Towards Supporting Holographic Services over Deterministic 6G Integrated Terrestrial & Non-Terrestrial Networks

Hao Yu
Oulu University
Oulu, Finland
hao.yu@oulu.fi

Tarik Taleb
Oulu University & Sejong University
Oulu, Finland; Seoul, South Korea
tarik.taleb@oulu.fi

Konstantinos Samdanis
Lenovo
Munich, Germany
ksamdanis@lenovo.com

JaeSeung Song
Sejong University
Seoul, South Korea
jssong@sejong.ac.kr

***Abstract*—Driven by the emerging mission-critical applications, the need for the deterministic networking (DetNet) capabilities of the current network infrastructure is becoming increasingly important, and that is to enable assured bandwidth, latency/jitter and reliability for these services. Recently, the technological advancements of non-terrestrial networks (NTNs) present great opportunities to provide such deterministic service provisioning for time-sensitive traffic, especially when integrated into terrestrial networks (TNs), which can be referred to as 6G-integrated terrestrial and non-terrestrial networks (6G-ITNTN). This article introduces the envisioned DetNet-enabled NTN architecture and explains the interplay of the bandwidth potential, computational power, and latency features of different NTN architecture options. Then, after analyzing the network performance requirements in terms of bandwidth, latency, and synchronization of remote holographic applications, we shed light on the NTN-based deterministic communication scenarios to support holographic services. Third, we discuss the deterministic network selection and routing (DNSR) scheme and propose a deep reinforcement learning (DRL)-based DNSR approach as a further step to control the end-to-end (E2E) delays for joint holographic flows in a deterministic way. Finally, we present a simulation to evaluate the performance of the proposed method of supporting critical holographic service flows over NTNs.**

***Index Terms*—Holographic service, Deterministic networking, 6G-integrated terrestrial and non-terrestrial networks, NTN, Artificial intelligence/Machine learning, and Metaverse.**

I. Introduction

6G is expected to introduce new communication experiences that blur the lines between real and virtual worlds. With advanced embedded monitoring and data collection technologies, Internet of Things (IoT) devices, such as holographic cameras, object detectors, and motion sensors, play vital roles in materializing virtual reality and augmented reality. For the mission-critical applications, e.g., holographic-type communication (HTC), delayed packets will be equally bad or even worse than those that are dropped. On the other hand, early packet delivery is also problematic. Today's low latency and high-bandwidth 5G enterprise services cannot support these futuristic application characteristics, which rely on deterministic bandwidth, bounded latency/jitter and high reliability. New technologies and network paradigms for deterministic Internet of Things (DIoT) applications are now being developed by different standard development organizations (SDO) to meet the deterministic service requirements.

Although significant achievements have been realized with 5G communications system in terms of higher date rate and lower latency, terrestrial communication technologies cannot provide all users with fair and deterministic quality of services (QoS). Not all users can access to high-quality network services at any time due to the scarce network capacities and limited coverage areas. For example, several time-critical applications are known for suffering from the lack of network broadband with stable latency guarantee in remote or rural areas. Therefore, non-terrestrial networks (NTNs) including satellites, unmanned aerial vehicles (UAVs), and high altitude platforms (HAPs, i.e., balloons), have gained attention to assure connectivity and QoS for 6G demands. As shown in Fig.1, two network segments comprising air and space networks at different altitudes are needed to complement the terrestrial networks so that IoT users can enjoy enhanced and flexible end-to-end (E2E) network connections. Combined with conventional terrestrial networks, this hierarchical systems are expected to comprise the 6G integrated terrestrial and non-terrestrial networks (6G-ITNTN).

Communication streams often need to wait until the appropriate network conditions are available. NTNs can enable traffic streams generated from remote ground nodes to be transmitted via more paths with less network congestion. However, a significant challenge is how to guarantee deterministic performance across such connections [1], especially when deterministic connectivity is a non-trivial task even for terrestrial networks. The Third Generation Partnership Project (3GPP) release 16 [2] includes 5G standard capabilities for ultra-reliable low-latency communication (URLLC) at both radio access networks (RANs) and 5G core [3]. However, transport network connectivity that remains crucial to assure E2E service level agreements (SLAs) is out of the 3GPP scope [4]. How to satisfy the bounded latency and jitter requirements of DIoT applications within NTN remains to be solved.

Compared with conventional terrestrial networks, the 6G-ITNTN is more complex to manage; thus intelligence is

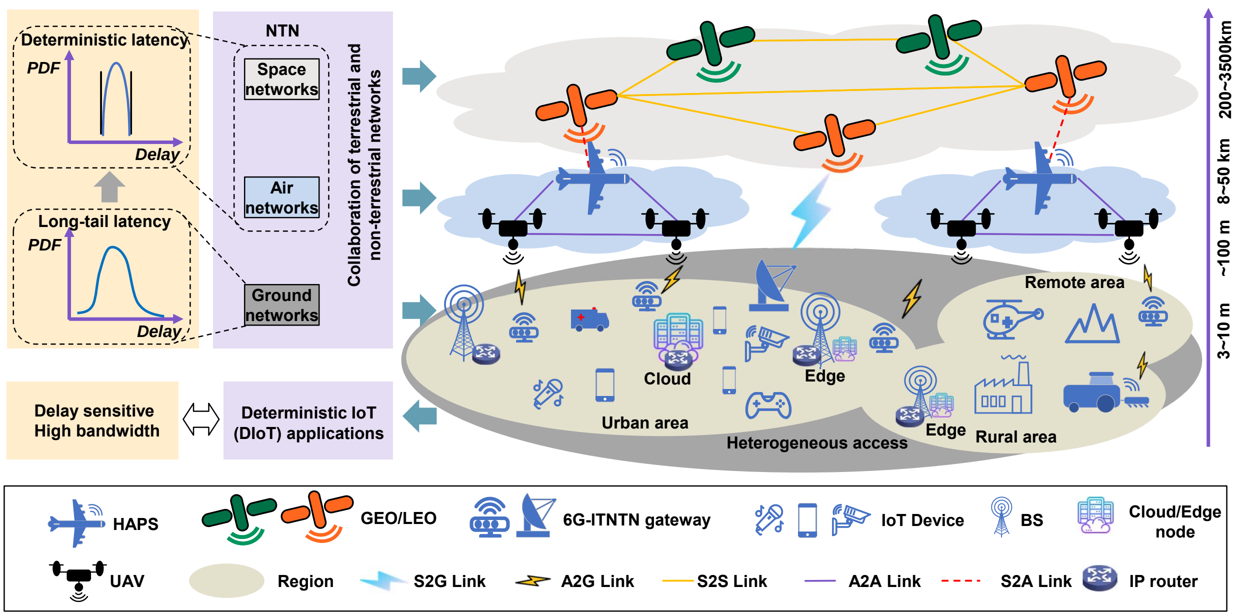

Fig. 1: 6G-ITNTN architecture for deterministic IoT services.

needed to improve its performance and efficiency. Artificial Intelligence (AI) / Machine Learning (ML) can be explored to enhance network performance [5]. Since high computational complexity of solver- and heuristic-based methods are usually enabled using human-handcrafted measures, highly scalable and intelligent scheduling approaches are promising solutions for scheduling the flows in a deterministic way.

