Accepted for presentation in: 2023 IEEE 97th Vehicular Technology Conference (VTC-Spring)

Data-Driven Digital Mobile Network Twin Enabling Mission-Critical Vehicular Applications

Hendrik Schippers, Stefan Böcker and Christian Wietfeld
Communication Networks Institute, TU Dortmund University, 44227 Dortmund, Germany
e-mail: {Hendrik.Schippers, Stefan.Boecker, Christian.Wietfeld}@tu-dortmund.de

***Abstract*—Future vehicular applications like Tele-Operated Driving (ToD) and Communication-Based Train Control (CBTC) pose demanding requirements on mobile communication networks. Despite continuous 5G technology upgrades and expansion strategies, mobile networks cannot provide a full-coverage service guarantee for the required mission-critical Key Performance Indicators (KPIs). However, application and location-specific Quality of Service (QoS) predictions are crucial to reliably meet the highest QoS compliance of emerging future smart city services.**

Therefore, this paper proposes a digital twin capable of merging connectivity data with arbitrary application domains to derive KPI predictions for mission-critical applications. The potential of the proposed approach is illustrated based on a case study in the urban area of Dortmund, Germany, considering data rate and latency predictions for mobile applications. In this context, a continuous data flow for the multi-dimensional mobile network twin is acquired using a massive, multimodal measurement campaign enabled by road and rail-based vehicles. This ever-growing database is utilized to analyze the KPI requirements of selected vehicular applications.

For an example ToD target zone, it is shown that a multi-Mobile Network Operator (MNO) approach increases the KPI fulfillment of direct control ToD from approximately 70 % to 90 % compared to a single MNO. By further restricting the ToD zone and combining two MNOs, a ToD-ready zone with 100 % fulfillment of the KPIs is reached.

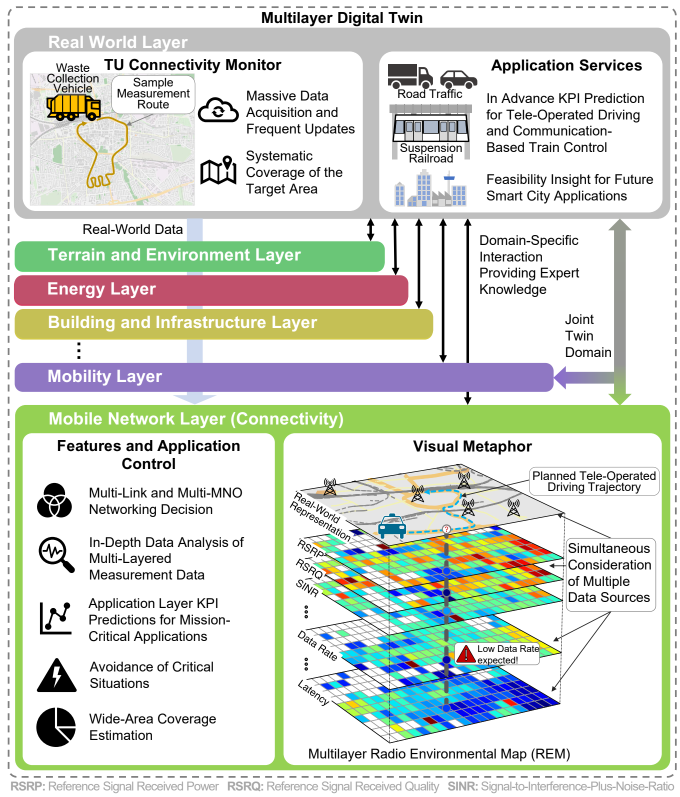


Fig. 1: Overview of a multi-layer smart city digital twin, focusing on the mobile network layer part for 5th Generation of Mobile Communication Networks (5G) networks providing expert insight to applications with the help of acquired real-world measurement data.

I. Introduction

Future vehicular applications rely on reliable and fast mobile networks, competing with an ever-growing demand by regular users [1]. Service levels are commonly not guaranteed in current 5G mobile networks, whereas client-based intelligence is needed to improve system reliability. However, typically, application domains lack the necessary expert knowledge. For this reason, a digital mobile network twin system is proposed in this paper. As shown in Fig. 1, city-wide real-world mobile network measurement data from multi-modal traffic systems is fed into the twin's multiple layers providing expert knowledge to applications as a service. Consequently, the leveraged information can be used to avoid critical situations and improve application Key Performance Indicator (KPI) compliance. That can be based on a domain-specific or a joint twin-domain interaction like, e.g., communication-aware mobility for in-advance trajectory planning.

