Digital Twin-driven End-to-End Network Slicing Towards 6G

Mahnoor Yaqoob, Ramona Trestian, Mallik Tatipamula, *Senior Member, IEEE,* and Huan X. Nguyen, *Senior Member, IEEE*

***Abstract*—The diverse use cases requirements and strict expectations of latency on shared infrastructure, from future networks, calls for networking paradigms that can provide efficiency and flexibility. Network slicing is a network paradigm introduced with 5G, which is a key in realizing diverse service requirements and meeting expectations of future 6G networks. A Network Digital Twin (NDT) model can benefit the management and orchestration of network slicing life cycle and provide a clear visual representation of the physical entities of a network slicing model while enabling experimenting with different resource allocation schemes without actually affecting the physical network. A NDT bridges the gap between physical world and digital world by having constant bi-directional communication. Hence, in this article we make a case for NDT as a key enabler of network slicing in 6G network while integrating Artificial Intelligence (AI) based network slicing strategy to enable end-to-end (E2E) Quality of Service provisioning.**

***Index Terms*—Network slicing, Digital Twin, 6G, Artificial Intelligence.**

I. Introduction

INDUSTRY 5.0 will enable autonomous manufacturing with human intelligence in and on the loop creating an ecosystem of coexistence between people, processes and machines as illustrated in Fig. 1. This is achieved by integrating various key enabling technologies, such as: Artificial Intelligence (AI), Big Data Analytics, next generation networks, edge/cloud/fog computing, Industrial Internet of Things (IIoTs), Digital Twins (DTs), etc [1]. The integration will help at extending the capabilities of both machines and people, to optimize efficiency and operations within various industries. The continuous industrial adoption of DT technology represents a great potential to reshape the future across diverse industries and elevate the level of virtual interactions with the physical environment. Even though this digital transformation across various industries will enable applications that serve different purposes, they all have something in common: dependency on reliable and strong connectivity. This ever-increasing service provisioning, demands for an adaptive, flexible and intelligent network architecture which is in direct contradiction of today's "one size-fits-all" network design paradigm.

According to the definition of 3GPP TR 28.801, a Network Slice (NS) consists of network functions and supporting network resources that are arranged and configured to form a complete logical network that meets specific network characteristics required by a service instance. The efficiency and flexibility in service provisioning is enabled by the integration of Software Defined Networking (SDN) and Network Function Virtualization (NFV) [2]. In order to achieve a desired capital expenditure (CapEx) and the operational expenditure (OpEx) for a service, particular set of virtual network functions (VNFs) and virtual links are embedded on substrate network forming a service function chain [2].

The vision of 6G promises a highly intelligent society that will deal with diversification of service platforms. Improving the resource consumption of network slicing will become vital to keep the operating cost as low as possible. Different verticals will have contradicting requirements deployed on the same shared infrastructure. In 6G we anticipate that VNFs will be cloud native distributed across heterogeneous technological domains such as radio access network (RAN), transport network (TN), core network (CN), cloud, and edge computing domains [3], [4]. This heterogeneity and diversity of technological domains and services will result in a high number of objects that need to be managed. In this context, the life cycle management (LCM) of E2E network slices requires self-optimizing solutions that minimize human intervention and provide efficient recovery in case of service degradation while maintaining resource optimization. Establishing Quality of Service (QoS) and Quality of Experience (QoE) based network slices is one of the most critical challenges even for the very promising 6G networks.

To overcome these challenges, it is vital for 6G networks to manage and orchestrate network slices in a hierarchical and distributed manner. Moreover, the increased number of services in 6G will result in massive number of network slices sharing same infrastructure with different requirements, this puts substantial strain on the management and orchestration systems and traditional centralized solutions and NFV cannot cope with such diverse requirements. In this article, we devise to overcome above mentioned challenges by proposing a novel decentralized DT framework that has potential to cope with the envisioned massive number of connected users and services in 6G while leveraging state-of-the-art Graph Neural Network (GNN) to ensure autonomy towards zero touch management. The proposed DT framework is novel because it replicates the physical network elements and functions in DT layer in order to deploy test network scenarios without disrupting the real network. The proposed DT is also capable of learning from network traffic while predicting the Key Performance

M. Yaqoob, R. Trestian, and H. X. Nguyen are with the London Digital Twin Research Centre, Middlesex University London, UK. H. X. Nguyen is also with International School, Vietnam National University, Hanoi, Vietnam. M. Tatipamula is with Ericsson Silicon Valley, USA.