This ambiguous landscape regarding the 6G-ITNTN, DIoT, and AI/ML has motivated the research work conducted in this paper. First, we provide a brief introduction to the perceived 6G-ITNTN architecture. Second, we take the holographic communication service as a DIoT example and underline the unique challenges that arise when implementing holographic communications with deterministic performance. Third, we investigate the deterministic network selection and routing schemes for joint holographic flow scheduling and apply a case study based on a deep reinforcement learning approach for service provisioning. Then, we measure the overall performance for evaluation. Finally, we conclude this paper.

II. 6G Non-Terrestrial Networks (NTN)

As shown in Fig. 1, 6G-ITNTN consists of heterogeneous communication infrastructures and multiple network segments in space, air, and ground for cost-efficient, flexible, and large-scale mission-critical DIoT applications. Owing to the enormous amounts of data and network traffic that are produced, exchanged and managed, conventional ground communications may need enhancements to meet the need of deterministic services. The presence of NTN platforms, such as satellites, airships, and UAVs [6], can assure exclusive and non-blocking network connectivity.

A. Spaceborne Networks

Spaceborne networks[1] are traditionally based on Geostationary Earth Orbit (GEO) satellites, which have a circular and equatorial orbits. They appear stationary from a ground observer's perspective and cover a large portion of the earth's surface. The typical beam footprint size varies from 200 km to 3500 km. Although, GEO satellites are widely available and can provide communication capacity over vast areas, their latency of 500 ms round-trip time (RTT) is not acceptable for time-critical IoT applications due to high orbit altitude. On the other hand, Low Earth Orbit (LEO) satellite constellations allow a lower latency and can be used for certain cases for voice and video transmissions (less than 100 ms RTT).

B. Airborne Networks

NTN includes High-Altitude Platform Stations (HAPS), e.g., airships and balloons. Notably, Unmanned Aerial Vehicles (UAVs) can also act as aerial forwarding nodes at altitudes of 8 to 50 km in support of HAPS and few hundred meters for UAVs above ground level [7]. Aerial platforms have great potential to establish Light of Sight (LoS) communication with the ground users and thus enhance the coverage and connectivity. Both HAPS and UAVs can be rapidly deployed in quasi-stationary condition and move on-demand according to the user distribution. Recently, Google initiated Project Loon, which aims to leverage high-altitude balloons as a means of provisioning broadband services to remote locations [8].

[1]The data in this section is derived from G. T. 38.821, "Solutions for nr to support non-terrestrial networks (ntn)," 2019.

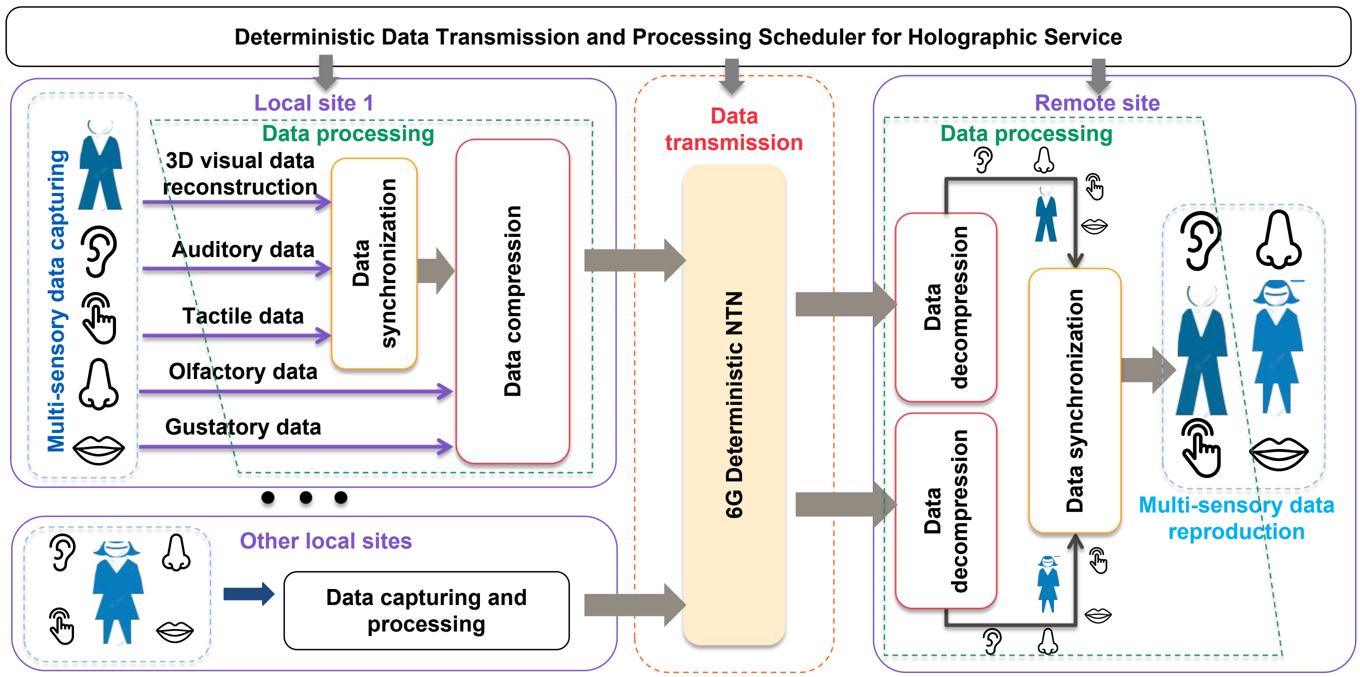

Fig. 2: Deterministic networking enabled Holographic-Type Communication (HTC) system.

