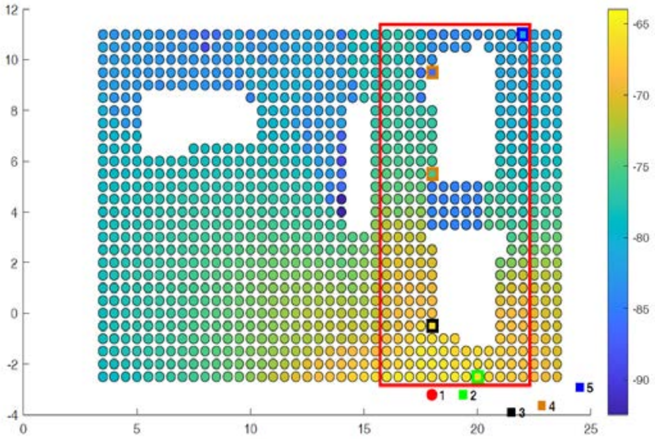


Figure 11: Recovered field of Tx_1 with $B=N_S=6$

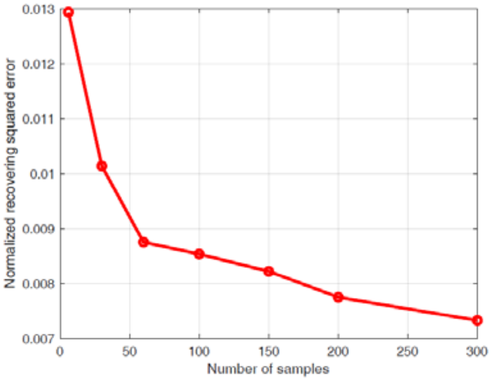


Figure 12: NSE versus number of samples with B=6

We also tested the robustness of our method in recovering the RSRP values obtained using the line graph learned from one data set, but reconstructing the RSRP field from values belonging to different data sets, obtained using slightly different transmitters' positions. Learning the line-graph from the first dataset and observing samples from the second dataset, we recovered the RSRP associated to transmitter 6 (red node in Figure 13) assuming α=13, β=2. The optimal pairs (Tx,Rx) in this case are: (7,117), (8,134), (8,167), (9,128), (9,157), (9,265).

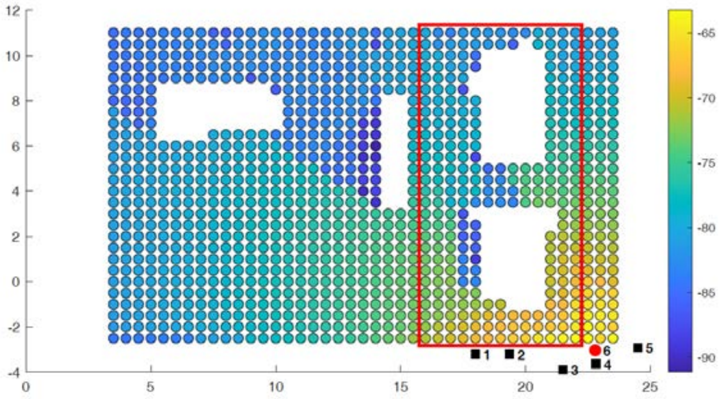


Figure 13: True RSRP field of Tx_6 with $B=N_s=6$

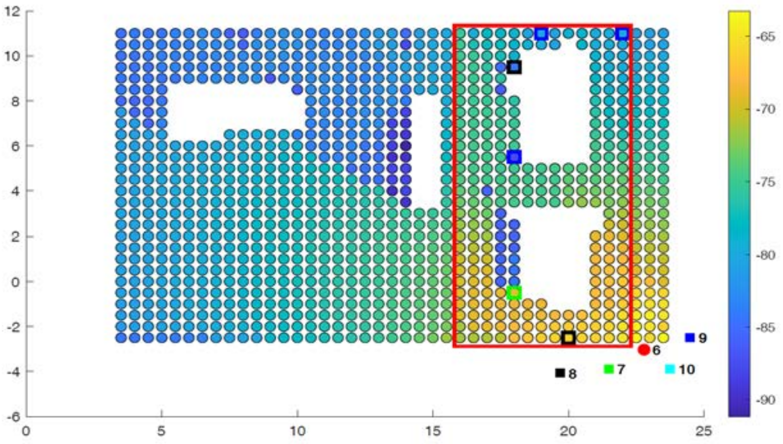


Figure 14: Recovered RSRP field of Tx_6 with $B=N_s=6$

In Figure 13 and Figure 14 we report the true and recovered RSRP associated to the transmitted 6. We can notice the effectiveness of the proposed method in recovering the RSRP field from the observation of a very small number of samples, in the presence of small variations of the transmitters' positions.

3.2 Extrapolation of throughput map from RSRP measurements

The second important modification with respect to D3.1 is that we want to build the throughput map using the line graph built using RSRP measurements and a few samples of throughput. Let us consider the RSRP map associated to Tx_1 illustrated in Figure 15 and the associated throughput reported in Figure 16. We focus on the recovery of the throughput by considering the N=234 receivers present in the red rectangle of Figure 16. We infer the line graph as in the previous section, using only RSRP values, and setting the thresholds as α=16.5 and β=1.75. Then, we build the throughput map starting from the observation of a small subset of throughput measures. In Figure 16 and in Figure 17, we report the true and recovered throughput (in the red rectangle) obtained by observing only N_s=10 samples. The 10 pairs of Tx-Rx sampled throughputs are identified with different colours: the squares around each sampled receivers (nodes) have the same colour of the associated transmitter. Specifically, we observe the following pairs of Tx-Rx: (1,81), (1,100), (1,104), (1,228), (3,93), (3,96), (3,125), (4,17), (5,22), (5,24). From Figure 16 and Figure 17, we can notice that the proposed method enables a good recovery of the throughput, in spite of the small number of samples even in the case where the line graph is inferred from the RSRP field. Finally, in Figure 18 we report the normalized sum of squared error in the recovery of the throughput signal, associated to transmitter Tx_1, versus the number of samples.
We can observe how the error is very small and decreases as the number of observed samples grows, as expected.