This work was supported by the funding under the EPSRC UK-India Future Networks Initiative (Grant ref - EP/W016524/1) led by the University of East Anglia (UK).

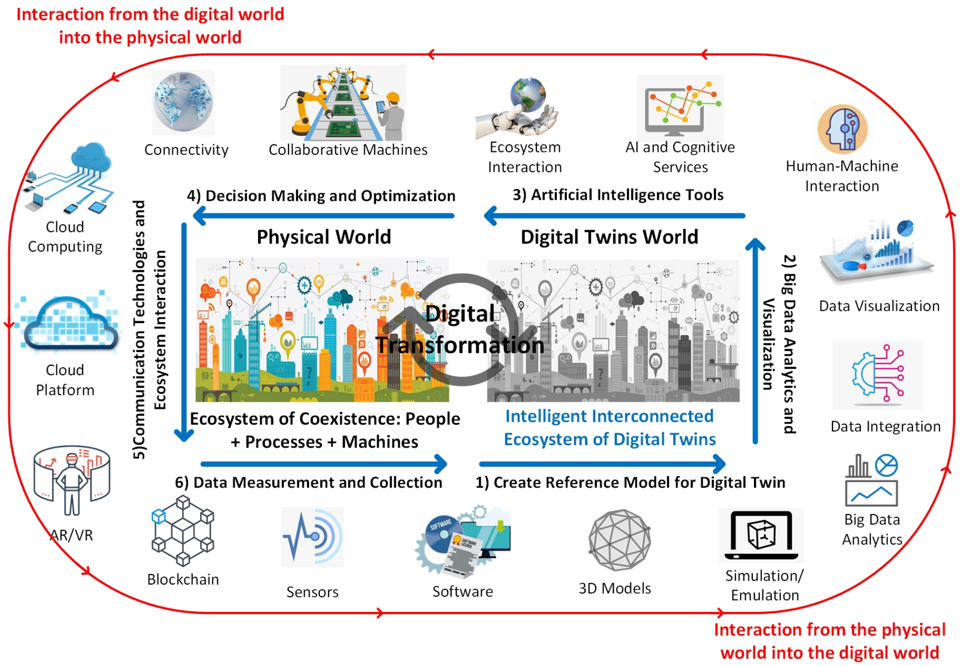

Fig. 1. Enabling Digital Transformation Through an Intelligent Interconnected Ecosystem of Digital Twins.

Indicators (KPIs) in order to ensure QoS provisioning.

II. Related Works

A. Network slicing in 6G

Ever since the early deployment of 5G networks, researchers have shifted their interest towards formalizing standards for future 6G networks. Recently, the European Telecommunications Standards Institute (ETSI) has launched two groups, Zero-touch Service Management (ZSM) [5] and Experiential Networked Intelligence (ENI) [6], with the aim to use AI and ML to achieve agile and completely automated orchestration and management of network resources. ETSI has launched a reference architecture that features distributed management and orchestration of resources. However, the proposed architecture ignores the management of network slices. ETSI ENI focuses on centralized framework to standardize different AI and ML policies to manage the network. Both groups are focusing on one part of the problem only, while ignoring how slices interact with each other as well as E2E service provisioning is also not catered for. To this extent, the proposed DT model can extrapolate useful information from the network data and learn the routing patterns to predict the E2E delay of network slices. This predictive capability makes the DT model novel because the accuracy of prediction is the first step towards anomaly detection and making preventive decisions without disrupting the existing service. Chergui et al. [7] proposed a federated learning approach for energy efficient network slicing for 6G. The federated learning engine makes predictions for slice level resources by learning in an offline manner while taking Service Level Agreement (SLA) constrains into account. Meanwhile, the authors in [8] have taken RAN slicing one step further by proposing self-sustained RAN slicing architecture for 6G. They split the network resources efficiently among different levels of architecture by introducing separate control policies for each network level, gNode level, and packet scheduling level. But they only focused on RAN slicing and other network domains are not included in the proposed solution. In this article we focus on a complete hierarchical DT architecture which captures the network behavior from end user to core network domain. Moreover, Deep Reinforcement Learning (DRL) is utilized in [9], for intelligent resource management in 6G for customized network slicing. The network scenario considered includes only 10 nodes which is not an accurate depiction of 6G network, in contrast our DT model caters for much more complex network scenarios with hundreds of connecting nodes. Wang et al. [10] proposed a network DT model based on GNN for efficient management of network slicing and to predict the E2E latency of individual network slices. This approach is a black box testing while we propose a complete DT of the network infrastructure to accurately map the network rather than using abstractions.