III. Deterministic Networking enabled HTC System

To show the potential of deterministic 6G-ITNTN in terms of supporting the emerging time-critical services, we take the holographic-type communication (HTC) applications as example, by giving a brief introduction of the HTC system in this section and discuss how deterministic 6G-ITNTN will facilitate the time-critical service provisioning in the following sections.

In order to design an HTC system with deterministic performance in terms of bandwidth, latency, etc., multi-sensory data transmission and processing at both local and remote sites must be coordinated for image/video capturing and reproduction. As shown in Fig. 2, a block diagram of a fundamental HTC system is depicted. Holographic users at local sites need to capture, process, and communicate the multi-sensory data, including 3D visual data, auditory data, tactile data, olfactory data and gustatory data, to the remote interlocutor. After being captured, the acquired data from various sensors are then subjected to a number of processing steps, including 3D visual data reconstruction, multi-source stream synchronization, and data compression. If various visual data sources are presented and the generated data need to be transmitted to one remote client, they must be synchronized. On the one hand, the user tracking function will capture dynamically the user's motions and gestures. Multi-view data alignment, object detection and extraction, etc., are supposed to be conducted with the 3D data reconstruction block at remote client side. On the other side, the synchronization between the different local sites' streams should be completed by synchronization block to reproduce the valid and harmonized video/image. A data compression function performing the data compression should also be used to reduce the enormous network bandwidth consumption before the heterogeneous data are emitted to the remote client side through the network segment. Unlike indoor holographic applications, remote holographic services pose huge challenges to the transport networks in terms of QoS guarantee. To ensure the holographic flows to be transmitted to the destination on time and the multiple holographic streams from different locations to be synchronized, a holographic service and network scheduler should be designed to coordinate the data transmission and processing for deterministic end-to-end performance. It will perform the network routing planning for time-sensitive holographic flows and reserve appropriate network and computation resources to accommodate the holographic traffic while satisfying their strict end-to-end QoS requirements.

In the next section, we will analyze the performance requirements of holographic services in terms of the bandwidth, latency and stream synchronization. Then we will discuss several NTN-based deterministic communication use cases for holographic services and propose a deep reinforcement learning based solution to solve the deterministic network selection and routing problem.

IV. Requirements of Holographic Application on the Networks

In this section, we will study the holographic-type communication (HTC) [9] by analyzing its network requirements in terms of data transmission, data processing, latency and stream synchronization.

By rendering information objects in three dimensions (3D) with more accurate light intensity, form, and texture than Augmented Reality/Virtual Reality (AR/VR), holograms produce visualizations in the fields of education, information, and industries. For example, holographic telepresence allows remote participants to be projected as holograms alongside local participants with real-time interactions. As a result, holographic avatars are multidimensional information duplicates of a real scene and its items. Presenting holographic data, particularly for transmission of moving objects, is a challenge that is stretching the limits of available technology.

The image capturing usually relates to a set of objects collected at different angles and tilts, a vast volume of data needs to be transmitted across the network. And the interactivity between users and holograms during immersive HTC requires ultra-low latency for real-time interactions. Furthermore, optimized streaming schemes for the coordination and synchronization of multiple concurrent streams at very high precision levels is also required to support truly immersive experiences. To this end, HTC system is supposed to be well-designed to support holographic transmissions and interactions between different remote locations across the network. Because

A. Data transmission requirement

The effectiveness of the pixels for a given unit surface area determines the quality or resolution of a hologram. With a rise in display sizes and frame rates, the size of continuously moving 3D hologram pictures will also expand. For instance, nearly 8.8 MB of raw data per frame, or 2.06 Gbps at 30 FPS can be the output of a camera sensor (e.g., Microsoft Kinect) to support a single 1080p image. With additional sensors of higher resolutions and frame rates, the required bandwidth will explode to between 100 Gbps to 1Tbps [9]. Using encoding and compression techniques, the bandwidth of the holograms to be transmitted can be significantly reduced. However, what should be transmitted or what should be compressed/dropped need to be determined depending on a variety of factors: compression algorithm, compression ratio, display size, and even the semantic information of holograms.

B. Data processing requirement

The main challenge of data processing in HTC systems is how to ensure the tradeoff between the lightweight computation and transmission of large amounts of data with low latency. The compulsory data processing procedures in HTC system usually include 3D data reconstruction and rendering, compression and decompression. Once the required data are loaded on the client device, non-parallelized loading of 30 frames of the video-based point-cloud compression takes, on average, 60 s, rendering frames at a frame rate of 0.5 FPS on a hexacore Intel Core i7-8850H CPU running at 2.60 GHz with 16 GB of RAM. So, loading and rendering a single item at the desired rate of 30 FPS would take a minimum of five CPUs assuming full parallelization [10]. With regard to this, leveraging Multi-access Edge Computing (MEC) resources is proposed in order to facilitate computing efforts by rendering the user's field of vision in the network rather than at the client side. As a result, huge transmission bandwidth demand will be induced. Although the bandwidth of volumetric media can be reduced by compression process, encoding and decoding will also bring latency increase significantly. Therefore, finding a tradeoff between the development of effective compression techniques, which would increase computing latency while reducing network bandwidth and latency, and vice versa, is crucial.

C. Latency requirement

Although 5G networks are expected to provide 1 ms round-trip latency, it remains impossible to achieve such low latency over a long distance. Most literature that evaluates experienced latency declares values of a couple of hundreds of milliseconds, which is too much for HTC. For example, fewer than 15 ms for motion-to-photon latency and fewer than 50–100 ms for end-to-end latency are acceptable according to [11]. To provide ultra-low latency, which is imperative for real immersive experiences, transport network architectures and protocols must be transformed to facilitate real-time communications. Many transport layer protocols promote reliable and fast data delivery, for example, Real-Time Media Protocol (RTMP) is the most widely adopted protocol among TCP-based solutions. Moreover, to eliminate additional latencies caused by continuous packet acknowledgements of TCP-based solutions, user datagram protocol (UDP) solutions may be better suited for low-latency services (e.g., Real Time Protocol). Despite efforts to strengthen the transport layer for real-time communications, the latency controls needed for extremely precise granularity at the packet or frame level is not supported. For instance, New Internet Protocol (NewIP) [12]–[14], which aims to revolutionize the legacy IP technology to support the packet forwarding with QoS guarantee by redesigning the network datagram format, specification and corresponding capabilities in data plane.