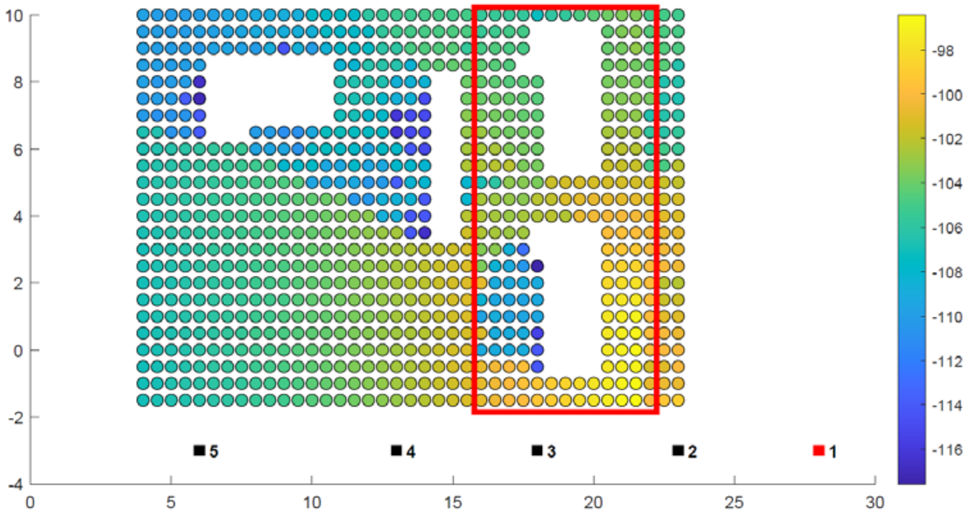


Figure 15: Observed RSRP field of Tx_1

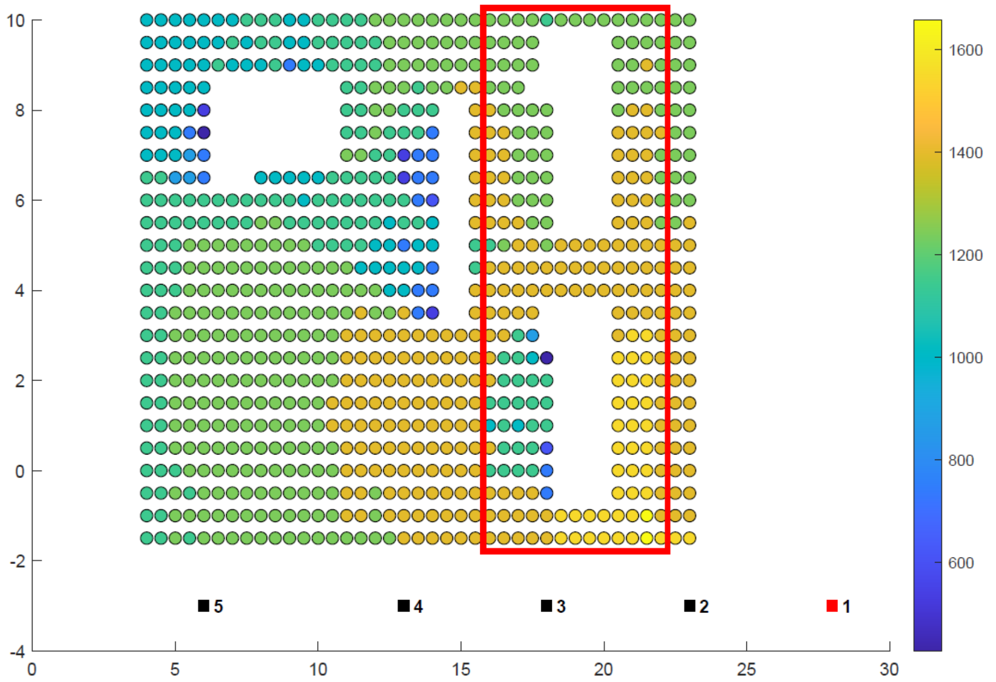


Figure 16: True throughput map associated to Tx$_1$

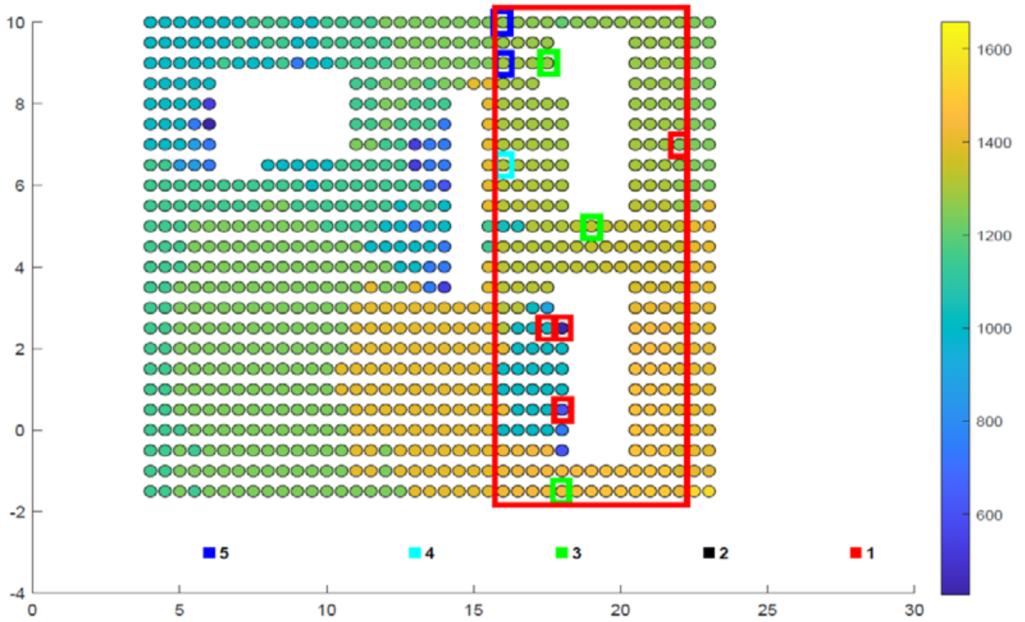


Figure 17: Recovered throughput map associated to Tx$_1$ with B=N$_s$=10

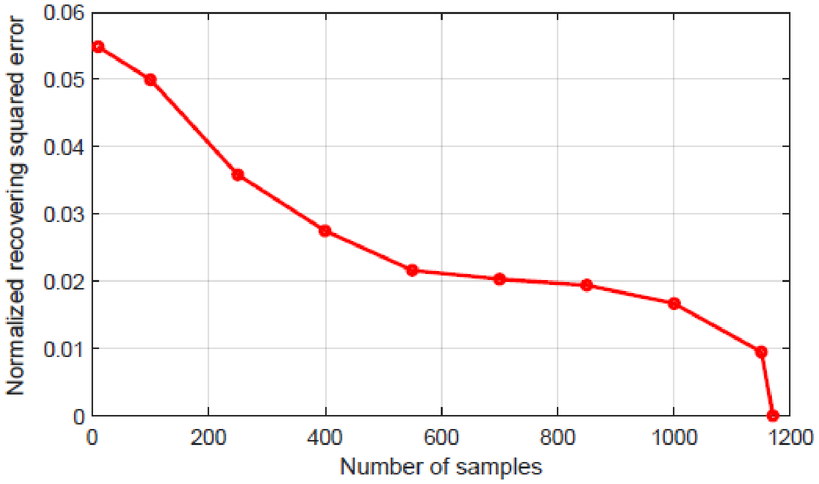


Figure 18: Normalized sum of squared errors in throughput recovery vs. the number of observed samples

4 Optimal service placement based on industrial channel models

In this section, we illustrate the innovations on optimal service placement with respect to the methods reported in deliverable D3.1. The problem is the optimal placement of radio and computation resources in the edge cloud and the association between mobile devices, radio access points and edge servers, including also the optimal routing across the edge cloud. The goal is to find an allocation of resources that makes possible to offload computations from the peripheral devices to the edge cloud, while guaranteeing the desired service delay, which includes both transmission time and computation time. With respect to the method analyzed in D3.1, to cast the proposed strategy in an industrial scenario, as foreseen in 5G-CONNI, SAP developed new algorithms incorporating the channel models developed by HHI through the measurement campaign conducted at Bosch's premises. Secondly, the new optimization problem has now been formulated as the joint optimization not only of the placement and routing variables, but also of the bandwidth allocated to each link and of the computation rate assigned to each task. We also modified the algorithm to better handle the integer constraints associated with the placement and routing variables, in a computationally efficient manner. The numerical results presented at the end of this section show how the new approach improves the previous solution.

4.1 Novel Problem Formulation

In this section, we propose an improvement of the *Joint Service Placement and Request Routing* (JSPRR) algorithm explored in the previous deliverable D3.1. We consider a scenario composed by a set of mobile devices $K = \{1, \dots, K\}$, sending data to either one of the N edge servers available in the industrial site, or to a remote centralized cloud, for processing. The application to be run on the data is supposed to belong to a set of *S* services, labelled as the set $S = \{1, \dots, S\}$, to be deployed at either some of the edge servers or at a remote centralized cloud. Each connected device k, with $k \in K$, sends an amount of bits b_s^k to an Access Point (AP), with the request of running a service s, with $s \in S$, involving the processing of a workload w_s^k, at the assigned server $n \in N$. Each AP is either co-located or simply connected to an edge server, endowed with communication, computation and storage capabilities. Our objective is to allocate services and network resources in order to minimize the overall service delay associated to all requested services. This total delay is computed as the sum of the following terms: the communication time spent to send data from the peripheral device to the AP, the time necessary to convey the data from the AP to either a server in the edge cloud or to the centralized cloud, and the computation time necessary to process the data received by the server to run the required service. The edge server has limited computational capabilities and it typically runs in a multitask fashion, so that only a portion of the CPU time is allocated to each service.

More specifically, the communication delay D_{kn}^c is computed using the Shannon expression of the transmission rate R_{kn}:

$$D_{kn}^{\mathrm{c}} = \frac{b_s^k}{R_{kn}} \qquad\qquad R_{kn} = B_{kn} \log_2(1 + \frac{P_k h_{kn}}{N_0 B_{kn}})$$

where B_{kn} is the portion of the overall available bandwidth allocated by a server n to enable the communication link with a device k; this bandwidth is now a new variable to be optimized, in contrast to the algorithms reported in D3.1, where the bandwidth was fixed for all links and all services.

The processing delay D_{kn}^p is computed as

$$D_{kn}^{\mathrm{p}} = \frac{w_s^k}{f_{kn}}$$

where f_{kn} is the portion of CPU clock rate, expressed in Gbps, allocated by server n to enable the computation of data offloaded by device k. Whenever necessary, to respect the service delay constraints, some requests can be routed to the central cloud. In such a case, the peripheral device gets access to the edge network through one of the available APs and then the data are routed through the core network using wired links. The communication delay experienced when connected to the central remote cloud is denoted by D_C. We model this value as a constant, because it is outside of our optimization procedure, which concerns only the data belonging to the edge cloud

In summary, the total service delay can be computed as:

$$\mathcal{D}_{kn}^{\mathrm{tot}} = D_{kn}^{\mathrm{c}} + D_{kn}^{\mathrm{p}} + D_{\mathcal{C}}$$

where the last term appears only when the request is routed to the centralized cloud. Our strategy is to limit as much as possible the access to the centralized cloud and enable all computations within the edge cloud, to keep the service delay under control.

The overall optimization problem is formulated as follows:

$$
\begin{array}{lll}
\mathbf{P}_0: \min\limits_{\Phi} & \sum\limits_{k\in\mathcal{K}} \sum\limits_{n\in\mathcal{N}_\mathrm{C}} y_{kn}\mathcal{D}_{kn}^{\mathrm{tot}} & \\
\text{s.t.} & a)\ \sum\limits_{n\in\mathcal{N}_k\cup\mathcal{C}_L} y_{kn} = 1, & \forall k\in\mathcal{K}; \\
 & b)\ y_{kn} = 0, & \forall k\in\mathcal{K}, n\notin\mathcal{N}_k\cup\mathcal{C}_L; \\
 & c)\ y_{kn} \le x_{sn}^{k} & \forall k\in\mathcal{K}, n\in\mathcal{N}; \\
 & d)\ \sum\limits_{s=1}^{S} x_{sn}\, m_s \le M_n & \forall n\in\mathcal{N}; \\
 & e)\ \sum\limits_{k=1}^{K} y_{kn}\, f_{kn} \le F_n & \forall n\in\mathcal{N}; \\
 & f)\ \sum\limits_{k=1}^{K} y_{kn}\, B_{kn} \le W_n & \forall n\in\mathcal{N}; \\
 & g)\ x_{sn}\in\{0,1\} & \forall s\in\mathcal{S}, n\in\mathcal{N}; \\
 & h)\ y_{kn}\in\{0,1\} & \forall k\in\mathcal{K}, n\in\mathcal{N}_C; \\
 & i)\ f_{kn}\ge 0 & \forall k\in\mathcal{K}, n\in\mathcal{N}; \\
 & j)\ B_{kn}\ge 0 & \forall k\in\mathcal{K}, n\in\mathcal{N};
\end{array}
$$

$$
\text{where } \Phi = \left[\{x_{sn}\}_{s,n}, \{y_{kn}\}_{k,n}, \{f_{kn}\}_{k,n}, \{B_{kn}\}_{k,n}\right].
$$

Φ represents the set of optimization variables and includes: the (integer) service placement variables, the routing variables, the bandwidths allocated to each link and the computing rates associated to each task (service). Each edge server is supposed to operate in a multi-task fashion, so that a percentage of its computing capability is allocated to each task.