B. Towards 6G with Digital Twin and AI

Digital Twin is an emerging technology represented by a system-of-systems which goes far beyond the traditional computer-based simulations and analysis [11] and has the potential to reshape the future of industries and society overall. A DT is essentially a digital replica of a physical object or entity, where DT evolves synchronously with physical entity throughout its life cycle. With the advances of supporting technologies, e.g., AI, IoT, data visualisation, DT has great potential to provide for a digital environment where future generation networks like 6G can evolve. Integrating DT within mobile networks is gaining popularity in industry from major tech companies like Ericsson, Huawei and Nokia [12]. In this context, DT has the capability to continuously monitor and analyze the performance of the network, predict any unanticipated failures and optimize the network performance by triggering intelligent decisions accordingly. Thus, allowing to lay a platform for 6G to achieve its visions in: i) *connected intelligence* (large-scale deployments of intelligence in the wider society); ii) deploying complex *network-of-networks* (through the DT system-of-systems concept); iii) *extreme user experience* (e.g., internet of senses, or fully immersive communication through virtual platform); and iv) *trustworthiness and sustainability* (ensure the confidentiality and integrity of E2E communications). DT for networks is different from DTs in industrial domain, as network DT has to merge the physical and virtual components since NS is already based on softwarized and virtualized components. These virtual components impact towards complexity of building the scalable DT. Fortunately, the advances in AI and ML have led to emergence of data driven techniques, that can learn from network traffic data instead of being programmed explicitly. AI especially Deep Learning (DL) [13]–[15] is undeniably one of the key drivers for NS in 6G networks. Hence, leveraging DT technology following a merger with AI algorithms, for complex network structures can significantly improve the performance of networks and help predict future failures.

III. Proposed DT framework for NS in 6G

The proposed DT framework is shown in Fig. 2 and includes three abstract layers namely: physical infrastructure layer, DT for network slicing layer and service layer.

A. Physical Infrastructure Layer

E2E network slicing and service provisioning in 6G can be achieved by virtualization of network resources, VNF placement, leveraging virtual machines (VM) placement and management while making intelligent decisions based on customized AI algorithms. Physical network infrastructure layer consists of three network domains, considering an E2E slice is composed of several interconnected VNFs from RAN, TN and CN. In the proposed framework all the components of 6G are set to be fully abstracted while supporting softwarization and virtualization technologies, in order to efficiently orchestrate and manage all network functions, network resources and network traffic data of 6G networks.

B. Digital Twin for Network Slicing Layer

The composition of DT layer is partitioned into several steps, each step having specific set of tasks. For instance step 1 is to create virtual counter parts of physical network, step 2 is to implement NFV and deploy VNFs while step 3, 4, 5, 6, 7 and 8 are a part of centralized SDN controller which is governed by AI algorithms. The role of each step is detailed as follows:

Virtual Twins: In DT layer the first step is to have a platform where counter twins of physical network domain can be created. This can be achieved by using existing emulation and simulation tools provided that there is communication between physical and virtual elements. DT can make use of emulated network traffic to mimic the network behavior in virtual world.

NFV in DT: We propose the NFV block in DT in compliance with European Telecommunications Standards Institute (ETSI) defined Management and Orchestration architecture. Each network domain has a resource pool and virtualization layer which is responsible for mapping the physical resources into logical or virtual resources (VNFs). The DT layer includes the two function blocks of VNF Management and Orchestration defined by ETSI: Virtualization Infrastructure Manager (VIM) and VFN Manager (VNFM). The goal of VIM and VNFM is to communicate with each other to manage the life cycle of VNFs which includes instantiating, migrating, modifying and decommissioning of VNFs.

AI based SDN Controller in DT: A centralized SDN controller powered by AI algorithms, is deployed which interacts with VNFM in order to support services with differentiated QoS requirements as well as to ensure programmable network configuration.

Network Life Cycle Management: With the help of SDN controller and NFV we can manage the life cycle of all network slices, for instance slice preparation, slice deployment, slice management and slice decommissioning.