In the mobile network layer, multi-MNO networking decisions can be assisted, multiple KPIs like the achievable latency and data rate at a specific location can be predicted, and a wide-area coverage estimation can be given based on in-depth data analysis. This layer is among other processing data containing reference signals like the Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ) but also results of active measurements like the achieved data rate at certain locations. Overall a multi-dimensional map with spatiotemporal data is constructed. Examples of future vehicular mobile network-based applications are Communication-Based Train Control

(CBTC) and Tele-Operated Driving (ToD). These require a set of specific KPIs to be full-filled all the time to work without interruptions [2], [3]. With the help of a full-stack digital twin, combining knowledge of multiple twin domains, these requirements could be predicted or even guaranteed.

The remainder of the paper is structured as follows: After discussing related works in Sec. II, the measurement approach for the data feeding the digital mobile network twin is described in Sec. III. Afterward, an overview of the methodological aspects is given in Sec. IV. Detailed results concerning instantaneous mobile network KPI predictions, selection of proper Operational Design Domains (ODDs), and multi-MNO-networking are presented in Sec. V. Finally, a conclusion and a brief outlook is given in Sec. VI.

II. Related Work

The related work part of this paper is divided into mobile network digital twin concepts and Machine Learning (ML) methods as part of the real-world simulation by digital twins.

A. Digital Network Twin Concepts and Usage

There is an emerging number of application areas for digital twins [4] and an exponential rise in publications covering digital twins [5]. Digital twins enable faster, safer and more cost-effective prototyping and optimization of products and systems [5]. One area for digital twins is mobile network behavior simulation, planning and prediction [6]. In [7], the advantages of a digital twin able to emulate 5G networks are discussed: With the help of a potent digital twin, a faster deployment by previous simulation and continuous monitoring and optimization of the network functions could be reached. That is why digital twins will also be an essential part of the future 6G networks [8].

B. Machine Learning for Non-Uniformly Distributed Data Sets

For meaningful mobile network KPI predictions, capable ML is needed. In our previous works, we gained extensive experience leveraging ML techniques for network-related predictions [9], [3]. In the case assessed in this paper, predictions for highly non-uniform data sets need to be considered. Common ML approaches for supervised learning assume uniform data set distributions. In the case of non-uniform data sets for supervised learning tasks and a model with a fixed loss function, either the data occurring with a higher probability has to be under-sampled for training, or the data occurring less often has to be over-sampled to get a uniform distribution for training synthetically [10]. Otherwise, the resulting model performance could deteriorate, depending on the exact use of the model. An algorithm for over-sampling by creating synthetic samples for classification tasks is Synthetic Minority Over-sampling Technique (SMOTE) [11]. In the case of regression tasks, Synthetic Minority Over-Sampling Technique for Regression with Gaussian Noise (SMOGN) can be utilized [12]. Both algorithms create new samples by combining the values in the vicinity of the feature space of each new sample in the lower training sample density regions.

Low-density regions in the training set feature space are often exceptional cases or anomalies that can not be generalized onto the whole data set. That is why the resulting labels of these samples can commonly not be predicted by a single strong learner. Instead, ensemble learners need to be utilized. By combining trained decision trees additive, several weak learners are combined into a strong learner, called boosting. Xtreme Gradient Boosting (XGBoost) is such a tree-based boosting method. It is a scale-able and high-performance ML method [13] originating to [14]. By minimizing the cost function C consisting of a loss part l dependent on the prediction error and a regularization term Ω punishing the complexity of the resulting model, an accurate prediction model, simultaneously preventing over-fitting, is created.

$$C = \sum_{n=1}^{N} l(\hat{y}_n, y_n) + \sum_{k}^{K} \Omega(t_k) \tag{1}$$

C is evaluated from iteration to iteration on the N training samples. The loss part of C is calculated by the sum of the loss of each sample $l(\hat{y}_n, y_n)$ based on the model output $\hat{y}_n$ and the ground truth value y_n. The regularization part is calculated as the sum of the regularization term for every tree t_k out of K trees. $\Omega(t_k)$ punishes the number and the weights of leaves. C is altered in the exact implementation to allow for an efficient calculation [13]. The resulting model of decision trees is evaluated by calculating the sum of the individual results of the trees, enabling reliable common and special case predictions.

III. Proposed Mobile Network Digital Twin Approach for Application Services

This work proposes the mobile network layer of a smart city mobile network twin architecture capable of providing expert knowledge of mobile networks to future applications. It relies on real-world input data to predict unseen events and control entities to improve overall system performance. As shown in Fig. 2, real-world and virtual-world objects are combined in a digital twin interface.