D. Stream synchronization requirement

Within one remote holographic application, multiple holographic streams originating from different sensors and locations must be synchronized and coordinated. Otherwise, misaligned streams of the same holographic application will result in badly rendered image display. Not just in-time provisioning (low latency) but on-time provisioning (deterministic latency) should be guaranteed to make the arrival of multiple streams aligned. Moreover, when users need to interact with their online content, stringent stream synchronization is imperative, which demands precise timing controls in underlying networks and sophisticated joint stream scheduling schemes in the control plane.

V. NTN-based Deterministic Communication Use Cases for Holographic Services

In this section, we examine a routing problem example of deterministic communications to support holographic services. As shown in Fig. 3, the inter-layer connections between air and

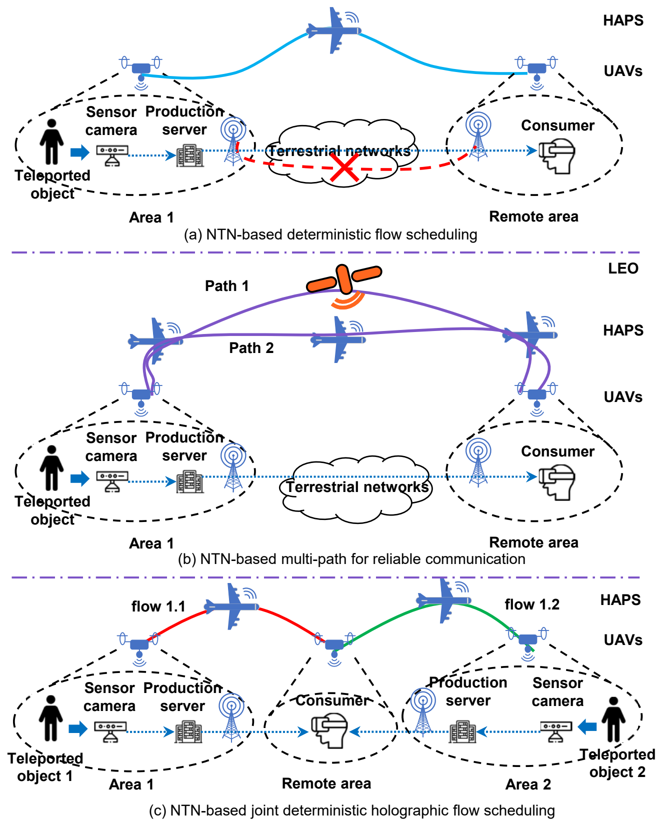

Fig. 3: NTN-based deterministic communication scenarios for holographic services.

space segments enable flow packets to be transmitted from IoT devices to destination holographic applications across multiple paths. Given that terrestrial networks suffer from non-uniform QoS, and an E2E path may comprise poor wireless connectivity, or transport network parts that do not support TSN/DetNet technology, the packets from Area 1 that are destined for remote applications can be transmitted via airborne or spaceborne networks. As shown in Use Case (a), the holographic service connection of BSs of Area 1 to the remote area travels through terrestrial networks in which the E2E service latency is difficult to predict and control. If the UAVs are assumed as the accessing nodes for holographic services, the stream may be transmitted directly via satellites with higher priority and fewer hops.

From the other perspective, UAVs can act as the relay node to deliver the holographic service as long as UAV-based networks can provide fast switching and end-to-end latency guarantee. However, UAVs' bandwidth and power capacities may not support massive traffic loads, especially for holographic flows with huge bandwidth demands. Thus, the hybrid HAPS and UAVs-based NTN is supposed to be a promising solution. UAVs perform as access points and gateways for a small or medium area to offload the data processing to the proper network elements. HAPS processes the data or perform as switching nodes to route the holographic traffic to the destination for a larger area.

When heavy traffic loads are transferred via the airborne networks, single path routing may lead to further network congestion. Assuming that the traffic demands of a certain area are easy to predict, and periodical bursts of traffic are generated from Area 1, the airborne network nodes that connect the IoT devices of Area 1 will be repeatedly congested. If the networks can learn the traffic patterns, then the airborne network nodes can transfer parts of their traffic loads to spaceborne network nodes for balancing. As shown in Use Case (b), multi-connectivity schemes can be also applied to the 6G-ITNTN system to enable high reliability communications and multiple-path routing for load balancing. Notably, deterministic IoT applications require ultra-high reliability to ensure the service continuity (i.e., no service disruptions and zero packet loss). Hence, multiple disjointed paths across multiple network segments can theoretically fulfill the reliability requirements. Furthermore, multiple-path routing can also achieve the load balancing needed to maximize the availability of network resources.

If a holographic service consists of multiple flows originating from different sources, as shown in Use Case (c), the system must empower the joint deterministic flow scheduling to ensure flow synchronization. That is, packets from different flows must arrive at the destination application at the same time. Compared with terrestrial networks, it is easier to decide the routing paths over NTNs with deterministic latency and jitter. We discuss deterministic network selection and joint holographic flow routing in the next section.

VI. Deterministic Network Selection and Routing (DNSR) for Holographic Services

Several emerging standards and technologies can be adopted to construct deterministic networks for DIoT. Whilst mature statistical multiplexed data networks often experience packet loss and significant variable latencies due to congestion, deterministic networking (DetNet) technologies are expected to deliver predictable network services over legacy infrastructures. In this section, we provide a brief introduction of emerging deterministic networking technologies. Then, we propose a communication scenario for joint holographic flow-scheduling and examine the interplay between network selection and routing to ensure deterministic QoS for holographic services in 6G-ITNTN.