DL for Expert Knowledge: Once the network resources are allocated and VNFs are placed to create the slices, this data goes through the AI block where DL algorithms are used to learn from the network behavior and further optimise the resource allocation. Based on the network data gathered from network life cycle expert knowledge is acquired by extracting network features that are a key in predicting performance measures of network.

KPI Prediction: After processing the network traffic through DL algorithms and acquiring expert knowledge we need ML algorithms to accurately predict the network KPIs like latency, delay, jitter and throughput, which is vital for E2E service provisioning in 6G. This can be done by leveraging GNNs [10] they are ideal to capture the complex network behavior and learn from it.

What if Scenarios: After successfully mimicking the network traffic and and KPI prediction DT can be used to forecast the future traffic, latency and bandwidth requirements by testing out scenarios like network traffic in hospital in case of outbreaks, strict latency requirements in case of remote surgery, enhanced user connectivity in case of virtual conferences. All these scenario can be tested beforehand in DT saving time as well as money. Because 6G has such higher expectations and strict latency requirements DTs are vital to forecast all the network behavior before deployment.

Customized Slices: Once the DT has learned from virtual scenarios slices can be customized based on the requirement of each type of slice. Dedicated AI algorithms can be used to deploy and optimize each type of slice. Because DL and AI algorithms require a lot of time in training, it would not be ideal to implement these algorithms in real time. Therefore by offline training in DT we can achieve the best results without affecting the real time service provisioning.

