# Real-Time Network Digital Twin: Experimental Setup for Proof of Concept

Oscar Stenhammar, Stefano Sorrentino, Fedor Chernogorov, Ricardo Blasco, Gábor Fodor, Carlo Fischione

***Abstract*—Network digital twins (NDTs) are virtual representations of network assets and processes, such as in 5G or 6G networks, synchronized with physical properties. NDTs can be leveraged for network monitoring, automation, and optimization tasks. Simulations and predictions from the NDT can be provided to scout the feasibility of important processes from a network connectivity quality perspective. The real-time modeling of a key performance indicator (KPI) enables continuous network management. This article presents results and insights from a proof-of-concept NDT for an enterprise use case, using real-time data from commercially available communication equipment. Neural networks, assisted by comprehensive feature selection and extraction, are integrated into the NDT to model the KPI, namely the downlink user throughput. KPI evaluations from the NDT are provided based on requests from the real-world generated demanding scenarios. The results provide general insights on an accurate real-time NDT that can support continuous network configuration updates.**

## I. Introduction

SINCE the beginning of 5G development, wireless communications have been targeting industrial markets, in addition to the traditional mobile broadband use cases [1]. Due to the rapid evolution of cellular networks, 6G will arguably enable more connected applications, demanding higher data rates, and improved intelligence [2]. To adequately address such increasing demands and complexities of the new generations, there will be a greater dependency on automated solutions for different layers of network functions.

Digital twin (DT) technology has recently shown benefits in several industrial domains, such as manufacturing, intelligent transport systems, and communication systems [3]. Unlike traditional network simulation and optimization methods, which rely on static or offline data, DTs dynamically synchronize with live network data, making them more effective for adaptive network management in complex and evolving environments. This has triggered the interest of DTs in academic and industrial research communities [1-12].

Network digital twins (NDTs) serve as virtual replicas of mobile network functions or physical infrastructures. More specifically, NDTs facilitate continuous monitoring, analysis, and prediction of network performance and can be seen as one of the tools for network automation [4]. With seamless testing and deployment of new technologies, NDTs help reduce the risk and cost associated with physical trials. This flexibility is crucial as networks evolve to accommodate the increasing demands of emerging technologies. Ultimately, the NDT has the ability to empower 5G and help prepare for 6G.

In cellular networks, the analysis part of the NDT consists of scenario simulation, network performance evaluation, predictive analysis, and optimization recommendations. Such analyses can be based on simulations, statistical methods, and artificial intelligence (AI). Among other benefits discussed in [5-8], one strong aspect of AI-based methods is the capability of real-time analysis of large amounts of monitoring data. By leveraging AI algorithms, NDTs can identify patterns and predict potential issues, enabling proactive maintenance and minimizing downtime. A real-time NDT can quickly model KPIs that are time-consuming to measure. Such information allows the network manager to promptly and dynamically adapt to changing network conditions by performing network optimization and resource allocation [1]. Hence, the integration of AI enhances decision-making processes by providing actionable insights and recommendations based on comprehensive data analysis.

In industries, NDTs can operate in combination with an enterprise digital twin (EDT). An EDT can provide a real-time model of enterprise assets and processes, such as the manufacturing process using automated mobile robotics. This allows the enterprise manager to perform monitoring, asset management, and process optimization tasks with high levels of automation. Collaboration between an NDT and an EDT can be advantageous for optimizing enterprise operations using network data or improving network services using inputs from the enterprise operations.

Through live experiments carried out in an industrial-like environment and practical implementations of AI, this paper aims to provide insights of the transformative potential of NDTs in future networks and enterprise applications. To summarize, the contributions of this paper are the following:

- a proof of concept setup of an NDT, based on an industrial indoor cellular network;
- an implementation of an AI-based real-time predictor as the analysis part of the NDT;
- a discussion of the end-to-end latency, potential benefits, challenges, and next steps for the proposed NDT, based on the insights obtained in the present work.

The remainder of this work is structured as follows: we review prior NDT research to identify gaps and contextualize

©2025 IEEE. Originally published in IEEE Wireless Communications [16]. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work.

Oscar Stenhammar, Stefano Sorrentino, Fedor Chernogorov, Ricardo Blasco, and Gabor Fodor are with Ericsson Research Sweden and Finland. Oscar Stenhammar, Gabor Fodor, and Carlo Fischione are with KTH Royal Institute of Technology, Sweden. Carlo Fischione is also affiliated with Digital Futures.