A. Deterministic Technologies in Networking

Time-sensitive Networking (TSN): Ethernet has been widely adopted as a common mode of networking connectivity due to its simple connection mechanisms and protocols. The best-effort Ethernet service reduces the network complexity and keeps protocol operations simple while reducing the product costs of Ethernet units. Despite its enormous success and widespread adoption, the Ethernet fundamentally lacks E2E deterministic QoS properties for flows. The IEEE 802.1 TSN TG standards extend the traditional Ethernet datalink layer by a set of standardized TSN mechanisms and principles,

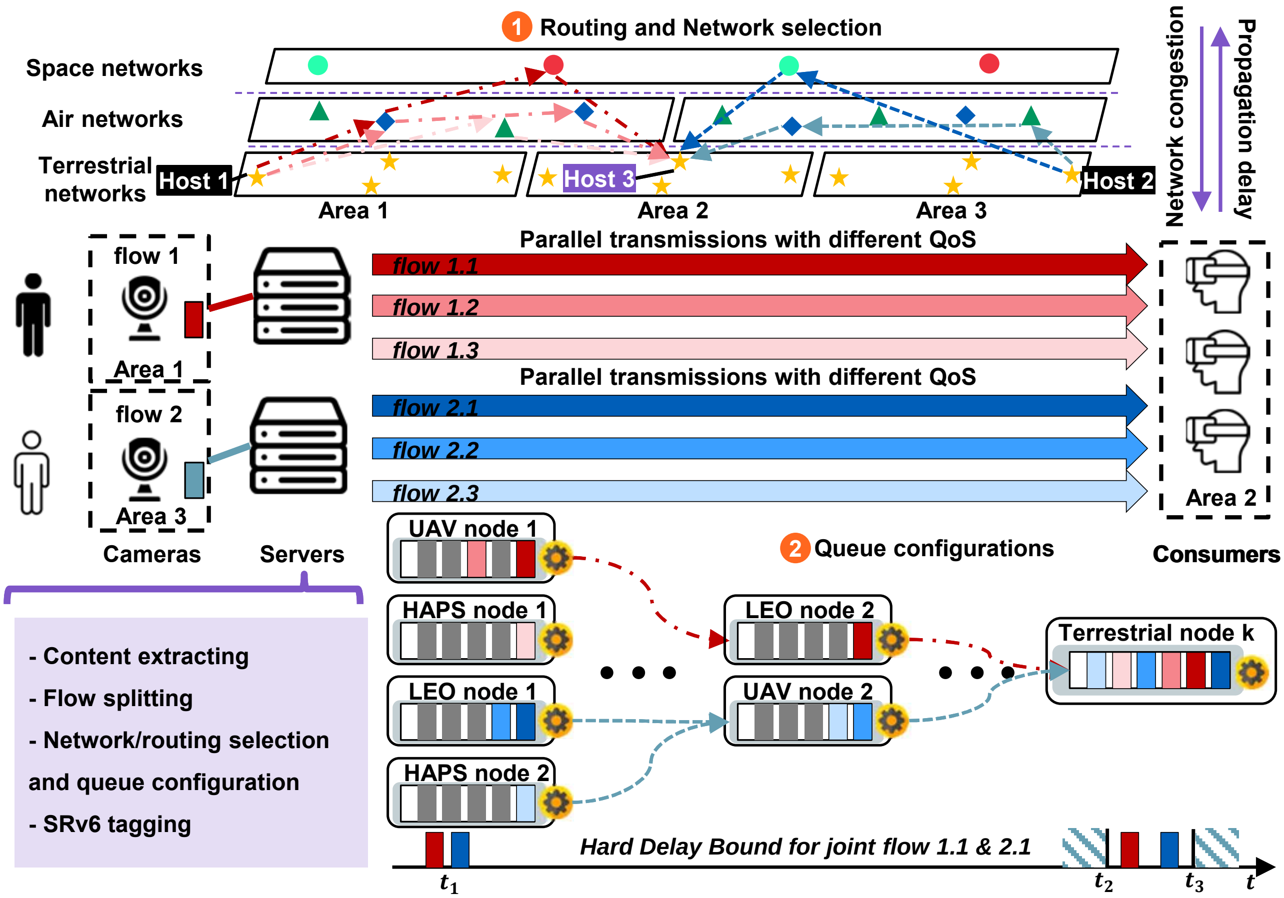

Fig. 4: Illustrative example of holographic-type communications for network selection and routing.

e.g., time-aware shaping, frame preemption, to guarantee time-sensitive flows with bounded ultra-low latencies, low delay variations (jitter), and extremely low loss, which are needed for industrial control and automotive applications.

Deterministic Networking (DetNet): Considering that the TSN TG mainly works on the Layer-2 networking, the Internet Engineering Task Force (IETF) DetNet Working Group focuses on deterministic data paths that operate over Layer-3 routed segments with bounded latency, jitter and high reliability and extends TSN technologies towards a larger network scale. Deterministic networking provides guaranteed latency on a per-deterministic-flow basis to ensure zero data-loss through congested networks. The deterministic forwarding method is accomplished by dedicating network resources (e.g., link bandwidths and buffer spaces) to DetNet flows via cycle-specified queuing and forwarding (CSQF) [15] mechanism, and by replicating packets along multiple disjointed paths. Unused resources are available to non-time-sensitive flows as long as all guarantees are fulfilled. Furthermore, explicit routing methods, e.g., segment routing, can be utilized to configure the flow scheduling and address the impact of the convergence of routing and bridging protocols[2].

[2]Referred to as SR-CSQF

Segment Routing IPv6 (SRv6): SRv6 relies on IP header to enable the packet steering of a particular flow using a set of instructions called segments. A segment represents a topological, local or service-based semantic item and can: (i) enforce a flow through a specified path, (ii) allow the selection of a resource, (iii) and enable service instruction (e.g., steer packets via a firewalls). SRv6 requires ingress and potentially selected intermediate node(s) to maintain the per-flow state needed for segment policy provision.

B. Deterministic Network Selection and Routing (DNSR) Scheme

In this section, we describe the network selection and routing scheme for deterministic joint holographic flow scheduling. This explains how multiple holographic streams originating from different sensors and locations can be synchronized and coordinated in the proposed network. Here, we assume that each NTN component can perform the DetNet functionalities as specified in [15], that is, configuring the route and cycle for packet transmission.

Because the object is captured from different angles at source, the consumer can select the viewing angle at the destination by changing the rendered shape, light intensity and texture to the object. This can be achieved by encapsulating the volumetric and other information through an attribute matrix, thus it allows the networks to selectively choose which part of the data should be transmitted with higher degrees of QoS and reliability, which part of the data may be dropped due to the congestion but still be able to reconstruct the information on the destination. As an example, it may be acceptable to drop some packets of the stationary background and ensure the packet delivery of moving objects when transmitting a hologram video.