The above problem is rather complex as it is a mixed-integer optimization problem, involving both integer and real optimization variables. In the following, we show how we handled this problem in a computationally efficient manner.

4.2 Proposed Solution

The first step to solve the proposed optimization problem, is to relax the binary variables x_{sn} and y_{kn} to be real (with new constraints instead of *g* and *h,* as already done in D3.1). Starting with initial constant values for CPU frequencies f_{kn} and bandwidth portions B_{kn}, then computing a fixed value for the total delay, we can solve a Linear Programming algorithm obtaining a fast, low complexity real valued solution for service placement and request routing variables. Leveraging the conditional gradient method, also known as the Frank-Wolfe (FW) algorithm, which provides a computational efficient alternative to projected gradient descent method to solve optimization problems, we update the real-valued solution $\widehat{z}_i$ at each iteration i, following the rule:

$$
\widehat{z}_{i+1} = \widehat{z}_i + \eta_i\left(z_i^* - \widehat{z}_i\right),
$$

where $\eta_i \in [0,1]$ is the step size, guaranteeing a feasible updated solution, and z_i^* is the solution of the relaxed optimization problem **P₀**. To further improve the optimization, after the linear optimization solution and the gradient updated step to obtain new placement and routing solutions, we solve two sub-problems to optimize communication and computation resources. With respect to the communication resources, we optimize the allocation of the portion of the total bandwidth, for each user device k, already routed to server n, which already stores data to deliver the required network service. The real updated solution of problem **P₀** allows us to divide the resource allocation optimization problem between different edge nodes. Then, we solve a convex optimization problem to find the optimal bandwidth allocation, for each server n, expressed as

$$\begin{aligned} \min_{\widetilde{\mathbf{B}}} \quad & \sum_{k=1}^{K} \frac{\tilde{b}_s^k}{\widetilde{B}_k \log_2(1+\tilde{a}_k/\widetilde{B}_k)} \\ \text{subject to} \quad & a)\ \widetilde{B}_k \geq 0, \qquad \forall k \in \mathcal{K}; \\ & b)\ \sum_{k=1}^{K} \widetilde{B}_k \leq W_n; \end{aligned}$$

where $\tilde{a}_k = y_{kn} P_k h_{kn}/N_0$, $\tilde{b}_s^k = y_{kn} b_s^k$ and $\widetilde{B}_k = y_{kn} B_{kn}$.

This sub-problem is convex and then we can easily find the optimal solution by using a numerically efficient algorithm, taking advantage of convex optimization tools.

In addition to communication resources, each edge server n, can also optimally allocate the CPU rates of every user devices already assigned to it, in order to run the requested application. This sub-problem can be formulated, for each server n, as:

$$\begin{aligned} \min_{\mathbf{f}} \quad & \sum_{k=1}^{K} \frac{\tilde{w}_s^k}{\tilde{f}_k} \\ \text{subject to} \quad & a)\ \tilde{f}_k \geq 0, \qquad \forall k \in \mathcal{K}; \\ & b)\ \sum_{k=1}^{K} \tilde{f}_k \leq F_n; \end{aligned}$$

where $\tilde{w}_s^k = y_{kn} w_s^k$ and $\tilde{f}_s^k = y_{kn} f_k$.

This optimization problem is also convex and its optimal solution can be derived in closed form as:

$$f_k^* = \frac{\sqrt{\tilde{w}_s^k} F_n}{\sum_{k=1}^{K} \sqrt{\tilde{w}_s^k}}$$

At each iteration, we alternate the updated solution obtained with the conditional gradient descent method with these two algorithms to find the transmission and computation rates. The algorithm stops when a specified threshold ε for accuracy is reached. The accuracy threshold is computed in terms of improvement of the objective function with respect to the initial objective function. After the last iteration, the possibly real-value solution must be rounded, if necessary, to obtain binary values for the placement and routing variables. If the number of actual values is small enough (in terms of the available computational capacity), we apply an exhaustive search algorithm to find the optimal combination, whenever this exists. Otherwise, we apply the *Approximation Algorithm* already illustrated in deliverable D3.1. Although the proposed randomized rounding algorithm has guarantees to obtain the optimal integer solution, it could require an excessive time to converge, especially for a large number of unknowns. To tackle this problem, we propose a heuristic solution that iteratively verifies which capacity limits have been violated. The first step of the proposed iterative rounding solution is to apply the Approximation Algorithm to the real valued solution only one time. Then, for each edge server, we check if the capacity constraints have been violated. When this happens, the request is re-routed to a neighbor edge node, if any with available sufficient capacities, or to the central cloud server. The request to be re-assigned is chosen following a specific sorting, depending on how much the bound has been violated. The proposed heuristic sorting mechanism is designed in order to minimize the impact on the objective function. At each iteration, the solution is updated considering the new assignment and the algorithm double checks again for other possible bound violations. The iterations stop when all constraints are satisfied.

4.3 Numerical Results

To assess the performance of the proposed optimized solution, named *JSPRR-RA,* we implemented, as a benchmark, a solution, named *MA (Matching Algorithm)*, which combines the conventional association between mobile devices and APs based on the SNR at the AP, and a matching theory algorithm [GAL62], to distribute the tasks among the available AP's and edge servers. We also compare the new method with the solution proposed in D3.1, here indicated as *JSPRR*. We considered three types of virtual machines as possible service network requests: micro, small and large, each type having its own requirements in terms of data units sent by the sensors to the edge servers, workload to be processed at the server and storage size of the virtual machines and data to be stored at the server to run the required services.

To cache the content needed by each service, we take into account the popularity profile for $S = 200$ services. We generated the popularity profile according to the Zipf distribution with shape parameter 0.8, that is a typical value for several network services. The considered area is 200m x 150m; 200 sensors are randomly distributed inside this area, divided between devices that can get access through LoS and NLoS wireless links. The portion of LoS vs. NLoS link is made variable to check performance.

The first group of numerical results involves different computation capabilities (number of CPU cycles per second) available at the edge server. In Figure 19, we compare the percentage of tasks that are allocated to the centralized cloud, vs. the computational capability of each edge server (assumed to be the same for each server, for simplicity). Clearly, the more CPU capacity is available at the edge, the smaller is the number of requests that gets routed to the centralized cloud. We can see from Figure 19 that, when the processing capability at the edge is scarce, both JSPRR and MA algorithms transfer a significant portion of the tasks to the cloud. Conversely, the joint optimization included in our new JSPRR-REA algorithms is able to reduce this proportion significantly, thus providing a method much more robust to run most of the tasks locally, in the edge cloud, under the same scenario.

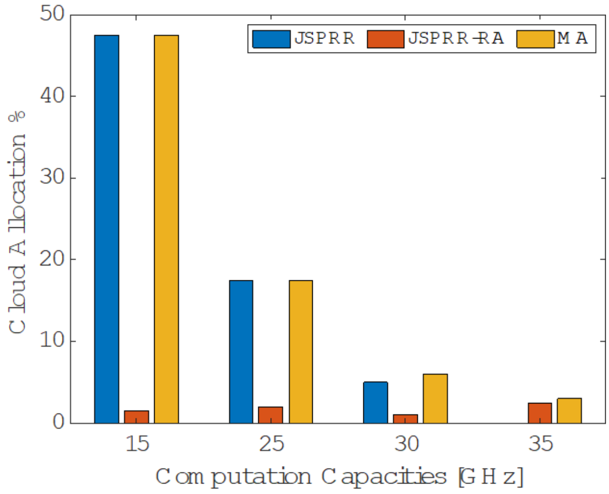


Figure 19: Percentage of sensors' requests routed to cloud, for different computation capacities

This different allocation and resource allocation strategy yields also a clear gain in terms of service delay experienced in offloading the required applications, as shown in Figure 20, where we compare the average service delay, as a function of the computational capabilities available at the edge servers, for the three methods. The figure shows that, again, the new JSPRR-RA offers a significant gain in terms of average service delay with respect to both MA and JSPRR methods. The *JSPRR* algorithm still offers some gain with respect to the SNR-based method, but it is clearly outperformed by the JSPRR-RA algorithm.