**TABLE I:** An overview of previous research in network digital twins (NDTs).

| Reference | Technologies | Use Cases | DT Design Targets | Role of AI | Data |
|---|---|---|---|---|---|
| **Nguyen, "Digital Twin for 5G and Beyond", [1]** | 4G, 5G, Narrowband IoT, Enhanced MTC | Massive/critical MTC, Enhanced MBB, Fixed wireless access | Shorten the validation time of 5G use cases, reduce costs, and support NW optimization | DT evolves to become more comprehensive and to achieve higher accuracy of reality | No data |
| **Jang, "DT for Intelligent Network" [6]** | 5G, 5G Advanced | NW planning, management, and optimization | Demonstrate DT from observations of cellular NWs | DT is used to train/test models of the real NW | Field trial employing three cellular BS |
| **Xia, "DT and AI for Intelligent Planning" [7]** | 4G, 5G with emphasis on massive MIMO | Enhanced MBB, URLLC, Massive/critical MTC | DT to offer a comprehensive view of cellular NWs in smart factories | DT and AI are used to plan, deploy, and operate the NW for energy efficiency | Simulations |
| **Groshev, "Toward Intelligent Cyber-Physical Systems: DT Meets AI" [8]** | 5G and WiFi 6E | Industry 4.0, real-time sensory data acquisition e.g., predictive maintenance | DT to facilitate the training of AI agents that require computation-intensive training | Movement prediction and task learning of robotic arm | Experimental validation using Nyrio One robotic arm with 802.11n WLAN |
| **Erman, "Generative Network Performance Prediction" [9]** | 5G | Use-case dependent service level agreement (SLA) compliance prediction | NDT to operate networks and to ensure SLA compliance | Generative adversarial NW methods to create/analyze "what if scenarios" | No data |
| **This work** | 5G | Industrial use case in symbiosis with an EDT (see Fig. 1) | Provide EDT with quality of service (QoS) predictions in requested scenarios | Model the feasible QoS | Real-time measurements in an industrial-like environment |

our study; describe our system model; present a case study with a live network-based NDT and real-time AI-driven KPI predictor; discuss benefits, challenges, and future directions; and conclude with a summary.

## II. Literature review

Compared to conventional simulation-based approaches, DTs offer a more responsive and data-driven methodology by integrating real-time measurements with analytics. Prior research has demonstrated the benefits of DTs in industrial and telecommunications domains [7-9]. NDT technology facilitates proactive assessment of data traffic and security vulnerabilities, improving operations and aiding the launch of innovative new services [1]. Several previous works envision AI as a key enabler for NDTs to facilitate anomaly detection, predictive analytics, or insights for network management [1], [3], [5-9]. Such functions have been used to support enterprise applications [8-10] and to enhance network performance [1], [6-7]. To give a comprehensive overview of the most significant studies within NDTs, we present them in relevant categorical columns in Table I. The table provides insights, such that NDTs have been considered for monitoring, operational, and predictive purposes. Furthermore, Table I illustrates that AI is the core of the NDT models in previous works.

An essential aspect to consider, which Table I captures, is whether real-world measurements or simulated data are used to train and evaluate the NDT framework. Simulated data with a communication model is more accessible. Additionally, generative AI holds the potential to replicate the behavior of a cellular network and to generate vast amounts of synthetical network data for an NDT [5]. However, measured data may provide data closer to reality, which gives a more rational scenario and more realistic results than simulated data [12-13]. To assess the generalizability of the predictive part of the NDT, the work in [13] has been experimenting with training a predictor with measured data from a source domain and evaluating the predictor on a separate target domain. The experiments show that training and evaluating the predictive model on different data domains limit the prediction performance, highlighting the importance of constructing representative datasets to properly train an NDT. This is an important and challenging aspect since constructing proper datasets by measurement campaigns tends to be time-consuming and costly [9, 10].

In summary, previous work on NDTs in Table I implements DTs to monitor, optimize and/or predict the dynamics of the network. The listed previous works in Table I are considerably different from our work in terms of use cases, design targets, the role of AI in meeting those targets, the source of training data, and the technologies that are in the scope of the proof of concept. In [7-8], NDTs are implemented to support enterprise use cases. In such scenarios, the DT concept can be extended by constructing an EDT as a virtual representation of the enterprise process to allow interaction between the NDT and EDT. We aim to showcase a proof of concept of such a model with real-time compatibility, that relies on collected data from realistic scenarios.

## III. System Model

To implement and evaluate the proposed real-time NDT, we define a system model that captures the data flows and interactions between the digital and physical network environments. We consider a system that exemplifies an enterprise process with stationary and mobile assets, such as automated guided robots, that perform safety-critical enterprise tasks. The assets are controlled by a central unit via wireless communication. The safety-critical nature of the enterprise process requires communication with high quality of service (QoS). When enterprise assets are connected via mobile networks to enable safety-critical processes, wireless communication performance needs to be assessed in the planning process. The common