As shown in Fig. 4, two remote holographic flows from different geographical locations (namely, Areas 1 and 3) are assumed to be transmitted to an identical holographic application host (i.e., in Area 2). As stated above, the hologram content extracting and encapsulating into sub-flows with different degrees of reliability are performed in the production servers at the source. For instance, data flows 1.1 and 2.1 represent the part of hologram content that should be transmitted with highest QoS, e.g., the position and action information of moving objects, data flows 1.2 and 2.2 denote the contents that are less essential to reconstruct the hologram at the destination, e.g., some fine-grained skin or hair texture. Data flows 1.3 and 2.3 can be the data of static background which will be unchangeable in a short period. Thus, data flows 1.1 and 2.1 are supposed to be transmitted simultaneously (t_1) and are expected to arrive at the destination node within the hard delay bound (t_2, t_3) and with zero packet drop, so that the key contents can be merged, decoded, and rendered at the application server for their holographic display. The other sub-flows can be transmitted with soft delay bound, i.e., with an acceptable probability of delay violation and packet drop. To accommodate these holographic flows, the UAV orbits are needed at fixed heights near the IoT devices to collect the latency-sensitive holographic service requests. Taking the network status of the UAVs, BSs, and HAPS into account, the ITNTN gateway can select the appropriate network segments by accessing different network components (i.e., nearby UAV, BS, or remote HAPS), e.g., to route the holographic data flows within the air network or forward the part of flows to space layer. Basically, transmitting through space networks will induce higher propagation delay, while the packet transmitting through air or terrestrial networks may be blocked due to network congestion.

The BS, UAV, and HAPS have different characteristics. For example, the BS has a high bandwidth capacity, whereas its coverage area is limited. On the other hand, UAVs have greater coverage but less bandwidth capacity. Certainly, the BS–UAV/HAPS link delay cannot be neglected. However, as UAVs can be deployed flexibly, they can act (when needed) as a complementary solution for the greenfield BS coverage. The HAPS can cover a larger area and act as an alternative solution for terrestrial networks with less forwarding latency due to less forwarding hops. However, considering the forwarding capacity of NTN components are limited, the NTN should be selected and prioritized for the network's time-critical flows. Therefore, holographic flows should be scheduled appropriately via different networks while accounting for the overall network environments comprehensively to ensure the deterministic latencies.

E2E service delays entail transmission, processing, propagation, and queuing delays. To simplify delay modeling, the overall networks can be modeled as a discrete cycle-based system with equal time duration, as specified in the CSQF [15]. If the bandwidth capacity of a network topology are known in advance, the transmission and protocol processing delay of a flow can be regarded as deterministic. Thus, queuing and propagation delay is not deterministic due to the resource competition and uncertain network status if the routing path is not decided. When there are not available network resources, that is, the data flows cannot be forwarded at once, they will be buffered in the forwarding queues of network components along the path, incurring additional delays. With the advancements of SRv6 and CSQF mechanisms, the data flows can be steered by planning the cycles for the forwarding in a deterministic way. By designating cycles for packet forwarding in advance, E2E service delays can be calculated according to the first cycle that the data flow enters the system and the last cycle that the data stream leaves the system. The corresponding forwarding cycle index can then be tagged in the SRv6 overhead of data streams. For the joint flow scheduling problem in which the packets originated from different sources and are expected to arrive at the destination at (almost) the same time (Fig. 4), we introduce the DRL-based network selection and routing scheme for joint holographic flow scheduling in the next section.

VII. DRL-based Deterministic Network Selection and Routing (DNSR) for Holographic Service

Assuming the complicated network environments and innumerable mission-critical application requirements, DRL technologies are key to operate on the flows collected from ground network segments and orchestrate the network resources and flow scheduling among the multiple layers in 6G-ITNTN to support deterministic service provisioning. The network selection and routing problem in 6G-ITNTN can be modeled as a Markov Decision Process (MDP) with a system state, action, and reward. The DRL agent, which can be hosted at some network elements with powerful computing capacity, e.g., a HAPS node, will collect the service requests, network topology, bandwidth and queue status through an available network controller to make the network selections and routing decision for each holographic sub-flows, $f \in \mathcal{F}_s$, within a holographic service, s. The system state, action, and reward are introduced next to describe the learning process.

State Space: The state, $\mathcal{S}$, reflects the network environment (i.e., network topology and queue utilization in each network node). The service requests are also included in the state and sent to the agent.

Action Space: The system action includes the network selection and routing strategy at each time step, $\mathcal{A}$. As in [16], we decompose the network selection and routing into multiple sub-actions to improve the scalability. $a_i = (e_i, c_i)$ is defined as the sub-action if the edge, e_i, (e.g., UAV–LEO

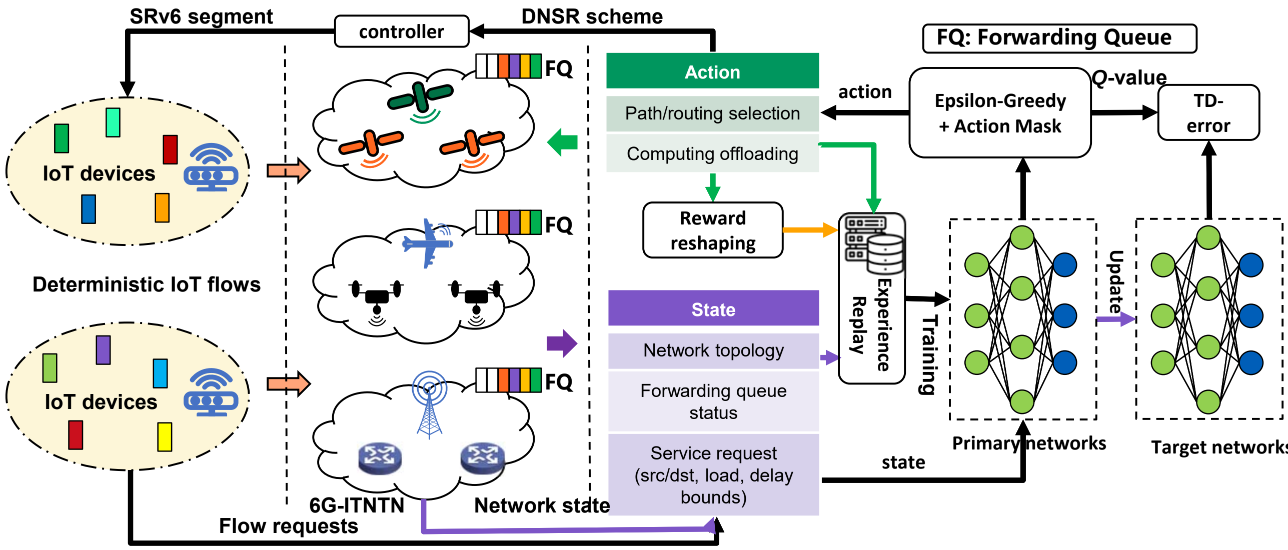

Fig. 5: Deep reinforcement learning-based network selection and routing for deterministic holographic services.

link) and cycle, c_i, (e.g., a forwarding cycle in an UAV node) are selected as parts of the path.