Real-world measurements of suspension railroad and waste collection vehicles are automatically inserted into the twin's intrinsic database with the help of the Connectivity Monitor (ConMon) application developed for this purpose. This way, the quality of public 4G and 5G Non-Standalone (NSA) networks is frequently monitored, and the twin is updated. Based on the accumulated data, specifically developed algorithms provide services for external applications like ToD or CBTC. This paper focuses on predictive QoS for the KPIs latency and data rate. The spatiotemporal predicted latency and data rate [9] combined with a proper ODD selection enables communication-aware mobility by exposing the prediction results to external applications needing to comply

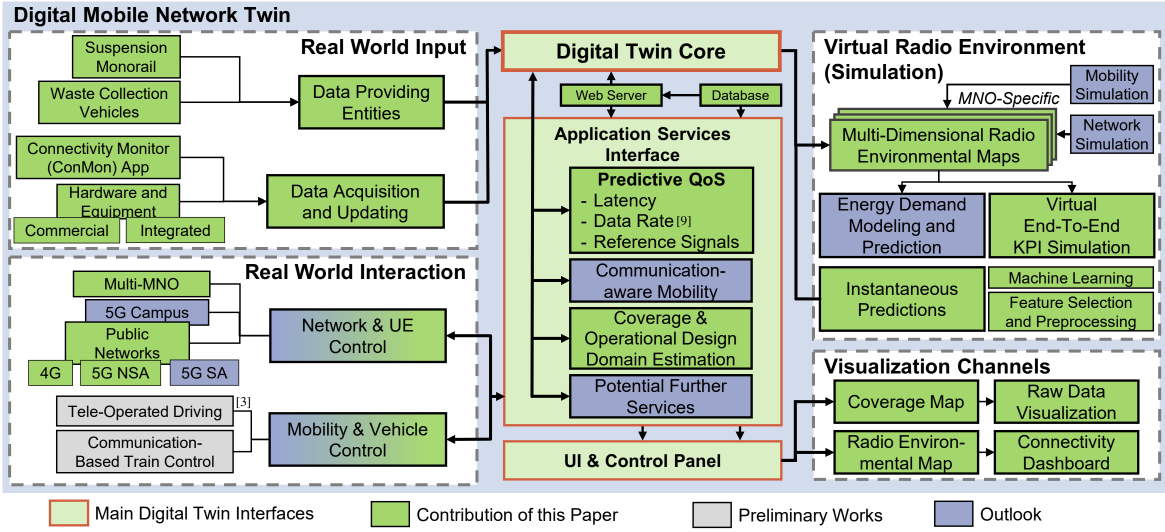


Fig. 2: System overview of the data-driven digital mobile network twin developed in this paper.

with strict KPI constraints. In [3], the data rate prediction for ToD applications based on multi-dimensional Individually Tuned Radio Environmental Maps (IREMs) has already been discussed. These MNO-specific REMs are constantly updated within the digital twin and form the core of the virtual radio environment simulation. Traffic and network simulation data can be included in these Radio Environmental Map (REM) layers to improve the simulation accuracy further. Virtual end-to-end KPI predictions relying on ML methods and a careful feature selection and preprocessing step are fed with the data of the REMs. If specific REM data is unavailable, instantaneous predictions for KPI simulations, as in [9], are possible. In the future, the discussed simulations will be extended by an energy demand modeling dependent on the radio channel.
Furthermore, this system could also be utilized for 5G campus networks. In this case, not only can the mobility of smart city applications be managed, but also autonomous or tele-operated logistics vehicles and production robots. Finally, the gained knowledge and control of the mobile network via the twin can be visualized to the user in a condensed form with the help of a Control Panel and User Interface (UI) entity.

IV. Methodology of the Measurement Campaign Proving Data for the Proposed Digital Twin

The proposed data-driven digital mobile network twin system can process multi-modal measurements to comprehensively cover the Dortmund city area. As shown in Fig. 3, two variants of measurement equipment have been developed. The first is based on commercial smartphones equipped with a capable measurement application. This `ConMon` application automatically measures the mobile network quality and regularly reports the results to the digital twin. Within the scope of a cooperation with the local waste disposal company, six smartphones equipped with `ConMon` are simultaneously installed in garbage collection vehicles. The phones are vertically mounted on the windshield of the vehicles. By systematically emptying the garbage cans of each household, it is guaranteed that a good exploration of the city area is reached and a frequent update of the measurement results is performed. The garbage collection vehicles repeat their measurement sectors every two weeks. These are fixed and mostly adjacent to each other, as shown in Fig. 4. Consequently, to cover all MNOs in all sectors of the target area, the used SIM cards are frequently exchanged between vehicles.