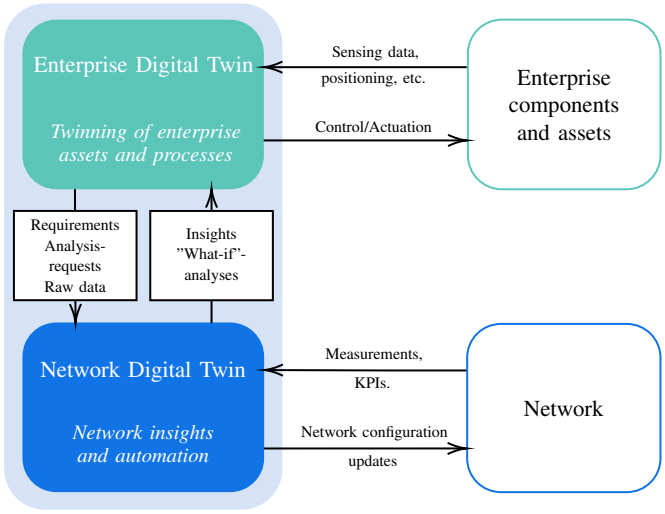



**Fig. 1:** A flow chart of the exchange of information between an NDT and an EDT. This work particularly aims to examine the blue NDT block of this figure.

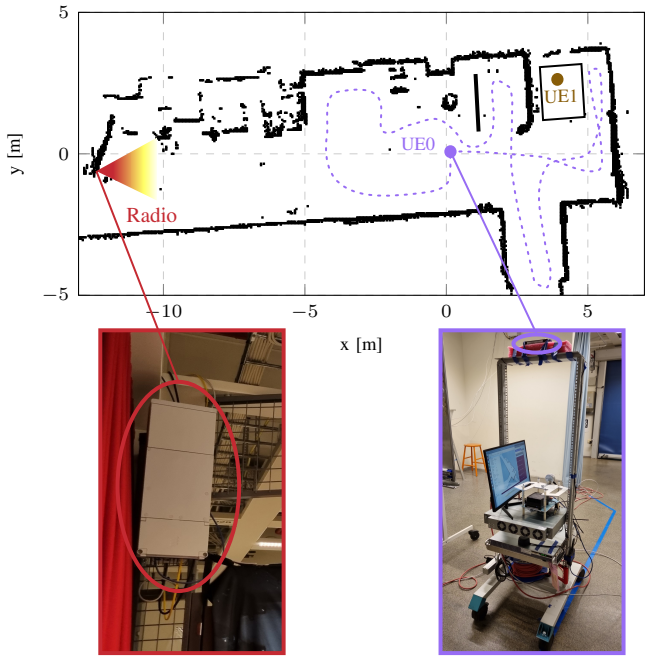



**Fig. 2:** An illustration of the experimental setup and indoor environment where the experiments from the case study are conducted. The plotted map is obtained from a LIDAR. A photo of the 5G base station (BS), encircled in red, is included. Another photo is included to illustrate the setup of the mobile UE positioned on a trolley, encircled in purple. An example of the trace of UE0 is visualized by the purple dashed line. The position of the static UE1 is also provided.

way of addressing this challenge consists of dimensioning the network for the most demanding cases, with obvious drawbacks in terms of deployment, operational cost, and efficiency.

To model the industrial system examined in this work, we consider a combination of an EDT and NDT, interacting over an interface as in Fig. 1. However, we are only interested in a proof of concept of the blue NDT block. To facilitate the necessary inputs for the NDT in Fig. 1, we must provide measurements from the network, and some raw data of an enterprise scenario. We focus on evaluating the analysis part of an NDT. Analyses from the NDT could potentially be used to send network configuration updates or insights to an EDT, according to Fig. 1.

A reasonable scenario of the model in Fig. 1 involves the EDT sending requests regarding the QoS for the enterprise assets to the NDT. The NDT provides feedback in the form of predictions and insights of the network assets as a response to the EDT. Predicted insights provided by the NDT can be leveraged in a connectivity quality guarantee perspective for the physical enterprise process. Such a model allows for a more proactive and dynamic utilization of network resources, where predicted network capabilities assist in deploying enterprise processes.

The requests from the EDT in this work are defined by the layout of the environment, the position of the industrial robots, and their requested QoS. Potentially critical scenarios of the enterprise process that are used to request feedback from the NDT, which generative algorithms could produce, are instead generated by real-world scenarios. This is to gain more control of the scenarios in our experiment and to deliver a proof of concept with real-world data. Based on the provided data, requests for real-time throughput predictions are sent to the NDT. The idea of this process is to ensure a sufficient QoS for the enterprise assets to perform critical processes.

## IV. Case-Study: Real-Time Network Digital Twin Proof of Concept

This study assesses the feasibility of real-time NDT analysis in enterprise scenarios using experimental data from commercial network equipment, integrating AI into the NDT framework. The system model from the previous section provides a structured foundation for implementing NDT with live network data. In this section, we describe an experimental setup reflecting industrial scenarios where users demand high QoS for safety-critical tasks. Two scenarios are examined with user equipment (UE) as potential enterprise assets, each with varying throughput targets.

### A. Experimental Setup

The experimental setup consists of one 5G new radio (NR) base station (BS) and two UEs, as defined in Table II. Two off-the-shelf UEs are used: UE0 is installed on a mobile trolley, while UE1 is placed on top of a server rack in a non-line-of-sight (NLOS) position. Fig. 2 illustrates the indoor experimental area, with the BS radiation sector represented by the orange triangle and UE0's typical mobility trace shown with the purple dashed line. Propagation scenarios and throughput levels are chosen to create a realistic yet challenging scenario for the NDT.