Reward Function: The agent continuously conducts sub-actions based on the current state until a valid path is formed (i.e., the edges are connected head-to-tail). Then, the agent obtains its final reward, r_f, based on the experienced latency of the flow (i.e., $c_N - c_1$). The latencies of flows in $\mathcal{F}_s$ result in lower jitter, and the reward is higher. The reward of each sub-action is obtained in a discount way (i.e., $r_i = r_i + i * r_f/N$), meaning that latter sub-actions are more essential than the former ones when forming a valid path for flow f. It is then stored in the replay buffer with state and sub-action.

Training Process: As shown in Fig. 5, a double deep Q-network (DDQN) based network selection and routing algorithm can be summarized as follows. The DDQN agent initializes the parameters θ of the primary neural network at first. Then, based on the current policy, θ, and state, s, sub-action a is generated with an epsilon-greedy strategy by the primary network. Action masking is used to filter the invalid actions in order to speed-up training convergence. When the agent manages to find a valid path for a flow f, the agent stores the corresponding obtained reward, r, and the next state into a replay buffer as a tuple. Third, the parameters of the target networks are updated from the primary networks after a period of training. The temporal-difference (TD) error between primary and target networks is used to update the primary neural networks with the gradient descent method.

VIII. Performance Analysis and Discussion

We evaluate the performance of the proposed DDQN-based network selection and routing scheme through simulation. We consider a simplified network structure as the considered scenario[3]. In the space segment, there are two GEOs, four

[3]Not depicted in this paper due to lack of space

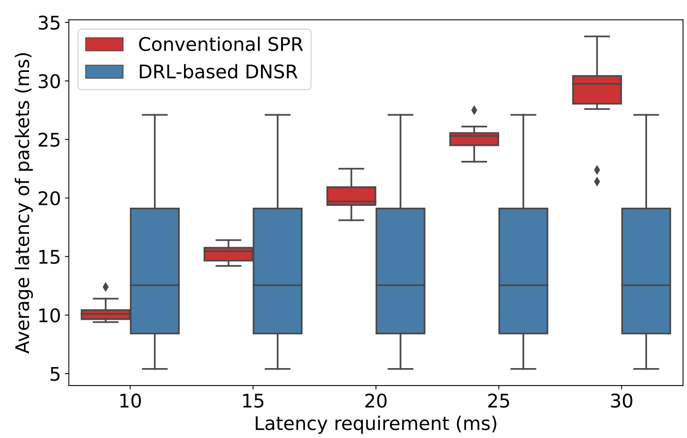

Fig. 6: Comparison of latency and jitter of sub-flows with hard delay bound between conventional shortest-path routing (SPR) and DRL-based DNSR.

LEOs, while the airborne segment is composed of eight HAPS and sixteen UAVs. The UAVs are distributed evenly, and each UAV acts as the access point to NTN platform for a certain area of DIoT users. If the user devices need to transmit the packets via the space segment, they will upload the packets to the UAVs. Therefore, the inter-layer links in the space layer as well as the air-space link are considered. The distances between different network elements are set as described in Section II, and the propagation latency is set as five μs/km. The link's available bandwidth is set between 500 Mbps and 2 Gbps; the time slot is 50 μs.

We randomly generate the holographic service requests. For each service request, two flows originated from different areas should be routed to the same application host. In the source of each flow, we assume three Kinect v2 sensors are capturing a dynamic scene from three directions at 30 fps.

Each sensor provides point-cloud data with 217,088 points per frame, which gives a total of 651,264 points per frame for three sensors. For each single point, geometry characteristics are represented by 32-bit X, Y, and Z values, and color attributes are described with 8-bit R, G, and B values. The calculation for the total amount of data at 30 fps is 651,264 × (3 × 32 + 3 × 8) × 30 = 2.344 Gbps. Then traditional video coding techniques are applied to compress the video stream. We assume they can offer lossy compression ratio of 1:200. The packet length follows the Ethernet standard 1,500 Bytes. Fewer than 50 ms end-to-end latency, according to [11], are defined as the latency requirement. We compare the proposed scheme to the conventional shortest-path routing (SPR) scheme.

As it is important to synchronize the multiple holographic images from different transmission paths, as shown in Fig. 6, after the training of the DRL agents with the generated service requests, the DRL-based DNSR scheme can achieve a deterministic performance with much lower jitters for the sub-flows with hard delay bound than conventional SPR schemes. This is due to the fact that the conventional SPR scheme tries to search for the shortest path for each flow, regardless of the jitter of joint flows within one holographic service. On the one hand, the RL-based routing and network method will coordinate the scheduling of flows within one single service with the objective of minimizing the jitter. On the other hand, the NTN platform also advances the deterministic routing scheme with fewer intermediate forwarding nodes. The flows with deterministic latency and lower jitter can thus provide a reliable guarantee for the execution of holographic services.