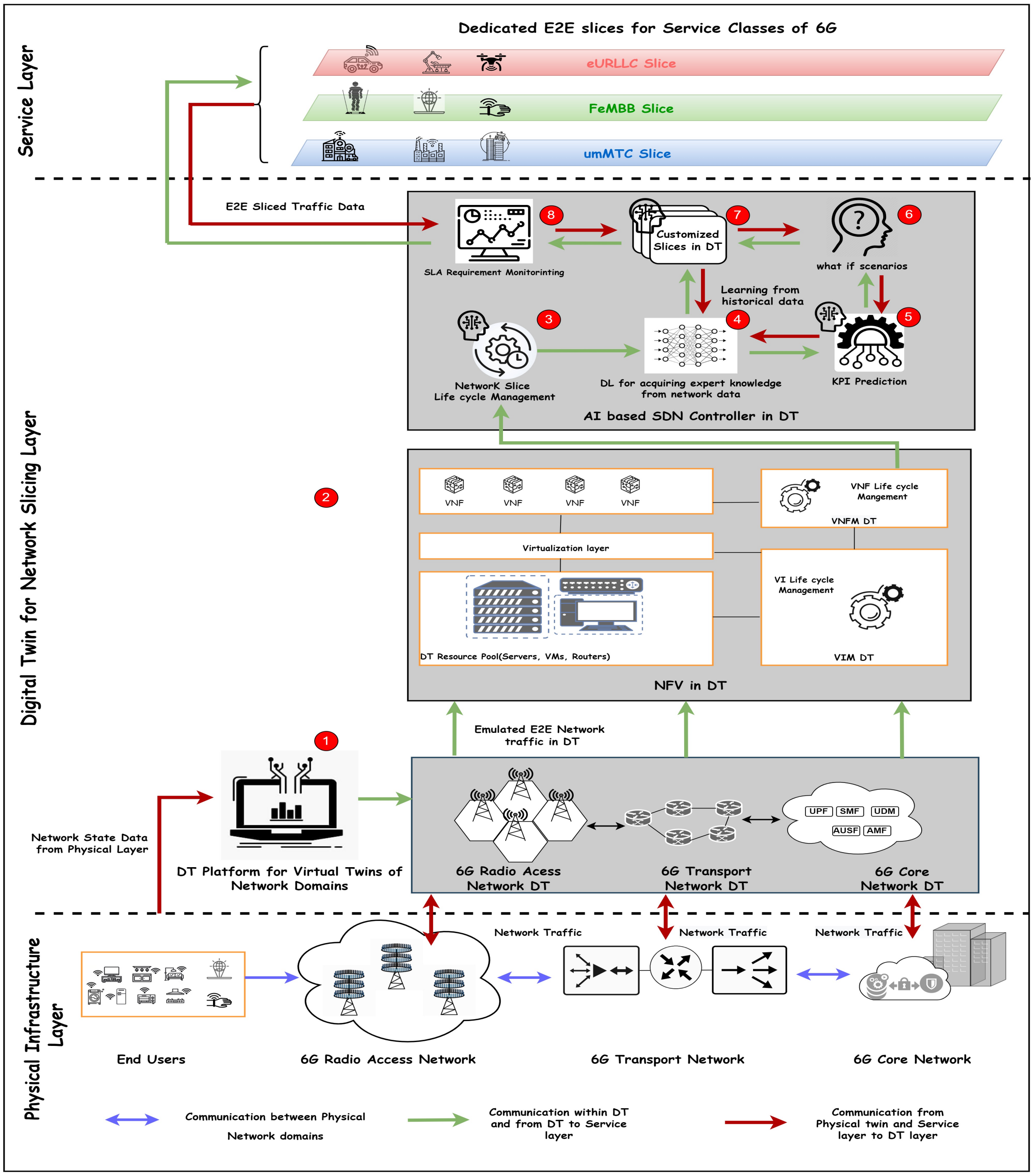

Fig. 2. Proposed DT framework enabling E2E network slicing using AI in 6G networks.

SLA Requirement Monitoring: Once the slices are customized it is vital to check SLA requirements. DT will make sure that each slice is complying with its SLA constraints, in case of violations alternative routes are chosen. DT has to continuously communicate with physical layer as well as service layer in order to become Self Organizing Network (SON), therefore the network traffic data is acquired from service layer as a feedback for self learning and optimization.

In Fig. 2, the DT layer will learn from 'what-if' scenarios and optimize performance measures like latency, jitter, bandwidth etc., enabling an intelligent networking paradigm. Even though the proposed architecture uses higher level of abstraction for network slicing model by making use of real time traffic to build the DT would be cost and time efficient.

C. Service Layer

In this layer SLA requirements are checked for deployed E2E slices by interacting with DT layer. According to the scope of network resources needed by NS in 6G, we define three types of Slices:

- *further enhanced mobile broadband (FeMBB) slice*; covers multimedia-rich applications like AR, VR, holographic meetings, etc. This type of slice would require a peak data rate of 100 times that of 5G which is at least $1Tb/s$.
- *ultra-massive machine type communication (umMTC) slice*; covers applications with high connection density, such as Industry 5.0-based scenarios, and large-scale IoE. This type of slice will require ten times the connectivity

density of 5G which can reach up to $10^7 devices/km^2$.

- *enhanced ultra-reliable and low latency communication (eURLLC) slice*; covers mission-critical applications such as autonomous intelligent driving, remote surgery, etc. that require high reliability, low latency and high availability. This type of slice will require latency of $10 - 100\mu s$.

By deploying slices on demand that comply with SLA requirement we can significantly reduce the OpEx and CapEx for service providers as well as shorten the deployment time for on demand network slices from weeks to hours. The (near) real-time capability of DT approach would facilitate the achievement of optimal trade-off among those extreme class requirements of different slices in future 6G networks.

IV. Use Case Scenario for network KPI prediction

As proof-of-concept, we validate the effectiveness of proposed DT model in terms of accurate prediction of E2E latency of network traffic generated in a SDN scenario, which is vital for guaranteeing QoS for end users.

A. Simulation setup

Due to infancy stage of 6G there have not been standard prototypes developed yet. Therefore to evaluate the proposed DT based NS architecture we leverage a packet level simulator OMNET++ for generating network traffic that represents the traffic of future generation 6G networks. The dataset generated by simulation setup contains results of delay, jitter, and packet loss for dynamically changing real world network topologies. The dataset is further divided into training and evaluation sets where training set contains relatively smaller networks with number of nodes ranging from 20 to 50, while validation set includes more complex topologies and relatively larger networks with nodes ranging from 50 to 300. Power-Law Out-Degree algorithm was used to generate these varying topologies artificially to represent traffic of 6G network, and the parameters have been extrapolated from repository of Internet Topology Zoo which is a real-world topology repository. The considered network scenario in dataset has the form of undirected graph containing nodes, links adjacency matrix, routing configuration port status and queue statistics. Each node is mapped as commodity servers, routers and VNFs with links joining each node and queuing policies are also configured for each node in SDN controller. Each node has resources like bandwidth, capacity, memory and storage. Nodes can send packets to any other node while each node has a routing table calculated based on Shortest Path First (SPF) algorithm. For each topology 100 different routing schemes have been defined eliminating any dependency on routing algorithm. The training set has 120000 samples where each sample has input files of network topology, routing scheme, traffic matrix and simulation results and one sample have results of 25 simulation runs. Similarly, validation set includes 3120 samples of similar configurations but with larger topologies. We then develop a GNN based solution in DT to train the above explained network scenario mapped as an undirected graph to make network metric prediction (latency). Accurate prediction of network analytics is the very first step towards building a complete NS DT for 6G networks. We implement AI based NS DT layer in python where Tensorflow, NetworkX and Keras libraries are used to model GNN based solution for acquiring expert knowledge and prediction of KPI.