UE0's trolley is equipped with a LIDAR-based simultaneous localization and mapping (SLAM) positioning system, which

**TABLE II:** An overview of the setup for the experiments in this study.

| Device | Setting | Measurements |
|---|---|---|
| **Base station** | Commercially available Non-Stand Alone 5G NR BS with analogue beamforming and dual polarization. **Duplexing mode:** Time-Division Duplex (TDD), 4:1 pattern (DDDSU). **Carrier frequency:** 28 GHz (band n257) **Bandwidth:** 100MHz | Total downlink throughput, Downlink/uplink radio resource utilization, Path loss distribution, MCS from both PDSCH and PUSCH, Downlink serving beam distribution, RSRP related to the SSB per beam, SINR of the PUSCH per beam |
| **User equipment** | Two commercially available UEs with 2 receiving antennas UE0 mobile on a trolley, UE1 fixed (see Fig. 2) | Position of the mobile UE (2D SLAM solution), Throughput for both UEs |

provides 2D-coordinates with an accuracy in the order of tens of centimeters. The mapping of the room provided by the LIDAR-based SLAM system is visualized in the plot in Fig. 2. The coordinates of UE0 are collected, together with timestamps, as part of the input features. The orientation of UE0 is kept constant during the experiments and, hence, is not measured. Effective downlink throughput is individually measured and logged for each of the UEs using purpose-built traffic generators based on the user datagram protocol. Depending on the experiment, full-buffer or fixed-rate traffic is produced. In both cases, the traffic generator avoids excessive buffering at the BS, to prevent traffic policing algorithms from intervening and creating memory effects in the datasets.

To mimic real-world industrial conditions, we carefully select parameters, using a 5G NR BS at 28 GHz to assess NDT feasibility in mmWave deployments relevant to 6G. The setup recreates mobility and varying propagation conditions by placing UE0 on a trolley to study movement effects and LOS-to-NLOS transitions as illustrated in Fig. 2. Network traffic is configured with full-buffer and fixed-rate conditions to isolate throughput analysis from external interference. These settings balance feasibility with rigorous evaluation, ensuring applicability to enterprise deployments.

### *B. Data processing*

Radio features are logged with a 1 second sampling period via the BS management interface, capturing numerous cell-specific measurements. Some are averaged over one second, while others are reported as histograms. BS measurements include distributions for downlink/uplink resource utilization, path loss, 3GPP table II modulation and coding scheme (MCS) distributions for the physical downlink shared channel (PDSCH) and physical uplink shared channel (PUSCH), serving beam dowlink, reference signal received power (RSRP) related to the synchronization signal block (SSB) per beam, and signal-to-interference-plus-noise ratio (SINR) of the PUSCH per beam. These measurements provide insights into the usage and quality of the radio resources within the specified cell and are intuitively correlated to link and system performance.

UE and BS data are collected via Message Queuing Telemetry Transport (MQTT) and processed centrally to extract input features for AI model training. Integrated into the NDT, the AI models system KPIs, specifically UE downlink throughput, predicting KPI values based on network statistics and UE data.

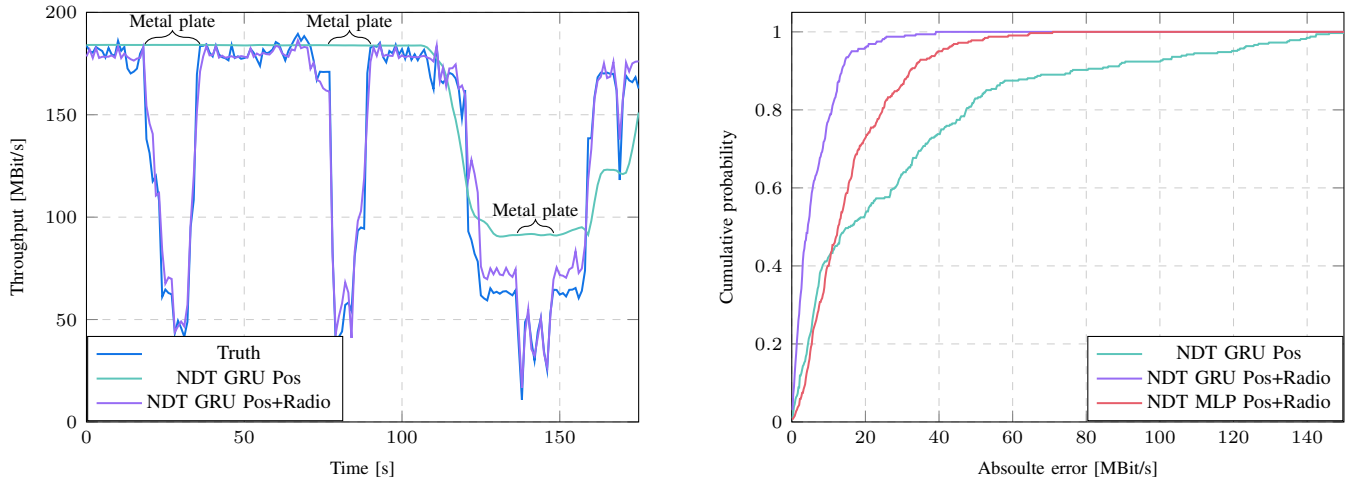
To preprocess and extract valuable information from the BS metrics expressed as distributions, statistical tools are applied. Specifically, the 10th percentile, 90th percentile, and the mean value are calculated for each distribution to serve as candidate inputs for the predictor of the NDT. Furthermore, the Jensen-Shannon divergence (JSD) [14] is pairwise applied to the distributions of the SINR, RSRP, and downlink/uplink radio resource utilization. The JSD provides valuable insights by quantifying differences between probability distributions of network metrics over time. These extracted features are included in the candidate input features to predict the throughput. By using the JSD between two distributions as input to the model, instead of using the entire distributions as inputs, we save inference time and computational power.