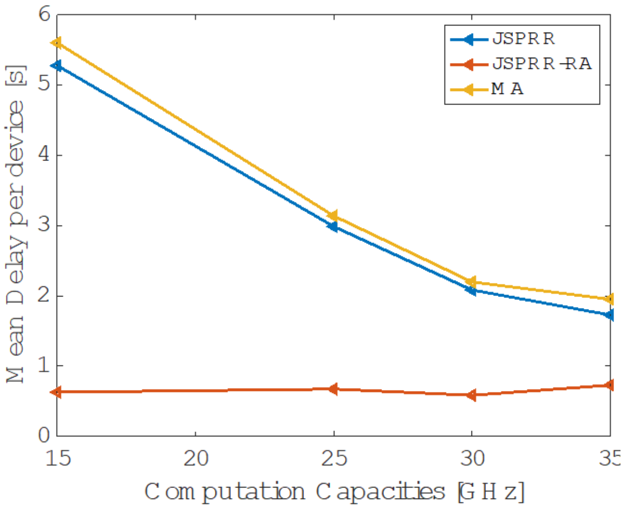


Figure 20: Average device delay, for different computation capacities

To help understanding the results, in Figure 21 we report the proportion of workload offloaded to the edge servers, as opposed to the central cloud. The *JSPRR* algorithm, aiming at minimizing the total latency experienced by each device, routs the heaviest requests to the central cloud, which has a much larger CPU capacity than the edge servers. Only in the last point of in Figure 21, the JSPRR method is able to keep all the tasks running in the edge cloud, as with JSPRR-RA, while the conventional MA method still routs some requests to the central cloud.

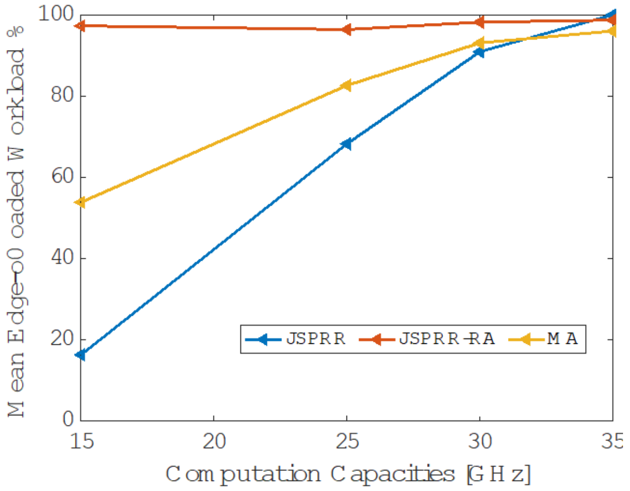


Figure 21: Percentage of required workload processed at the edge, for different server capacities

The second part of numerical results shows the performance as a function of the overall bandwidth available for accessing the edge cloud. In Figure 22 and Figure 23, we show again the percentage of requests routed to the central cloud and the average service delay (as for Figure 19 and Figure 20), vs. the overall bandwidth. We can see that, when the bandwidth is scarce, many requests are routed to the cloud, but this percentage drops as the bandwidth increases. This happens because when the bandwidth is scarce the transmission from the mobile device to the RAP takes more time and then, to satisfy the overall service delay constraint, the server needs to run in a smaller time and this requires the involvement of the centralized cloud, which has much more computational capabilities than the edge servers. We can also see that the new JSPRR-RA method significantly outperforms the other two methods. JSPRR-RA can in fact properly manage the resources to handle most requests within the edge cloud, also in critical situations, thus obtaining a larger reward in terms of average service delay.

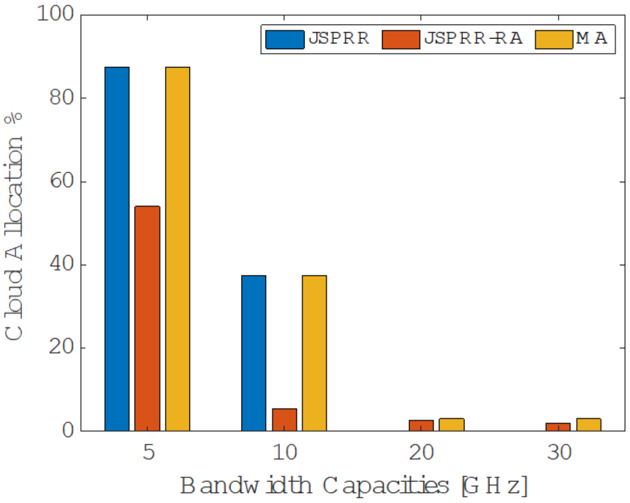


Figure 22: Percentage of sensors' requests routed to cloud, for different bandwidth capacities

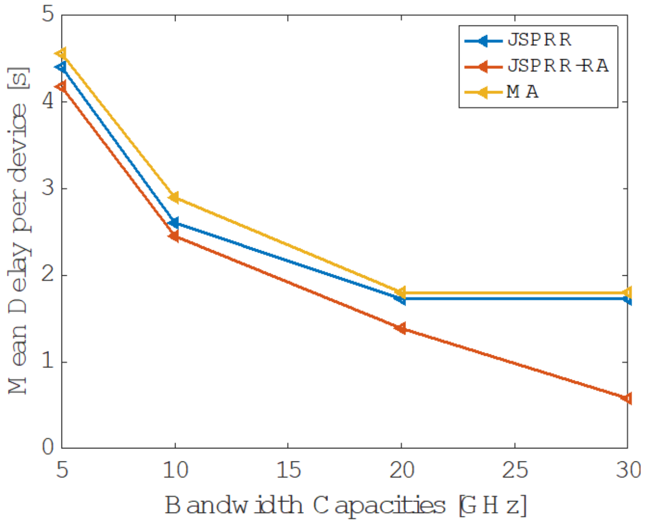


Figure 23: Average device delay, for different bandwidth capacities

5 Network monitoring performance

The private 5G system in 5G-CONNI project consists of RAN, including CPE, MEC and 5G core network. In order to check if the system meets the design critera and have a better knowledge of how the 5G system performs, each of the network components will monitor the corresponding KPI.

The section includes the following monitoring subtasks:

- RAN monitoring
- Core network monitoring
- Edge cloud monitoring

5.1 RAN monitoring

The new tools to monitor the RAN (Radio Access Network) have been developed by ANI. The RAN is composed of several components such as RU (Radio Unit) and DU/CU (Distributed Unit / Central Unit). RAN will connect to 5GC (5G Core) and managed with NETCONF (Network Configuration Protocol). These elements are shown in Figure 24.

The task of RAN monitoring focuses on the operations, administration and maintenance (OA&M) of 5G DU/CU. The protocol of NETCONF Server is implemented in the DU/CU; and NETCONF Client is in the Management server.

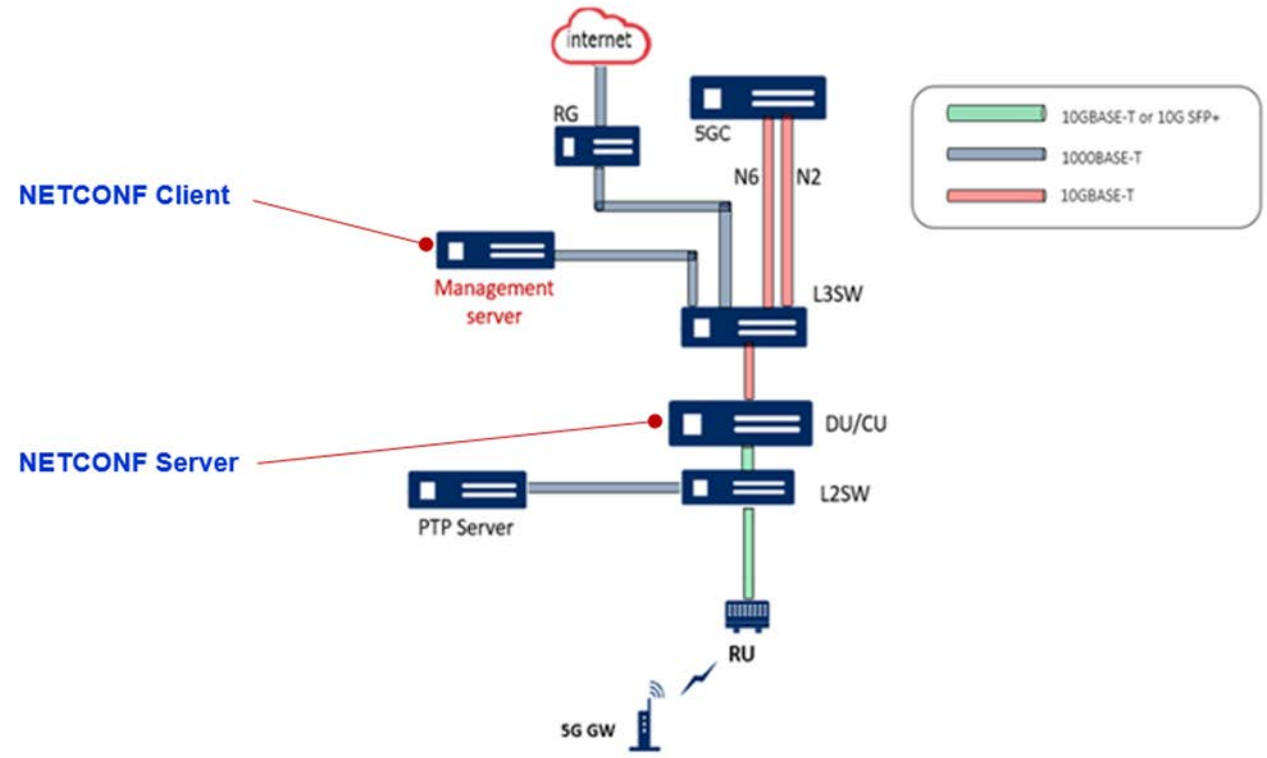


Figure 24: Network Architecture and OA&M in RAN

The NETCONF Server in CU/DU is to provide CM/PM/FM interfaces with the YANG model to NETCONF Client. A text mode NETCONF Client is implemented in the management server.