Different from the terrestrial network, the device mobility and the number of non-terrestrial devices will have an effect on the Quality of Experiences (QoE) and should not be neglected. Therefore, it is important to discuss the mobility impact of each type of NTN nodes. For the Satellite Network, Geostationary Earth Orbit satellites will keep stationary relatively from the ground station and have a larger coverage. LEO satellites will move very fast from a ground reference point. To cover the Earth, the LEO satellites will be organized in form of constellation. The mobility of LEO satellites raise many challenges, including long distance radio communication, accurate laser tracking technologies, as well as networking and routing. In practice, a LEO satellite constellation consists of certain number of orbits and each orbit has its own orbit elements, such as altitude and inclination angle. Each orbit will deploy a certain number of satellites and all the satellites in the same orbit will also evenly distributed in the ranges of 360°. Due to the fast moving of LEO satellites, the communication between LEO satellites and terrestrial devices will not be steady over the time. To understand their characteristics, we next analyze the satellites' motion and its impact on the network in terms of topology, link lifetime, link metrics. Typically most satellite is moving within its orbit with a speed faster than 7 km/s. The moving speed of terrestrial station is at 463m/s with the Earth self-rotation, which will result in: the links between LEO satellites and terrestrial devices keep flipping every couple of minutes ($\sim$5 minutes for LEO satellites at 550km of altitudes.) Despite fast moving of LEO satellites, the position and track of a certain satellite can be calculated and predicted over time. In addition, Inter-Satellites Link (ISL) allows the communication between LEO satellites to forward the packets in the space. Thus, if the orbit parameters (Altitude and Inclination angle) of LEO satellites can be shared within the whole LEO satellite network through novel network protocol (such as new IP) or obtained by centralized network controller, the corresponding deterministic routing path can be derived according to these information or updated if the parameters change to ensure the seamless quality of experience. Note that, we only analyze the impact of device mobility on routing and the corresponding solution from the perspective of networking. The signal attenuation due to device mobility is out of the scope of this paper. As for HAPS such as ballons and airship, they are supposed to be equipped with enough power to keep motionless in the air. For UAV (swarm), we also assume that UAVs only act as the access points for holographic services. If the ground nodes need to transmit the packets via the space segment, they will first upload the packets to the UAVs.

IX. Conclusion

This paper sheds light on the 6G-ITNTN concept with deterministic abilities and justify why 6G-ITNTN will bring the DIoT vision closer to reality by providing network connectivity with deterministic latency characteristic to mission-critical applications in large and remote areas. First, the motivation of introducing deterministic network capabilities into 6G integrated terrestrial & non-terrestrial networks (6G-ITNTN) is clarified and a brief introduction of 6G-ITNTN architecture is presented. Then, we showcase the different architecture options of merging NTN with mobile network components and the bandwidth, compute and latency requirements of each options are discussed qualitatively. Next, we take a remote holographic application as an example and introduce its requirements on transport networks. As a step further, we provide our vision on several NTN-based deterministic communication scenarios for holographic services and propose a DRL-based deterministic network selection and routing scheme with the CSQF mechanism. Finally, the performance evaluation in a simulated 6G-ITNTN environments shows that the proposed DNSR scheme outperforms the conventional shortest path routing, which provides bounded service latencies for time-critical flows.

In this paper, we only consider latency factor in the reward function, because the focus of the proposed DRL solution is learning how to schedule the sub-flows of holographic service jointly with the objective of minimizing jitter between sub-flows. But the other characteristics, e.g., bandwidth, of holographic service as well as distributed intelligent flow control algorithm to make the network device make fast routing and packet scheduling decision by themselves, which can benefit for the time-sensitive holographic services, will be investigated in our future work.

ACKNOWLEDGMENT

This research work is partially supported by the European Unions Horizon 2020 Research and Innovation Program through the Charity project under Grant No. 101016509; the Academy of Finland 6Genesis project under Grant No. 318927 and the Academy of Finland IDEA-MILL project under Grant No. 352428.

REFERENCES

[1] H. Yu, T. Taleb, and J. Zhang, "Deterministic latency/jitter-aware service function chaining over beyond 5g edge fabric," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2148–2162, 2022.

[2] S. Parkvall, Y. Blankenship, R. Blasco, E. Dahlman, G. Fodor, S. Grant, E. Stare, and M. Stattin, "5g nr release 16: Start of the 5g evolution," *IEEE Communications Standards Magazine*, vol. 4, no. 4, pp. 56–63, 2020.

[3] X. Lin, S. Rommer, S. Euler, E. A. Yavuz, and R. S. Karlsson, "5g from space: An overview of 3gpp non-terrestrial networks," *IEEE Communications Standards Magazine*, 2021.

[4] J. Prados-Garzon and T. Taleb, "Asynchronous time-sensitive networking for 5g backhauling," *IEEE Network*, vol. 35, no. 2, pp. 144–151, 2021.

[5] N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE wireless communications*, vol. 24, no. 3, pp. 146–153, 2016.

[6] E. T. Michailidis, S. M. Potirakis, and A. G. Kanatas, "Ai-inspired non-terrestrial networks for iiot: Review on enabling technologies and applications," *IoT*, vol. 1, no. 1, p. 3, 2020.

[7] X. Cao, P. Yang, M. Alzenad, X. Xi, D. Wu, and H. Yanikomeroglu, "Airborne communication networks: A survey," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 9, pp. 1907–1926, 2018.

[8] B. Moision, B. Erkmen, E. Keyes, T. Belt, O. Bowen, D. Brinkley, P. Csonka, M. Eglington, A. Kazmierski, N.-h. Kim *et al.*, "Demonstration of free-space optical communication for long-range data links between balloons on project loon," in *Free-Space Laser Communication and Atmospheric Propagation XXIX*, vol. 10096. SPIE, 2017, pp. 259–272.

[9] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. De Turck, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, 2020.

[10] J. van der Hooft, M. T. Vega, T. Wauters, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, "From capturing to rendering: Volumetric media delivery with six degrees of freedom," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 49–55, 2020.

[11] T. Qualcomm, "Vr and ar pushing connectivity limits," 2018.

[12] R. Li, "Towards a New Internet for the Year 2030 and Beyond," in *Proc. 3rd Annu. ITU IMT-2020/5G Workshop Demo Day*, 2018, pp. 1–21.

[13] Li et al., R., "New IP: A Data Packet Framework to Evolve the Internet: Invited Paper," in *Proc. of IEEE 21st International Conference on High Performance Switching and Routing (HPRS)*, 5 2020, pp. 1 – 8.

[14] R. Li, U. S. Chunduri, A. Clemm, and L. Dong, *New IP: Enabling the Next Wave of Networking Innovation*. M. Boucadair and C. Jacquenet, IGI Global, 1 2021, ch. 1, pp. 1 – 42.

[15] J. Krolikowski, S. Martin, P. Medagliani, J. Leguay, S. Chen, X. Chang, and X. Geng, "Joint routing and scheduling for large-scale deterministic ip networks," *Computer Communications*, vol. 165, pp. 33–42, 2021.

[16] H. Yu, T. Taleb, and J. Zhang, "Deep reinforcement learning based deterministic routing and scheduling for mixed-criticality flows," *IEEE Transactions on Industrial Informatics*, 2022.