With the Shapley additive explanations (SHAP) feature selection method, the most influential radio features are selected to serve as the input for the NDT predictor. Furthermore, we experimentally explore the importance of the BS radio features, and the position and throughput targets of the UEs.

### *C. Predictor*

Data from the three most recent time indices are included as input to the predictor for each selected feature. From the entire collected dataset, approximately 85% is used to train the NDT predictor, while the remaining 15% establishes the test dataset to evaluate the performance of the predictor. The AI model integrated into the NDT to model the throughput is the recurrent neural network-based gated recurrent unit (GRU), which has previously shown great potential in time series prediction and DTs [15]. The choice of the GRU model over other more advanced deep learning models, is motivated by a trade-off between computational efficiency and predictive accuracy. GRUs offer comparable high performance while requiring fewer parameters, leading to less computational complexity, which is crucial for real-time NDT applications.

The neural network is constructed by two GRU layers with 750 hidden states, followed by a dense layer to calculate the output. We find the number of hidden states by testing how many hidden states we can add to increase prediction accuracy. Using the Adam optimizer and the mean squared error loss function, the predictor is trained for 50 epochs, using a batch size of 32 and a learning rate of 0.0001. We find the number of epochs by applying early stopping, which detects that convergence is reached. The learning rate is found by excessive experimental testing. To avoid overfitting, a dropout rate of 0.2 is implemented after both GRU layers. To evaluate the effectiveness of the GRU predictor, we use the multilayer perceptron (MLP) as a benchmark. The MLP is a neural network that has been proposed for predicting network performance in several previous works, for example in [13]. Four hidden layers are implemented in the MLP, with 750,




**(a)** A sequence of the throughput from the test dataset, with obstructions by a metal plate, and the output from the two predictors.

**(b)** The CDF of the absolute prediction error for each predictor. The NDT GRU predictor including radio data as input is superior.

**Fig. 3:** Results from the evaluation of Scenario 1. The NDT is evaluated for the different predictors. One predictor only includes the position of UE0 as input. The other predictors also include the extracted BS radio features.

500, 250, and 125 neurons. Otherwise, the hyperparameters for the GRU are reused.

### *D. Scenario 1*

As mentioned earlier, we consider two different scenarios in our case study. In this subsection, we describe the first scenario. This scenario corresponds to a case in which the EDT would send queries to the NDT to predict the maximum downlink throughput for UE0 at a given position, as the only active UE in the cell.

The training dataset is collected by feeding the BS with full-buffer traffic while moving UE0, illustrated by the positional trace in Fig. 2. The data in the training dataset spans a broad range of propagation conditions, including LOS to NLOS conditions as the UE moves. A test dataset is separately collected in the same environment. To verify the performance of the NDT predictor, and to manually evaluate a scenario where the EDT requests cases of unexpected obstructions, the test dataset contains propagation challenges that are not included in the training dataset. Specifically, a metal plate is manually used to block the LOS path between the BS and UE0 at position (0, 0). The metal plate also interferes with the NLOS communication when UE0 is located at position (3, -4). Three different predictors are compared for this scenario, where the first predictor performs throughput predictions solely based on UE0's position, while the other two predictors also include radio measurements among the input features.

### *E. Scenario 2*

The second scenario reflects a case where an EDT considers two concurrently active UEs performing demanding tasks, competing for shared radio resources at specific locations. UE0 is moved to different propagation conditions, just as in Scenario 1, while UE1 is stationed at an NLOS position as depicted in Fig. 2. In contrast to Scenario 1, each UE is fed with a configurable maximum throughput, which we call the target throughput rate. We perform experiments with different combinations of target throughputs for UE0 and UE1. Training data is obtained for the following target throughput combinations for respectively UE0 and UE1: {5, 40}, {15, 40}, {15, 60}, and {20, 50} Mbps, while testing is done for {20, 50} Mbps.