Additionally, an embedded measurement equipment for outdoor purposes has been developed. It is based around a System on a Chip (SoC), connected to a 5G modem and a Global Navigation Satellite System (GNSS) module providing accurate location information. The modem is connected to a vehicular 4x4 Multiple Input Multiple Output (MIMO) and GNSS antenna. As visible in Fig. 3, extensive shielding measures have been taken to protect the GNSS signal from USB3 irradiation. This equipment is installed on the local suspension railroad, as shown in Fig. 3 on the bottom right. The suspension railroad is autonomously driving on dedicated suspended tracks, which could - like other rail systems - otherwise not be covered by garbage collection vehicles. This multi-modal approach ensures a superior coverage of the target area compared to single-vehicle approaches. In the future, it is planned to extend the measurements further to cover an even larger area of the city.

The measurement applications are set up to measure a multitude of parameters, as listed in Tab. I. These indicators can be divided into passive and active indicators. While the acquisition of passive indicators like reference signals is comparably fast and does not require any data transmissions, active parameters like the data rate rely on energy-intensive measurements taking up to several seconds to perform. Passive parameters can be further divided into time-varying reference signals and cell-specific parameters, which are fixed for each cell. Furthermore, other non-mobile network-related parame-

Fig. 3: Demand-driven connectivity measurement approach for application in multi-modal traffic systems to realize a data-driven digital mobile network twin in the Dortmund city area.

ters like the GNSS location and time are also logged.

More elaborate active measurements are executed less frequently to reduce the environmental impact. Data rate measurements that are even more impact-full on the shared mobile channel are further reduced compared to latency measurements, see Fig. 4. Additional measures to save energy and dispense unnecessary measurements are adopting the measurement frequency by the current velocity and the interruption of active measurements if the current location can not be accurately determined.

As described in Sec. IV, the measurement vehicles cover a set measurement sector each working day. This procedure results in a daily driven distance of more than 200 km. The driven trajectories are part of the systematical garbage collection, and thus it is guaranteed that a good exploration of the city is reached by every driven kilometer. Mobile network measurements frequently cover over 1 300 km of the Dortmund city area. The measurements consist of over 17 000 passive reference signal measurements every day, as shown in Fig. 4. By extrapolating the number of measurements per day, it becomes clear that the scope of the measurement campaign is massive and can be utilized to adequately feed a digital twin of the mobile network in the target area. Coverage of over 70 % is already reached after approximately three months of measurements (80 m grid). An even higher coverage has been reached depending on the needed resolution or cell width of the measurements.

V. Reliability Evaluation for Remote Control Applications in the Dortmund Smart City Area

The results of the measurement campaign shall be evaluated in the scope of a case study about Tele-Operated Driving (ToD) and Communication-Based Train Control (CBTC) as examples of smart city applications. As described in [3], concerning the 5G Automotive Association (5GAA), ToD requires a latency from as low as 20 ms in high speed `direct control`

TABLE I: Parameters and Repetition Rate of ConMon Application.

Parameters	Description and Interval
Passive Network Parameters	Every second & on change
RSRP, RSRQ, SINR, SS-RSRP, SS-RSRQ, SS-SINR, RSSI, CQI, TA, neighboring cell information, ...	Time-varying reference signals for 4G and 5G cells
Bandwidth, frequency, cell index, physical cell index, tracking area code	Cell-specific parameters
Active Parameters	Every 5 seconds
Latency	ICMP latency to 5 servers
Data Rate	Up- & downlink, TCP & UDP, Varying file sizes for TCP
Other Parameters	Every second
GNSS location, velocity, bearing, device name and status, timestamp	Device, location and context information

and up to 300 ms in slow driving `indirect control` scenarios. This paper assumes a value of 100 ms for urban ToD scenarios. In [2], a maximum latency of 100 ms that needs to be reached for 99.999 % of the time in a configuration of two independent connections at the ends of the train [2] is named for CBTC to be required. These performance requirements are shown in Tab. II. Summarized, CBTC with fewer degrees of freedom is slightly less demanding for the mobile network as ToD.

In contrast to CBTC, one or multiple live video feeds in combination with precise steering capabilities are needed for ToD requiring higher data rate requirements especially in the uplink direction. For the case study, four video streams with a bit rate of 8 MBit/s each, as described in [15], are assumed for `direct control` ToD, resulting in a required data rate of 32 MBit/s. With the help of the measurement approach introduced in Sec. III, the feasibility of the more demanding ToD shall be evaluated based on the KPIs latency and data rate. As the downlink data rate requirements for ToD are much lower than in the uplink direction, only the uplink data rate is considered.