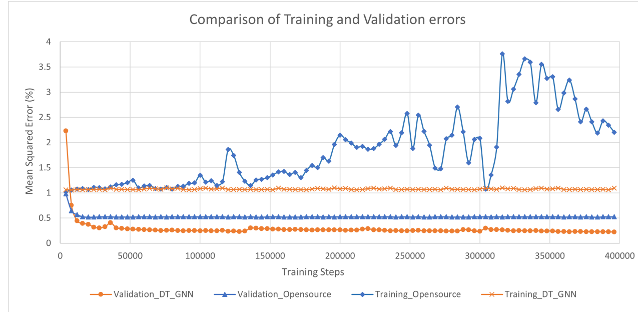

Fig. 3. Evaluating the generalization capability of the DT for predicting path delays on different network topologies.

B. Performance Evaluation

The GNN in DT will take traffic matrix as input to learn the features of network traffic in order to predict the overall path delays (latency) in scenarios not seen during the training process. We use mini batches of size 100 and the training with one of such batches forms a step. The DT model is trained to minimize the Mean Squared Error loss function between true latency and predicted latency. Adam optimizer is used as gradient-based optimization function while RELU is used as activation function. Figure 3 showcase the potential of DT model to learn from the historical data and make generalization in case of scenarios that were not involved during the training process. In order to evaluate the performance of our proposed solution we compared it with traditional open-source model which is design to learn the routes of the networks. Results prove that our DT model performs better for both training and validation set and converges much faster than the traditional offline approach. In Fig. 3 we have plotted training loss across training steps. The difference in accuracy is because of the capability of DT model to accurately map the relationship between network entities acquired from each layer in DT model as well as its ability to extrapolate the useful information from network data. Then results show that GNN DT model is clearly able to generalize to the unseen network scenarios in comparison with traditional data-driven model. In Fig. 4 we show the regression plot of results on 100 randomly selected latency pairs. The dots represent the predicted latency across true latency values while yellow lines represent 95% confidence level. The regression plot demonstrates that the overall prediction error is considerably low, but the errors of high latency values is more noticeable than the low ones. This problem can be solved by using a training set that is inclusive of wide range of latency distributions.

V. Challenges and Open Issues

In the following, we identify and envisage pressing issues with regards to network slicing for 6G and leveraging DT for ensuring E2E service provisioning.

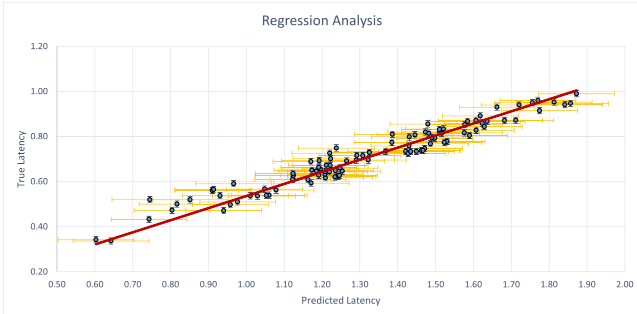

Fig. 4. Regression plot of true and predicted latency values with confidence level.

Diverse Traffic and Slice Isolation: Diverse traffic and immense number of connected devices in 6G, gives rise to the challenge of translating QoS requirements into network demands. Network Slicing DT has to provide dynamic and efficient slicing policies while keeping SLA violations to minimum. Moreover, it is vital to ensure efficient management of spectrum sharing and interference management schemes to achieve optimal utilization of available resources.

Real Time Traffic Monitoring and Data Collection: In context of networking processing and collection of real time data is challenging and expensive. In order to model network DT that is production-ready it must be capable of monitoring network health in real time, must interact with various twin objects, obtain logs from real twin and must evolve by learning from network behaviour to become SON.