Through the NETCONF protocol and interfaces defined in the YANG model, the user in management server is able to access the configuration of DU/CU. DU/CU can report alarms to management server when events are triggered. As for the network performance, the DU/CU will calculate and report to management server in a regular period.

The details of the YANG model are shown in Figure 25.

Configuration Management (CM)	Performance Management (PM)	Fault Management (FM)
DeviceInfo Model Name Software Version Serial Number Uptime FW Upgrade RF Upgrade Log Upload **CellConfig** AdminState gNBAddress eGTPUAddress AMFServerList CellIdentity FIVEGSTAC PLMNID	**PerfMgmt** Enable URL PeriodicUploadInterval PeriodicUploadTime **PerfMgmtStatus** RRCReconfigurationSendNum RRCReconfigurationCompeleteReceiveNum DLPdcpThroughputIgress DLPdcpThroughputEgress ULPdcpThroughputIgress ULPdcpThroughputEgress DLMacThroughputIgress DLMacThroughputEgress	**FaultMgmt** SupportAlarmNumberOfEntries MaxCurrentAlarmOfEntries HistoryEventNumberOfEntries SupportedAlarm CurrentAlarm HitoryEvent **AlarmEvent** gNBSystemStart gNBSevericON FWUpgrade SystemHalt

NRArfcnDL NRArfcnUL SsbAbsoluteFreq TDDConfigPattern MaxMimoLayers DL256QAMSupport	ULMacThroughputIgress ULMacThroughputEgress DLAverageMcs ULAverageMcs DLRbNumber ULRbNumber DLPdschBler ULPuschBler	

Figure 25: the CM interfaces in YANG model

For items in Configuration Management (CM), **DeviceInfo** stands for device information which presents the basic information of the 5G gNB. The parameters Uptime is how long the system is alive. This parameter is used for monitoring the system reliability.

CellConfig stands for the cell configuration. To make cell setup properly, parameters below **CellConfig** including network IP such as gNBAddress, eGTPUAddress; frequency such as NRArfcnDL, NRArfcnUL and SsbAbsoluteFreq and cell ID such as CellIdentity and PLMNID in this list should be set correctly.

For items in Performance Management (PM), parameters under **PerfMgmt** are the settings for the system to upload performance status to certain URL (management server). Refer to Figure 26, parameters under **PerfMgmtStatus** presents the calculation of throughput for each layer in the stack. From up to bottom, parameters which have RRC wording presents the RRC reconfiguration status. Parameters which have **Pdcp** wording presents the traffic to enter (Ingress) and exit (Egress) PDCP layer. Parameters which have **Mac** wording presents the ingress and egress throughput for MAC layer. ULRbNum, DLPdschBler and ULPusch-Bler is the statistic for physical layer.

For the Fault Management (FM), when events such as RAN system starts, service is enabled, firmware upgrade or system halt is triggered. The NETCONF server will report to NETCONF client accordingly.

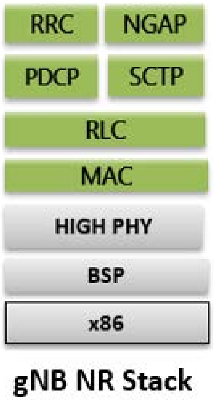


Figure 26: gNB stack.

5.1.1 KPI of throughput, latency and reliability

In 5G-CONNI project, working group 5 (WG5) is about the integration test in the field. We use the RAN monitoring system to monitor throughput and reliability. As for the latency, since it is the measurement of the round trip time (RTT) between two network elements, e.g. 5G

CPE and certain equipment behind 5GC, the approach of PING and observing the RTT is used to monitor the latency.

To verify the implementation of RAN monitoring, the tasks in WG5 are leveraged. There are three use cases (UC) in WG5 Taiwan site and the corresponding KPI requirement are listed in Table 6.

Table 6: KPI requirement in WG5 field site in Taiwan

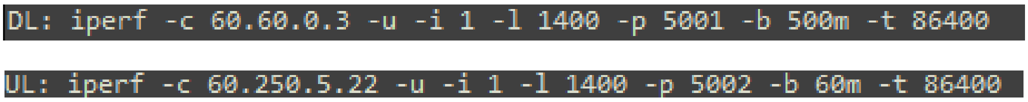
UC & Application	KPI (Throughput)	KPI (Latency)
UC1 Data Collection	UL TCP > 16Mbps	(no requirement)
UC2 AR Diagonostic	UL TCP > 3Mbps DL TCP > 50 Mbps	E2E latency < 30ms Motion to photon latency < 50ms
UC3 Flexible Workholding	UL TCP > 170kbps DL TCP > 335kbps	E2E latency < 20ms

Table 7: KPI and Yang model parameters

KPI RAN Monitoring	Throughput	Reliability
	DLPdcpThroughputIgress DLPdcpThroughputEgress ULPdcpThroughputIgress ULPdcpThroughputEgress	Uptime

To check the KPI result of Throughput and Reliability, iperf tool is used to emulate the traffic in the RAN system. Figure 27 shows the command for UL and DL traffic simulation. 500Mbps traffic is pumped in the downlink (DL) and 60Mbps traffic is for uplink (UL).

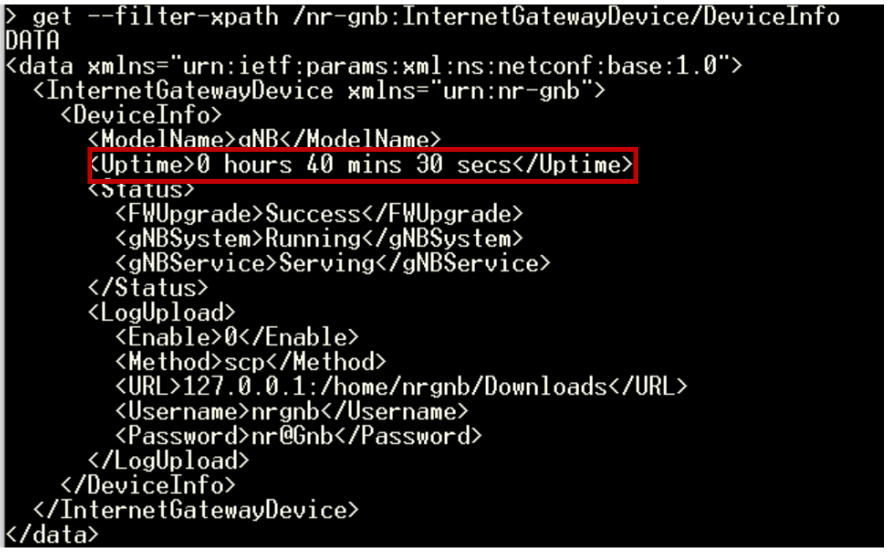
The RAN parameters listed in Table 7 can be accessed with commands shown in the tables below. The Uptime can be observed in Figure 28 and the PDCP throughput for DL and UL can be observed in Figure 29. Not to count the header length of high level protocols, the throughput measured is close to the traffic generated by iperf.



```
DL: iperf -c 60.60.0.3 -u -i 1 -l 1400 -p 5001 -b 500m -t 86400
```

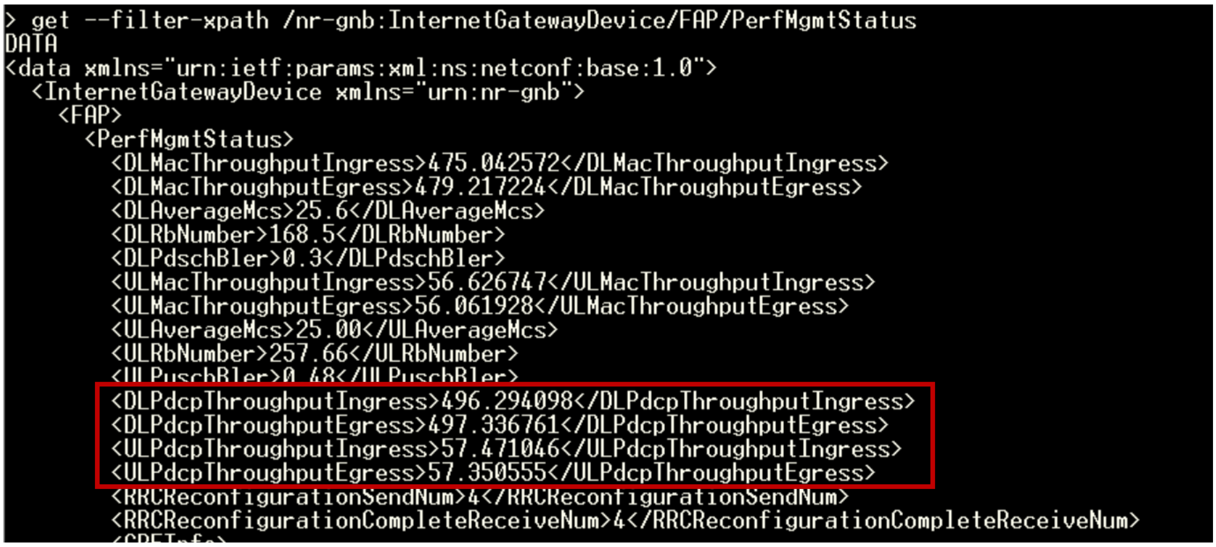
```
UL: iperf -c 60.250.5.22 -u -i 1 -l 1400 -p 5002 -b 60m -t 86400
```