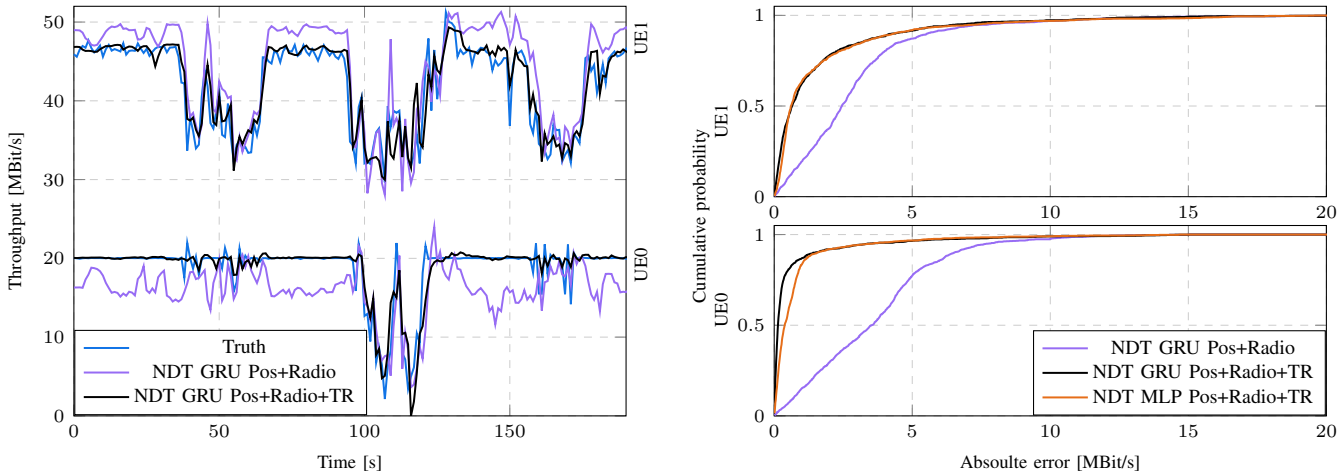
## V. Results

### *A. Scenario 1*

To analyze the NDT case study results, predictors with two different input feature sets are considered. As described in the previous Section IV-D, the first predictor only considers the position of UE0 as the input, while the second predictor also includes the extracted BS radio data features. The purpose is to investigate the importance of the radio data versus the location data in terms of prediction accuracy.

We evaluate the test data in Fig. 3 and delve into details on a sequence from the test data in Fig. 3a. For the first 110 seconds in Fig. 3a, UE0 is positioned at a fixed spot, approximately at coordinates (0, 0) in Fig. 2, with high throughput under normal conditions. Both GRU prediction algorithms, using different input feature sets, predict the throughput accurately at this point, as indicated by the first 20 seconds in Fig. 3a. However, since the position of UE0 is fixed, the output for the predictor solely based on position is unable to identify small fluctuations over time. When the metal plate blocks the LOS between the BS and UE0, starting from time 20s, the throughput drops as indicated by Fig. 3a. Although such obstructions are not present in the training data at position (0, 0), the predictor using radio data as input features can predict throughput close to the ground truth, based on the state of the system.

After 110 seconds, UE0 throughput starts to decline (see Fig. 3a) as the user moves towards position (3, -4) for blockage of the LOS for the NR cell (see Fig. 2). Performance observations between 110 and 160 seconds are characterized by




**(a)** A sequence of the throughput from the test dataset and the output from the two predictors. The three upper graphs and the three lower graphs correspond to UE1 and UE0, respectively.

**(b)** The CDF of the absolute prediction error for the two GRU predictors and the MLP predictor. The upper plot and the lower plot correspond to the performance of UE1 respectively UE0.

**Fig. 4:** Results from the predictors for Scenario 2, with two UEs and natural obstructions of the signal. One GRU predictor includes the BS radio features. The other predictors, MLP and GRU respectively, also include the target throughput rate (TR).

NLOS propagation conditions, including an ancillary NLOS scenario for the time slot between 137 and 148 seconds when the metal plate blocks additional wireless paths between the NR cell and UE0. The predictor exclusively based on the position of UE0 shows an average absolute error of 26 Mbit/s for a basic NLOS condition, and 52.5 Mbit/s average absolute error when the metal shield is present. Correspondingly, for the predictor which uses a combination of radio and positioning data, the average absolute prediction error is 10.3 Mbit/s for the basic NLOS situation and around 2 Mbit/s for the NLOS with metal shield blockage. Here we observe two phenomena. Firstly, the predictor considering the radio data is more accurate for any NLOS condition encountered in the testing data. Secondly, situations with spontaneous NLOS or ancillary NLOS conditions caused by a metal plate, which are not present in the training data, can only be predicted when radio measurements are taken into account by the AI predictor. By comparing the overall performance of the two predictors, it becomes obvious that positioning data by itself is not sufficient for accurate predictions of the UE0 throughput. In Fig. 3b, the 90th percentile of the cumulative distribution function (CDF) for the absolute prediction error for the GRU predictor only using positioning is 120 Mbit/s, while the 90th percentile of the absolute error for the GRU predictor based on both positioning and radio data is 17 Mbit/s. The mean absolute percentage error (MAPE) calculated for these two predictors show the same trend: 43% and 7.0% for the former and latter predictors correspondingly. Moreover, we include the MLP predictor to evaluate the GRU predictor and notice that the GRU outperforms the MLP predictor for UE0.