A. Measurement Results in the Target Area

The results of the measurement campaign yield that a Round-Trip Time (RTT) of under 100 ms needed for ToD is reached over 98.2 % to 98.9 % of the cases, depending on the MNO. The distribution of the minimum RTT of four subsequent `ping` measurements varies, especially in the range over 40 ms, resulting in more high latency occurrences at MNO B, which is also depicted in higher percentiles (see

TABLE II: Application requirements for Tele-Operated Driving (ToD) and Communication-Based Train Control (CBTC).

Application	Latency	Reliability	Data Rate Uplink	Data Rate Downlink
CBTC [2]	100 ms	99.999 %	30 - 150 kbit/s	
Direct ToD [15]	~100 ms	99.999 %	3 - 50 Mbit/s	0.25 - 5 Mbit/s
Indirect ToD [15]	300 ms	99 %	8 - 30 Mbit/s	< 0.3 Mbit/s

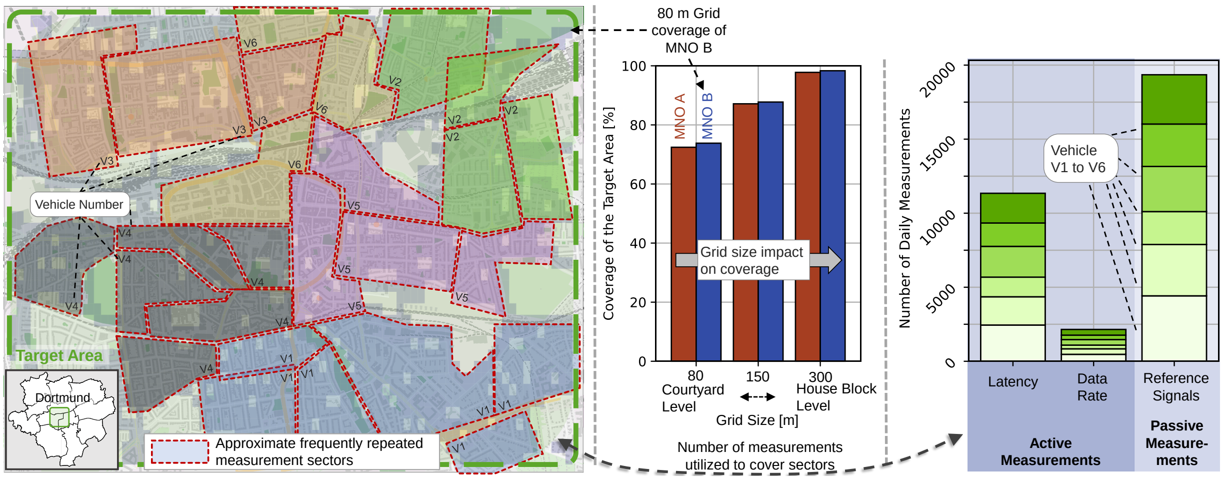


Fig. 4: Target area of the digital twin with approximate repeated daily measurement areas of the six measurement vehicles V1 to V6, already covered percentage of the target area and the daily number of conducted active and passive measurements to achieve this coverage.

Fig. 5). RTT measurements to four public Domain Name Servers (DNSs) (`8.8.8.8`, `8.8.4.4`, `1.1.1.1` and `1.0.0.1`) via the `ping` command were conducted to exclude errors due to individual latency anomalies for a specific server. The minimum RTT of four subsequent `ping` measurements is utilized in this paper. In a dual-connectivity configuration via MNO A and MNO B, the resulting reliability of a latency of under 100 ms is 99.98 %, provided the latency at both MNOs is statistically independent. This reliability still needs to be improved in the future to comply with the requirements in [2], [15]. However, in the case of ToD, in our measurements, the assumed data rate constraints were more critical than latency constraints.

B. Latency Prediction for Mission-Critical Mobile Applications

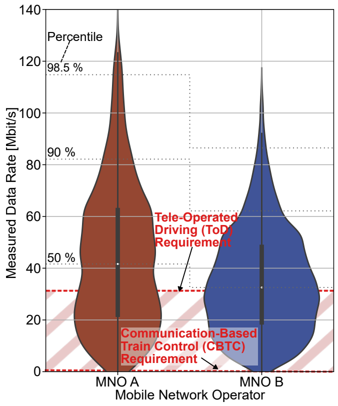
The latency and the data rate can be predicted based on passively measured parameters. Mission-critical applications must adapt to communication channel changes, especially deterioration of the achievable data rate and latency. That is why predicting the future latency could improve these applications' Quality of Experience (QoE) by in-advance warning the user if a high latency control phase is incoming. In this case, for example, in a ToD scenario, the vehicle's velocity could be adopted to prevent a latency-triggered emergency stop. Effectively, the failure probability of the system could be reduced by the portion of correctly predicted samples.