Figure 27 iperf commend used to simulate the traffic in the RAN



```
> get --filter-xpath /nr-gnb:InternetGatewayDevice/DeviceInfo
DATA
<data xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <InternetGatewayDevice xmlns="urn:nr-gnb">
    <DeviceInfo>
      <ModelName>gNB</ModelName>
      <Uptime>0 hours 40 mins 30 secs</Uptime>
      <Status>
        <FWUpgrade>Success</FWUpgrade>
        <gNBSystem>Running</gNBSystem>
        <gNBService>Serving</gNBService>
      </Status>
      <LogUpload>
        <Enable>0</Enable>
        <Method>scp</Method>
        <URL>127.0.0.1:/home/nrgnb/Downloads</URL>
        <Username>nrgnb</Username>
        <Password>nr@Gnb</Password>
      </LogUpload>
    </DeviceInfo>
  </InternetGatewayDevice>
</data>
```

Figure 28: Uptime



```
> get --filter-xpath /nr-gnb:InternetGatewayDevice/FAP/PerfMgmtStatus
DATA
<data xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <InternetGatewayDevice xmlns="urn:nr-gnb">
    <FAP>
      <PerfMgmtStatus>
        <DLMacThroughputIngress>475.042572</DLMacThroughputIngress>
        <DLMacThroughputEgress>479.217224</DLMacThroughputEgress>
        <DLAverageMcs>25.6</DLAverageMcs>
        <DLRbNumber>168.5</DLRbNumber>
        <DLPdschBler>0.3</DLPdschBler>
        <ULMacThroughputIngress>56.626747</ULMacThroughputIngress>
        <ULMacThroughputEgress>56.061928</ULMacThroughputEgress>
        <ULAverageMcs>25.00</ULAverageMcs>
        <ULRbNumber>257.66</ULRbNumber>
        <ULPuschBler>0.48</ULPuschBler>
        <DLPdcpThroughputIngress>496.294098</DLPdcpThroughputIngress>
        <DLPdcpThroughputEgress>497.336761</DLPdcpThroughputEgress>
        <ULPdcpThroughputIngress>57.471046</ULPdcpThroughputIngress>
        <ULPdcpThroughputEgress>57.350555</ULPdcpThroughputEgress>
        <RRCReconfigurationSendNum>4</RRCReconfigurationSendNum>
        <RRCReconfigurationCompleteReceiveNum>4</RRCReconfigurationCompleteReceiveNum>
        <CPEInfo>
```

Figure 29: DL and UL PDCP Ingress/Egress Throughput

5.2 Core network monitoring

The new tools to monitor the Core network monitoring have been developed by III. The Core Network Management architecture is reported in Figure 30. The 5G Core Network Manager (CNM) has been introduced to provide a single clean, consistent management interface regardless of network element type. The graphical interface was introduced to reduce training requirements, and allow operator personnel to quickly drill down to the source of any issue to keep the network running at optimal efficiency.

5G CNM consolidates OA&M operations for all network elements and provides:

- Fault Management of all network elements to provide best-in-class GUI capabilities by integrating with OAM
- Simple graphical data fill editor for Configuration Management
- Consolidated Performance Management of all network elements to provide simplified GUI with advanced features

- Robust Security Management to set user/group level controls to provide access to network elements

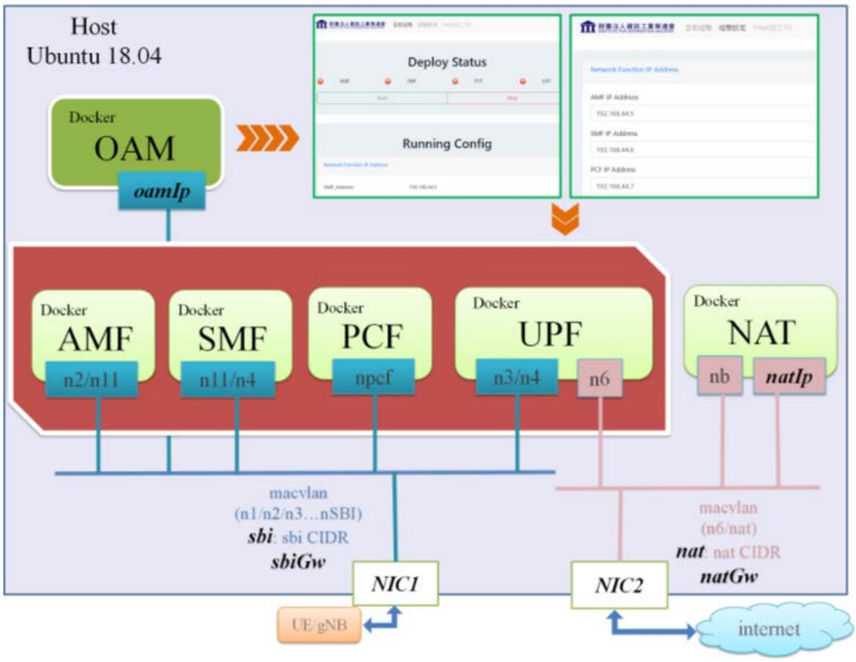


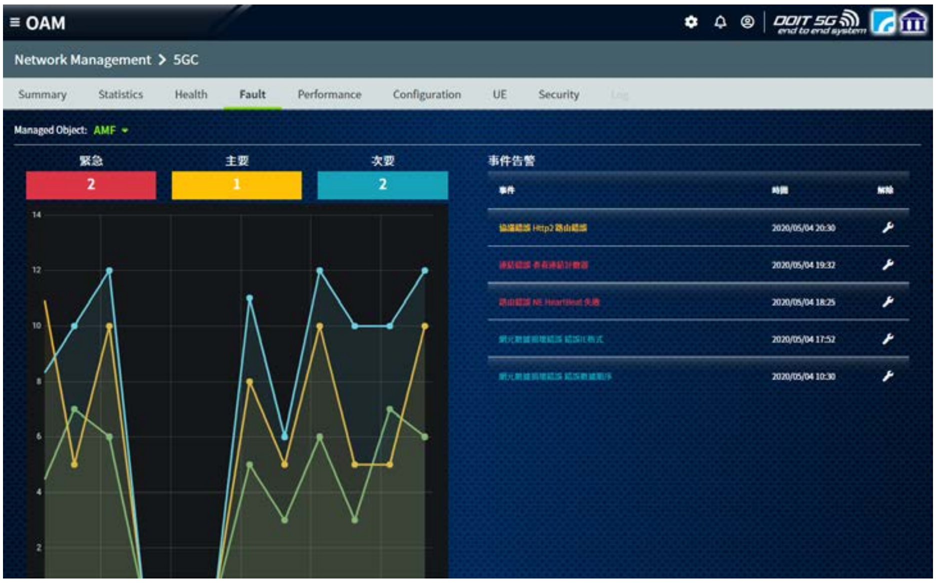
Figure 30: The Core Network Management architecture

The 5GC CNM GUI consolidates the following OA&M operations for the core network:

- General Information: shows general 5G Core execution information that includes information like CPU usage, memory usage and healthy management, which had been reported in D3.1.
- Performance Monitoring: manages overall performance of the network system to avoid potential performance bottlenecks, which had been reported in D3.1.
- Fault Management, Configuration Management, Accounting Management and Security Management functions are new functionalities that have been developed in 2022 and are described below.

Fault Management (FM)

With fault management, the potential problems are identified to keep the network system operational and minimize the downtime. The FM alert system sets three kinds of warning event with different color to distinguish their severity. For example, if the UE authentication failure will be alerted as the most serious event, it will marked as red. An example of operation of the Fault Management tool is shown in Figure 31.



Managed Object: AMF

緊急	主要	次要
2	1	2

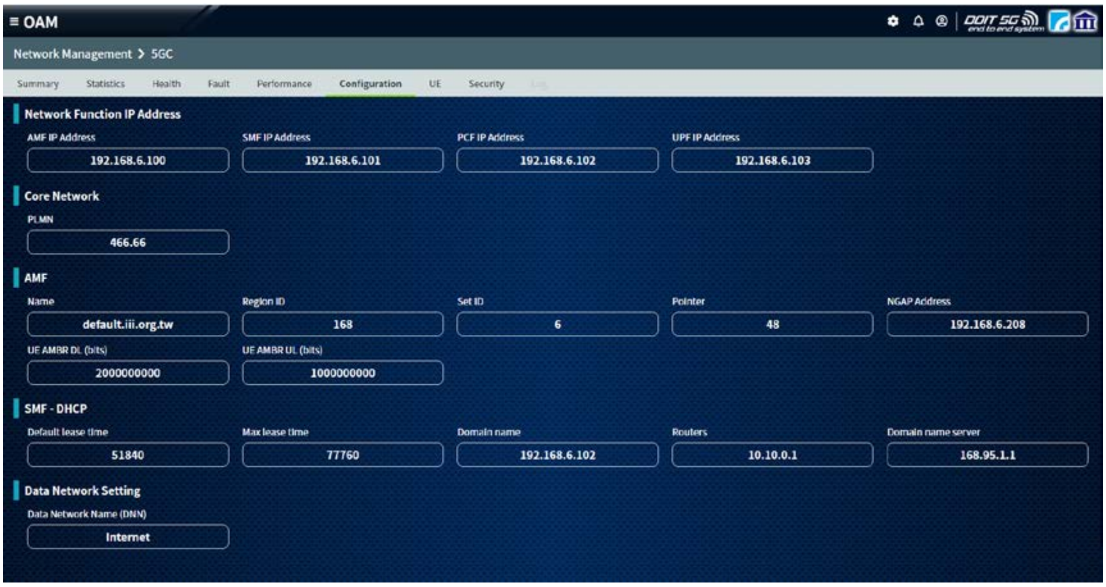
事件告警

事件	時間	解除
[illegible] Http2 [illegible]	2020/05/04 20:30	
[illegible]	2020/05/04 19:32	
[illegible] NE Heartbeat [illegible]	2020/05/04 18:25	
[illegible]	2020/05/04 17:52	
[illegible]	2020/05/04 10:30	

Figure 31: The FM of 5G Core OAM

Configuration Management (CM)

In this configuration management section, the network operation is monitored and controlled. Including the IP address of every 5G core network functions, PLMN, Data Network name and the QoS parameter such as UE AMBR for DL and UL can be set up. An example of operation of the CM tool is reported in Figure 32.