VI. CONCLUSION

This article has introduced a concept of network slicing DT for 6G and its reference architecture. We have argued that a DT that can accurately predict performance and understand the relation between network configuration, is essential for optimizing performance for 6G service classes. We also unveiled potential of DT along with AI to retain the information between network metrics and to predict network KPIs. We believe that by leveraging real time data and existing technologies like SDN and NFV along with DT, we can achieve dynamic and fully automated network orchestration, optimization, and management that will promote the advancements towards autonomous self-evolving 6G environments.

REFERENCES

[1] P. Yang et al, "6G Wireless Communications: Vision and Potential Techniques," in *IEEE Network*, vol. 33, no. 4, pp. 70-75, July 2019.

[2] S. Wijethilaka and M. Liyanage, "Survey on Network Slicing for Internet of Things Realization in 5G Networks," in *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 957-994, 2021.

[3] Z. Zhang et al, "6G wireless networks: Vision, requirements, architecture, and key technologies," in *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28-41, Sept. 2019.

[4] H. Yang, et al, "Artificial-Intelligence-Enabled Intelligent 6G Networks," in *IEEE Network*, vol. 34, no. 6, pp. 272-280, Nov. 2020.

[5] ETSI Zero Touch Network & Service Management (ZSM); www.etsi.org/committee/zsm, accessed May 2021.

[6] ETSI, "Experiential Networked Intelligence (ENI)"; www.etsi.org/committee-activity/eni, accessed Jan. 19, 2021.

[7] H. Chergui et al., "Zero-Touch AI-Driven Distributed Management for Energy-Efficient 6G Massive Network Slicing," in *IEEE Network*, vol. 35, no. 6, pp. 43-49, Dec. 2021.

[8] J. Mei et al, "An intelligent self-sustained RAN slicing framework for diverse service provisioning in 5G-beyond and 6G networks," in *Intelligent and Converged Networks*, vol. 1, no. 3, pp. 281-294, Dec. 2020.

[9] W. Guan et al, "Customized Slicing for 6G: Enforcing Artificial Intelligence on Resource Management," in *IEEE Network*, vol. 35, no. 5, pp. 264-271, Oct. 2021.

[10] H. Wang et al, "A Graph Neural Network-Based Digital Twin for Network Slicing Management," in *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1367-1376, Feb. 2022.

[11] S. Mihai et al., "Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects," in *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2255-2291, Sept 2022.

[12] H. X. Nguyen et al, "Digital Twin for 5G and Beyond," in *IEEE Communications Magazine*, vol. 59, no. 2, pp. 10-15, Feb. 2021.

[13] P. Almasan et al., "Network Digital Twin: Context, Enabling Technologies, and Opportunities," in *IEEE Communications Magazine*, vol. 60, no. 11, pp. 22-27, November 2022.

[14] H. Chergui, et al. "Toward Zero-Touch Management and Orchestration of Massive Deployment of Network Slices in 6G." in *IEEE Wireless Communications*, vol. 29, no. 1, pp. 86-93, February 2022.

[15] M. Bagaa, et al. "Toward Enabling Network Slice Mobility to Support 6G System." in *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10130-10144, December 2022.

M. Yaqoob is a PhD Researcher with Middlesex Univ., London, UK. She received her BS degree in Software Engineering from Fatima Jinnah Women University Pakistan in 2016 and MS degree in Computer Engineering from METU-NCC in 2020. Her current research interests include DT, ML/AI, wireless communications, etc.

R. Trestian is an Associate Professor with Middlesex Univ., London, UK. She received her Ph.D. degree from Dublin City Univ., Ireland in 2012. Her research interests include mobile and wireless communications, ML, QoE, DTs. She is an Associate Editor of the IEEE Commun. Surveys & Tutorials.

M. Tatipamula is the CTO of Ericsson Group Function Technologies & Architectures, focusing on 5G, distributed cloud/edge compute, SDN/NFV, AI/ML and IoT. Mallik received PhD, Master's and Bachelor's degrees from University of Tokyo, IIT (Chennai) and NIT (Warangal) India, correspondingly. Mallik is a Fellow of Canadian Academy of Engineering.

H. X. Nguyen is a Professor of Digital Communication Engineering at Middlesex Univ., London, UK, Director of the London Digital Twin Research Centre and Head of the 5G & IoT Research Group. He is also a Visiting Professor at International School, Vietnam National University, Hanoi, Vietnam. He received his Ph.D. degree from the Uni. of New South Wales, Australia, in 2007. He leads research activities in digital twinning, 5G/6G/IoT systems, and digital transformation solutions.