### *B. Scenario 2*

In the second scenario, UE1 with a fixed position is added to the experimental setup. In this scenario, the target throughput rates for the UEs are varied as described in the previous section. The true throughput of the test data in Fig. 4a is plotted in blue, where the target rate is 20 Mbit/s for UE0 and 50 Mbit/s for UE1. Notably, the target rate of 50 Mbit/s for UE1 is not achievable due to radio conditions, UE capability, and BS configuration. However, the target rate of 20 Mbit/s is achieved for UE0. Furthermore, it can be observed that the throughput of the fixed UE1 is to a larger extent affected by the movement of UE0, than the throughput of the mobile UE0. This is because once UE0 moves into a location with poor coverage, it starts consuming a larger portion of the spectral resources. This deteriorates the performance of UE1, since UE1 has a greater target throughput rate, and all radio resources are consumed to achieve the desired target rates. This dependency on UE performance and their respective radio conditions shows that an NDT can be helpful in finding intricate relations in the real-time data and providing insight into the network quality.

Based on the learnings from the first scenario, it is concluded that the predictor solely based on positioning data can be excluded from the analysis and that the predictor based on radio and positioning data is kept. However, as illustrated in Fig. 4a, the predictor jointly using positioning and radio data presented by the violet graph is not accurate.

An additional predictor that includes the target throughput rate from both UEs as input is implemented to the NDT, presented in black color in Fig. 4. The true throughput in Fig. 4a is especially well captured by this GRU prediction algorithm when the actual throughput is moderately static, even for UE1, which does not reach its assigned target throughput rate. Without the target throughput rate as an input feature, the NDT lacks knowledge of the targeted proportion of resources divided between the UEs. This recognition can be obtained by inspecting Fig. 4a at a specific point in time, realizing that the sum of the predicted throughput for UE0 and UE1 are rather similar for both predictors.

The CDFs of the absolute prediction error are presented in Fig. 4b, where the upper plot represents the CDF for UE1,

and the lower plot represents the CDF for UE0. The CDFs are obtained by cross-validation using data from all target throughput rates in the experimental setup of Scenario 2. The CDFs for the GRU and the MLP predictors show similar results in Fig. 4b. However, the GRU predictor performs slightly better for UE0. The CDFs for the GRU predictors in Fig. 4b, reveal that the absolute error for the 90th percentile is trimmed to 1.5 Mbit/s for UE0 and 4.2 Mbit/s for UE1 by including the target throughput as an input feature, compared to 5.5 Mbit/s for both UEs using the predictor not considering the target throughput. Furthermore, the use of target throughput rate as an input feature to the GRU predictor reduces the MAPE from 22% to 5.8% for UE0, and from 5.3% to 4.0% for UE1. The results of the two error metrics indicate that the prediction accuracy is significantly improved by including the target throughput rate as an input feature.

## VI. Discussion

Transitioning an NDT from a controlled laboratory environment to a real-world setting introduces several challenges, such as increased complexity of network conditions, dynamic interference, and varying user densities. Scalability is another challenge, as larger and more complex networks generate significantly higher volumes of data, necessitating more efficient data processing. As seen in the scenario with two UEs in this work, features that reveal how radio resources among UEs are shared are crucial for the NDT. This was captured by the target throughput rate in this work. To enhance insights into how radio resources are dynamically split among UEs, future work should include additional UE-specific radio features. Such data would be beneficial for NDTs in more complex networks. Yet another main challenge in building a data-driven NDT model for a large-scale deployment is obtaining diverse datasets for proper generalization. In this work, we considered the interplay between an NDT and EDT to manually generate measurements in various scenarios. An approach to obtain more data is to include generative AI for the purpose of generating data.

With access to large, heterogeneous datasets and the set of principles in this paper, the NDT can scale to larger networks. To handle higher device densities and varied conditions, several scalability enhancements can be considered. Optimizing the data processing pipeline by leveraging distributed edge computing and parallelized data handling can help manage increased data volumes while maintaining low-latency performance. Additionally, implementing hierarchical NDT architectures, and EDT architectures in industrial settings, where local DT instances operate at the edge and synchronize with a central model, could improve scalability without overwhelming computational resources. Future implementations could also train models across network segments using federated learning to reduce the need for centralized data aggregation. These strategies would enable the NDT framework to scale effectively and support larger deployments.