Network Function IP Address

AMF IP Address: 192.168.6.100
SMF IP Address: 192.168.6.101
PCF IP Address: 192.168.6.102
UPF IP Address: 192.168.6.103

Core Network

PLMN: 466.66

AMF

Name: default.iii.org.tw
Region ID: 168
Set ID: 6
Pointer: 48
NGAP Address: 192.168.6.208
UE AMBR DL (bits): 2000000000
UE AMBR UL (bits): 1000000000

SMF - DHCP

Default lease time: 51840
Max lease time: 77760
Domain name: 192.168.6.102
Routers: 10.10.0.1
Domain name server: 168.95.1.1

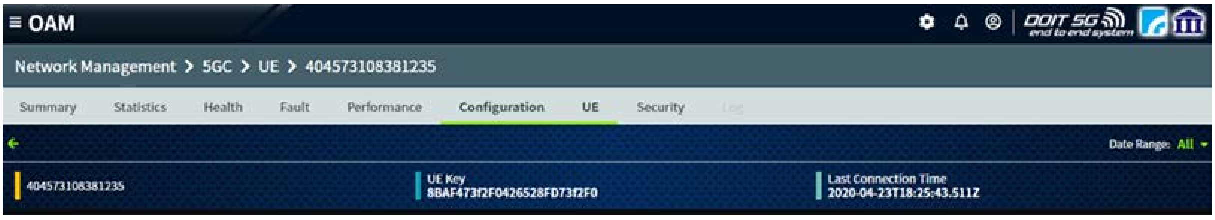
Data Network Setting

Data Network Name (DNN): Internet

Figure 32: The CM of 5G Core OAM

Accounting Management (AM)

The purpose of AM is to monitor and allocate the resources optimally and fairly among UE subscribers. In addition to make a more effective usage of the system's resources available in order to minimize the operation cost, the AM page shows also the information about data plane status like throughput and packets per second value of specific users. An example of the AM interface is reported in Figure 33.



404573108381235

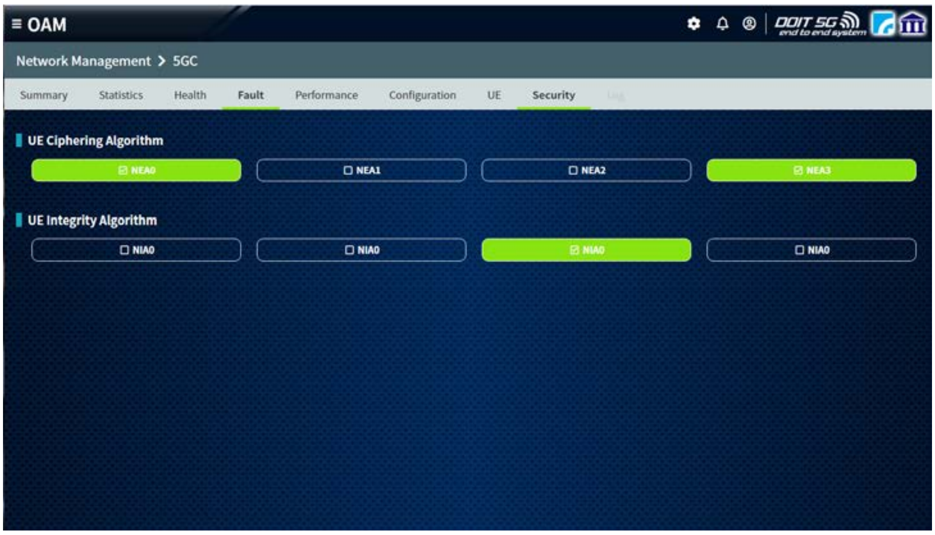
UE Key
8BAF473f2F0426528FD73f2F0

Last Connection Time
2020-04-23T18:25:43.511Z

Figure 33: The AM of 5G Core OAM

Security Management (SM)

At the security management, the network is protected against unauthorized users and deciphered RRC connection. The privacy of user information is maintained where necessary or warranted. The SM function shows the NR encryption algorithm and integrity algorithm for 5G system like NEA and NIA parameters. An example of the SM interface is reported in Figure 34.



UE Ciphering Algorithm

☑ NEA0 ☐ NEA1 ☐ NEA2 ☑ NEA3

UE Integrity Algorithm

☐ NIA0 ☐ NIA0 ☑ NIA0 ☐ NIA0

Figure 34: The SM of 5G Core OAM

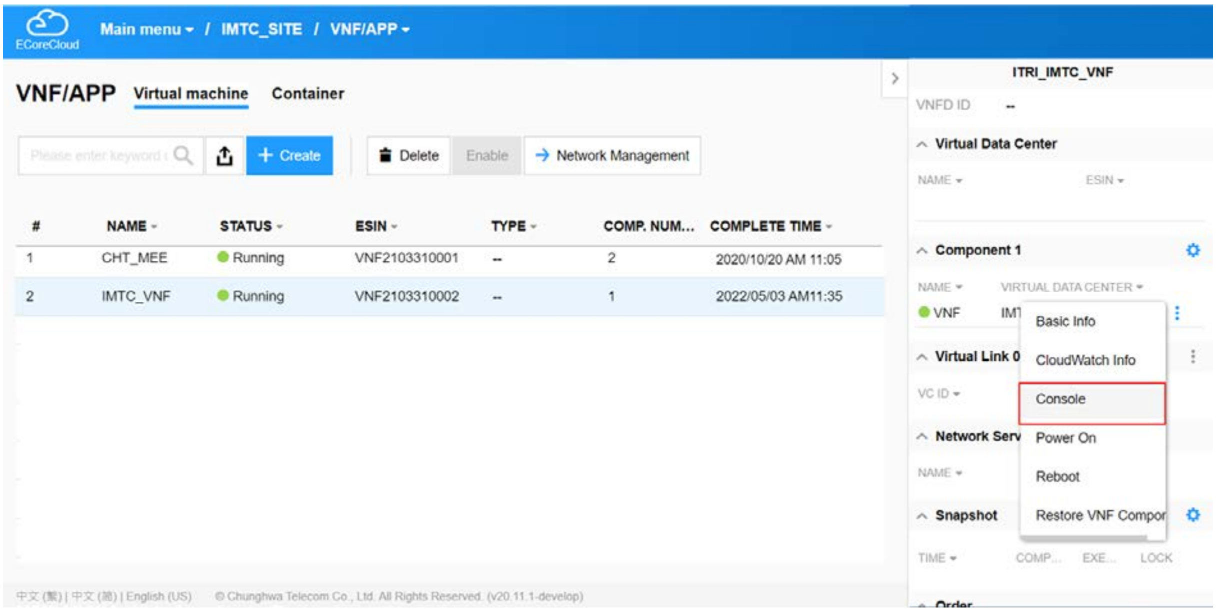
5.3 Edge cloud monitoring

The new tools to monitor the edge cloud have been developed by CHT. The deployment of private 5G networks in smart factories without interruption of industrial applications is a necessary requirement. Production line interruptions can result in substantial monetary losses. Therefore, the monitoring of the continuous operation of the industrial application can be realized through the edge cloud. Deploying industrial applications on the edge cloud can reduce latency and bring the advantages of rapid application onboarding and resource management. MANO test environment has been developed followed by the ETSI standards in

5G-CONNI's Taiwanese testbed, as described in D3.1 as well, MANO test environment can monitor the VNF. In this testbed, The edge cloud onboard the MEE VNF for traffic steering and ITRI IMTC's cloud-based controller of a fixture system application (cf. The additional use cases proposed in D1.1, Section 2.5).

ITRI IMTC's cloud-based controller of a fixture system application has low latency requirements and the scenario of 5G characteristic is URLLC. So we choose this application to virtualize remote computing of the cloud controller to VNFs on the edge cloud. The edge cloud is deployed as close to the ground controller as possible to reduce the data transmission time. Besides, there are two parameters in the cloud-based controller application to make sure continuous operation, including motion command buffer size and command transmission delay. Our MANO Test MANO can monitor the primary resource usage such as CPU, memory, disk and network interface. It also provides the SNMP and RestAPI interface to monitor the customized parameters of the applications and enables different industrial applications to be monitored.