As the latency is not equally distributed over its value range, see Fig. 5b, SMOGN is applied to the training set of the ML algorithm, which contains 80 % of the data. This procedure aims to improve the prediction of rare high-latency cases and reduce underestimations, as described in Sec. II. As rare high latency events are not predictable by strong learners, XGBoost is utilized in this work. Hyperparameters of the ML setup are tuned using a random grid search approach with 20 iterations, including a ten-fold cross-validation each. After this, the best estimator is used to predict the unseen 20 % test set. The same procedure is used for data rate predictions.

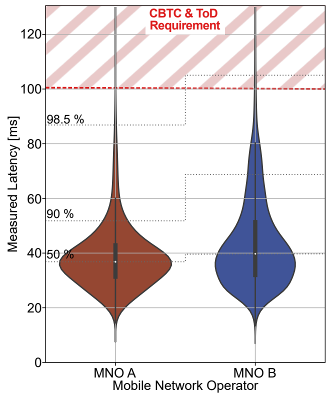
In Fig. 5d, the measured latency for MNO B is scattered above the predicted latency of the test set. It can be seen that the prediction accuracy decreases at high latency values. This behavior can be explained by the rare occurrence of high latency values in combination with effects beyond the radio access network. Despite a potential high activity in the mobile network being, e.g., reflected in a reduced RSRQ, an overall higher activity in the backbone of the internet can not directly be derived from measured values at the User Equipment (UE). In the case of data rate predictions, the correlation between the size of the label and the prediction error is smaller (see Fig. 5c).

In Fig. 6, the feature importance of the ML model is evaluated by SHapley Additive exPlanations (SHAP). The framework SHAP, introduced in [16], enables the interpretation of the importance and the influence of features based on their feature value. The features are ordered by their impact on the model output from top to bottom. Regarding the data rate prediction, the RSRP and SS-RSRP are highly important. That could be due to the signal strength being an indicator for possible Modulation and Coding Schemes (MCSs) and such for the achievable data rate. However, features like the time of day, the cell index and the timely averaged RSRQ, which give insight into the mobile network utilization, also gain considerable importance. These features representing the utilization of the mobile network are even more critical for latency predictions. Here, the time of day, cell index and RSRQ all gain high importance. In [17], latency prediction for LTE measurements by Random Forests (RFs) and decision trees were performed, resulting in a single high dependence on the RSRP with a mean impurity decrease of over 0.89. This relationship cannot be shown in the performed 5G NSA latency predictions.

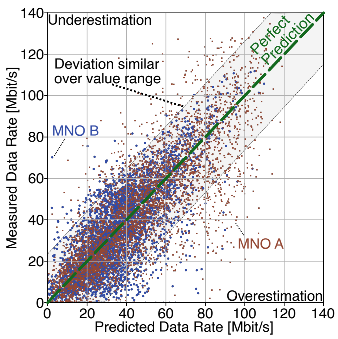
When predicting the 5G latency, the Secondary Signal



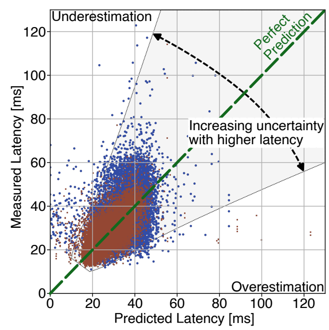
(a) Measured UDP Uplink Data Rate



(b) Measured ICMP Echo ("Ping") Round-Trip Time



(c) UDP Uplink Data Rate Prediction



(d) ICMP Echo ("Ping") Round-Trip Time Prediction

Fig. 5: Measured MNO-dependent data rate and latency of the complete data set compared with application layer requirements on the top. Scatter plots of the measured KPIs over predicted values (hold-out data set) on the bottom.

RSRP (SS-RSRP) and the Timing Advance (TA) behave partly contrary to the case of the data rate prediction. The SS-RSRP is not only an indicator for the modulation order but also an indirect measure of the kind of environment. As the Internet Control Message Protocol (ICMP) RTT measured with the `ping` command uses a small payload size, the latency is less dependent on a high MCS and, thus, less dependent on the signal strength. Low SS-RSRP is much more likely to occur in less crowded suburban or rural areas, which might result in reduced latency. This connection also exists for the TA, as a high TA is rarely measured in urban areas with an elevated cell density.