To ensure the reliability of real-time NDTs in industrial applications, it is essential for the total latency to be smaller than the sampling time. Depending on the safety-critical tasks that the enterprise process requires, a short sampling time is important to evaluate critical scenarios promptly. When scaling the real-time NDT to a larger network, this becomes an important design aspect. In this work, the predictors are optimized for efficient execution by using feature selection to include only the most influential input parameters and employing batch inference. The preprocessing and predictions of the NDT reach a latency of 0.19s for the GRU and 0.17s for the MLP predictor. Polling and receiving the measurements further add to the total latency, around 0.5s in the scenario with two UEs. Model compression methods, parallel processing, and GPU acceleration can be leveraged to significantly reduce computational latency compared to CPU execution. Combined with the suggested strategies for scalability enhancements, these optimizations will benefit the end-to-end latency, making the proposed framework suitable for real-time NDT applications in industrial scenarios.

## VII. Conclusions

This paper presents a proof of concept for a real-time AI-based NDT deployed in an indoor cellular mobile network environment to provide throughput predictions for enterprise processes. Results demonstrate the NDT's strong ability to predict the UE downlink throughput based on positioning data, radio measurements, and target throughput rate of the enterprise equipment, despite different types of variabilities of the communication paths and varying target throughput rates. The results indicate that insights and predictions from an NDT can help enterprise managers to dimension the cellular network, perform monitoring, and process optimization tasks. Furthermore, this study illustrates the importance of capturing how multiple UEs dynamically share the available radio resources. To model this process more accurately, future work should consider including UE-specific radio information as input features. The study discusses and gives clearly stated directions to overcome challenges connected to the transition of an NDT from a controlled laboratory environment to a real-time, large-scale setting with low-latency requirements.

## Acknowledgement

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation. The work of G. Fodor was funded by the Swedish Strategic Foundation Future Software Systems Project SAICOM, grant no: FUS21-0004.

## References

[1] H. X. Nguyen et al., "Digital twin for 5G and beyond," *IEEE Commun. Mag.*, vol. 59, no. 2, pp. 10-15, Feb. 2021.

[2] S. Mihai et al., "Digital twins: A survey on enabling technologies, challenges, Trends and Future Prospects," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 4, pp. 2255-2291, Fourthquarter 2022.

[3] Y. Wu, K. Zhang and Y. Zhang, "Digital twin networks: A survey," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13789-13804, Sep. 2021.

[4] P. Öhlén et al., "Network digital twins – outlook and opportunities," *Ericsson Technology Review*, vol. 2022, no. 12, pp. 2-11, Dec. 2022

[5] P. Almasan et al., "Network digital twin: Context, enabling technologies, and opportunities," *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 22-27, Nov. 2022.

[6] S. Jang et al., “Digital twin for intelligent network: Data lifecycle, digital replication, and AI-based optimizations”, in IEEE Commun. Mag., vol. 61, no. 11, pp. 96-102 , Nov. 2023.
[7] D. Xia et al., “Digital twin and AI for intelligent planning and energy-efficient deployment of 6G networks in smart factories”, *IEEE Wirel. Commun.*, vol. 30, no. 3, pp. 171-179, Jun. 2023.
[8] M. Groshev et al., “Toward intelligent cyber-physical systems: Digital twin meets artificial intelligence”, *IEEE Commun. Mag.*, vol. 59, no. 8, pp. 14-20, Aug. 2021.
[9] B. Erman and C. Di Martino, “Generative network performance prediction with network digital twin”, *IEEE Netw.*, vol. 37, no. 2, pp. 286-292, Mar. 2023.
[10] I. U. Din, et al., “Resilient production control using digital twins in the industrial internet of things”, *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3204-3211, Feb. 2024.
[11] Y. Lu et al., “Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks,” in *IEEE Trans. Ind. Inform.*, vol. 17, no. 7, pp. 5098-5107, Jul. 2021.
[12] M. M. Roselló, et al., “Network digital twin for non-nublic networks,” *IEEE WoWMoM*, Jun. 2023, pp. 495-500.
[13] A. Rao et al., “Generalizable one-way delay prediction models for heterogeneous UEs in 5G networks,” *IEEE NOMS*, pp. 1-9, May 2024.
[14] M. L. Menéndez et al., “The Jensen-Shannon divergence,” *J. Frankl. Inst.*, 334.2: 307-318, 1997.
[15] S. K. Singh et al., “GRU-based digital twin framework for data allocation and storage in IoT-enabled smart home networks,” *Future Gener. Comput. Syst.*, 153: 391-402, Apr. 2024.
[16] O. Stenhammar, S. Sorrentino, F. Chernogorov, R. Blasco, G. Fodor and C. Fischione, ”Real-Time Network Digital Twin: Experimental Setup for Proof of Concept,” *IEEE Wirel. Commun.*, vol. 32, no. 3, pp. 114-121, June 2025, doi: 10.1109/MWC.001.2400462.