To realize the virtualization and NFV management of cloud controller applications, we must consider industrial application requirements, VNF transformation, and VNF with monitoring. At first, Industrial application requirements often include packet latency and data throughput, which must remain the same after virtualization. Therefore, the virtualization platform as an edge cloud to achieve high transmission and low latency by using SRIOV and PCI passthrough technologies for VNFs. After that, the platform virtualizes cloud controller applications to VNFs following a structure based on ETSI MANO architecture. Then, the enterprise achieves VNF management such as lifecycle management, health, and scaling through the virtualization platform. Moreover, the virtualization platform also has to support customized monitoring to keep the availability of VNFs. At last, the enterprise could easily keep up with the development trend of 5G smart factories by implementing the virtualization platform. The cloud controller application is virtualized on the edge cloud platform, as shown in Figure 35.



#	NAME	STATUS	ESIN	TYPE	COMP. NUM...	COMPLETE TIME
1	CHT_MEE	● Running	VNF2103310001	--	2	2020/10/20 AM 11:05
2	IMTC_VNF	● Running	VNF2103310002	--	1	2022/05/03 AM11:35

VNFD ID --

Virtual Data Center

NAME	ESIN

Component 1

NAME	VIRTUAL DATA CENTER
● VNF	IMT

Virtual Link 0

VC ID

Network Serv

NAME

Snapshot

TIME	COMP...	EXE...	LOCK

Order

Figure 35: Virtualization cloud controller application

To ensure continuous operation and keep every mechanism in complete running order in 5G smart factories, edge cloud must accurately monitor smart factories and efficiently notify operators. To this end, the cloud controller application needs to periodically collect machine operating parameters and status from the factory for use by the edge cloud platform. After that, the edge cloud platform receives necessary raw data from the cloud controller application through SNMP or Restful API interface for monitoring parameters analytics. When monitoring parameters exceed thresholds defined by operators, the edge cloud platform must notify the enterprises immediately via email, SMS, or the user portal. Therefore, enterprises can effectively maintain the continuous operation of smart factories with the edge cloud platform.

In the Taiwanese deployment of the 5G CONNI project, we implemented a 5G smart factory with a monitoring mechanism, as shown in Figure 36. Our edge cloud platform built the cloud controller application with PCI pass-through, remote control interface, and compatible drivers, according to the structure based on the ETSI MANO architecture. The Mobile Edge Enabler VNF on edge cloud platform provides the traffic steering for cloud controller and ground controller. The edge cloud platform monitor the MEE VNF status described in D3.1. It supported the customized monitoring parameters for the application through the previously mentioned monitoring mechanism, such as motion command buffer size and command transmission delay. The edge cloud platform provides a complete monitoring solution to ensure the continuous operation of application services at smart factories. Moreover, the cloud controller must guarantee a buffer size greater than 100 and a latency time less than 30 ms to ensure efficient and stable machining processes. Therefore, we monitor the above two monitoring parameters by our edge cloud platform to maintain the continuous operation of the cloud-based controller of a fixture system application. When the monitoring parameters exceed the proposed requirements, the edge cloud platform will send an alarm notification to the administrator, as shown in Figure 37. At last, the edge cloud platform can provide private 5G networks at smart factories with real-time operation, critical alarms, maintenance time reduction, and many operational benefits.

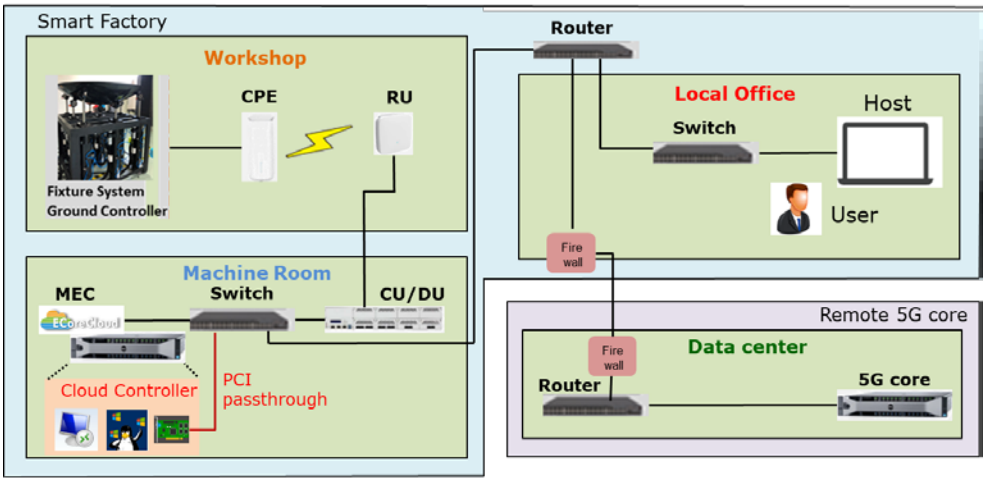
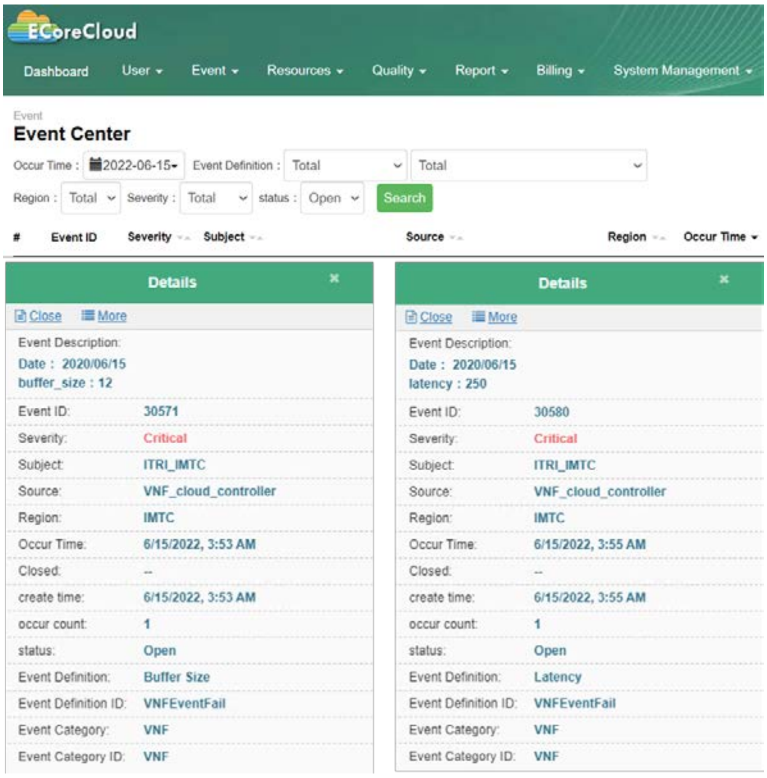


Figure 36: Edge cloud platform of 5G CONNI Taiwanese testbed



Event

Event Center

Occur Time : 2022-06-15 Event Definition : Total Total

Region : Total Severity : Total status : Open Search

#	Event ID	Severity	Subject	Source	Region	Occur Time

Details

Close More

Event Description:
Date : 2020/06/15
buffer_size : 12

Event ID:	30571
Severity:	Critical
Subject:	ITRI_IMTC
Source:	VNF_cloud_controller
Region:	IMTC
Occur Time:	6/15/2022, 3:53 AM
Closed:	--
create time:	6/15/2022, 3:53 AM
occur count:	1
status:	Open
Event Definition:	Buffer Size
Event Definition ID:	VNFEventFail
Event Category:	VNF
Event Category ID:	VNF

Details

Close More

Event Description:
Date : 2020/06/15
latency : 250

Event ID:	30580
Severity:	Critical
Subject:	ITRI_IMTC
Source:	VNF_cloud_controller
Region:	IMTC
Occur Time:	6/15/2022, 3:55 AM
Closed:	--
create time:	6/15/2022, 3:55 AM
occur count:	1
status:	Open
Event Definition:	Latency
Event Definition ID:	VNFEventFail
Event Category:	VNF
Event Category ID:	VNF

Figure 37: Alarm Information

6 Conclusions

The channel models developed in WP3 and described in Section 2 are used throughout the whole project to assess the performance, assuming a realistic channel model to be used in an industrial scenario. The optimal resource allocation and service placement derived in WP3, in the static case, forms the basis for its generalization to handle the dynamic case in WP4. The monitoring and alerting functions developed in Section 5 have been used in a live industrial environment in WP5 to ensure the smooth operation of the use case and 5G system. In addition, virtualization and NFV management of the edge cloud monitoring described in Section 5.3 make possible to explore the advanced capabilities of MEC in WP4. The Edge Cloud monitoring tool is also applied within WP5's activity.

7 References

[GAL62] D. Gale and L. S. Shapley, “College Admissions and the Stability of Marriage,” The Amer. Math. Month., vol. 69, n. 1, pp. 9-15, 1962.

[SHU13] D.I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” IEEE Signal Process. Mag., vol. 30, no. 3, pp. 83–98, May 2013.

[TSI16] M. Tsitsvero, S. Barbarossa and P. Di Lorenzo, "Signals on Graphs: Uncertainty Principle and Sampling," in IEEE Transactions on Signal Processing, vol. 64, no. 18, pp. 4845-4860, Sept. 2016.