The RSRQ can partly depict the mobile network activity, as the signal quality decreases with the number of active users in a cell. However, the RSRQ is only coarsely quantized for a cell utilization prediction, and the cell utilization can fluctuate strongly with time. That is why the averaged RSRQ over 5 s and 500 s are added as features for the prediction model. The averaging effect increases the resolution of this feature, which could be why the ML model also relies on the 5 s averaged signal quality and the 500 s averaged (SS-)RSRQ in addition to the values directly obtained from the `Android` API. It has also been tried to utilize the available neighboring cell reference signals for latency predictions. As these parameters only yield an insignificant feature importance, the influence of neighboring cells is likely not strongly affecting the measured minimum RTT in the considered target area.

C. *Case Study: ToD Target Zones in a Multi-MNO Approach*

Current consumer-oriented mobile networks are not primarily designed to serve ToD applications. Consequently, at some locations for both MNOs in the Dortmund ring area, representing a potential ToD target zone, ToD requirements are not met in our measurements. That could result in service interruptions and emergency stops without further evolution of the mobile network. KPI predictions can prevent the latter - as shown in Fig. 5 - but a significant QoS restriction would be unavoidable. That is why a ToD-ready zone is defined to cover parts of the Dortmund city ring, where ToD requirements could already be fulfilled when the data was gathered.

The concept of ODDs is used in the context of autonomous driving [18] and is adopted for this paper to indicate geographical zones in which ToD requirements are fulfilled. For each 25 m grid cell, the direct ToD scenario's data rate and latency requirements are compared with the mean of the measurements and are visualized in Fig. 7. While most parts of the Dortmund city ring might be part of a future ToD ODD, at the moment, a potential ToD-ready zone could, e.g., already extend over parts of the northern ring, fulfilling the direct ToD scenario's requirements in a multi-MNO setup. In the ToD-ready zone, both MNOs alone reach a coverage of 79 % in our measurements. However, the combination of both MNOs yields a coverage of 100 % in this scenario. Different scenarios have been evaluated as results differ with the mobile networks and the exact requirements. While the challenging Direct ToD Scenario requirements can already be fulfilled in 90 % of the grid cells in the ToD target zone with the network deployment measured at the beginning of 2023, the less demanding CBTC

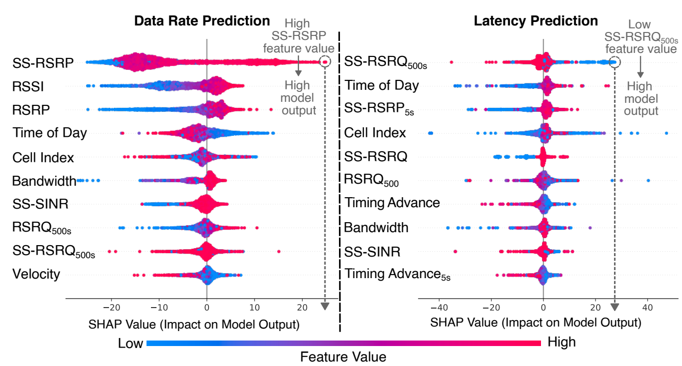


Fig. 6: SHAP feature impact for MNO B on the model output for the ten most important features comparing data rate to latency predictions.

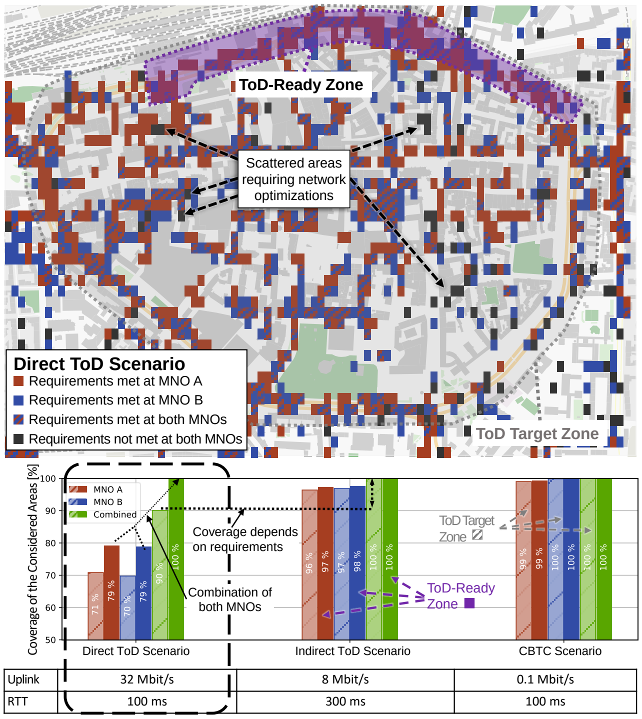


Fig. 7: Case study of a possible Tele-Operated Driving target zone covering the Dortmund city ring and an already ToD-ready zone for different scenarios. Compliance with KPI requirements for both MNOs in a 25 m grid is evaluated.

Scenario requirements are fulfilled in 100 % of the grid cells in both the defined zones.

For realistic deployments, our methodology enables the selection of ToD capable network areas, indicates areas requiring further network optimizations and shows the potential of combining multiple networks to increase reliability and extend the area of potential ToD zones. Consequently, this methodology might accelerate the deployment of ToD-capable mobile networks.

VI. Conclusion and Outlook

This paper introduced a data-driven digital mobile network twin enabling mission-critical vehicular applications by providing crucial mobile network KPI predictions. These spatiotemporal predictions are a challenging but crucial task to enable critical applications securely. Effectively predicted network KPIs enable proactive adaptions of services and networks to increase the overall QoS. Additionally, selecting proper ToD ODDs in combination with multi-MNO approaches increases the feasibility of future ToD applications.

Due to the promising results, it is planned to extend the measurement campaign to additional data sources to cover other times of day and mobility behaviors and improve prediction accuracy. Especially the prediction of rare high latency cases will need to be addressed to reliably prevent sudden QoS degradations even more efficiently.

Further improvements of the achieved KPIs could be achieved via network slicing and extended multi-MNO networking. However, advanced network planning and immense network extension measures are needed for mass-market ToD applications in extended areas. In the future, 6G multi-X (X substitutes bands, radio access technologies, etc.) strategies shall be evaluated to be used for mission-critical applications. Whereas this publication concentrates on public mobile networks, future work may also consider dedicated private 5G networks and apply the discussed REM-based digital twins to them.

Acknowledgment

This work has been supported by the Federal Ministry of Transport and Digital Infrastructure (BMVI) in context of the project *Virtual integration of decentralized charging infrastructure in cab stands* under the funding reference 16DKVM006B and by the Ministry of Economic Affairs, Innovation, Digitalization and Energy of the state of North Rhine–Westphalia (MWIDE NRW) in the course of *5Guarantee* under grant number 005–2008–0046.

References

[1] Ericsson, “Ericsson mobility report november 2022,” Stockholm, Sweden, Report, Nov. 2022.
[2] ETSI, “ETSI TR 103 442 V1.1.1,” ETSI, Tech. Rep., May 2016.
[3] H. Schippers, C. Schüler, B. Sliwa, and C. Wietfeld, “System modeling and performance evaluation of predictive QoS for future tele-operated driving,” in *2022 IEEE International Systems Conference (SysCon)*, Apr. 2022, pp. 1–8.
[4] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, “Digital twin in industry: State-of-the-art,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.
[5] M. Singh, E. Fuenmayor, E. Hinchy, Y. Qiao, N. Murray, and D. Devine, “Digital twin: Origin to future,” *Applied System Innovation*, vol. 4, no. 2, p. 36, May 2021.
[6] Y. Wu, K. Zhang, and Y. Zhang, “Digital twin networks: A survey,” *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 13 789–13 804, Sep. 2021.
[7] H. X. Nguyen, R. Trestian, D. To, and M. Tatipamula, “Digital twin for 5G and beyond,” *IEEE Communications Magazine*, vol. 59, no. 2, pp. 10–15, Feb. 2021.
[8] T. Wild, V. Braun, and H. Viswanathan, “Joint design of communication and sensing for beyond 5G and 6G systems,” *IEEE Access*, vol. 9, pp. 30 845–30 857, Feb. 2021.
[9] B. Sliwa, H. Schippers, and C. Wietfeld, “Machine learning-enabled data rate prediction for 5G NSA vehicle-to-cloud communications,” in *2021 IEEE 4th 5G World Forum (5GWF)*, Oct. 2021, pp. 299–304.
[10] Y. Yang, K. Zha, Y.-C. Chen, H. Wang, and D. Katabi, “Delving into deep imbalanced regression,” in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 139. PMLR, Jul. 2021, pp. 11 842–11 851.
[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002.
[12] N. Kunz, “SMOGN: Synthetic minority over-sampling technique for regression with gaussian noise,” PyPI, 2020.
[13] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 785–794.
[14] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Statist. 29(5)*, Apr. 2001.
[15] T. Linget, “C-V2X use cases volume II: Examples and service level requirements,” 5GAA Automotive Association, Techreport, Oct. 2020.
[16] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS’17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 4768–4777.
[17] A. S. Khatouni, F. Soro, and D. Giordano, “A machine learning application for latency prediction in operational 4G networks,” in *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, Apr. 2019, pp. 71–74.
[18] 5GAA, “Safety treatment in connected and autonomous driving functions report,” 5GAA Automotive Association, Tech. Rep., Sep. 2020.