Digital Twin-Driven Network Architecture for Video Streaming

Xinyu Huang, Haojun Yang, Shisheng Hu, and Xuemin Shen

Abstract

Digital twin (DT) is revolutionizing the emerging video streaming services through tailored network management. By integrating diverse advanced communication technologies, DTs are promised to construct a holistic virtualized network for better network management performance. To this end, we develop a DT-driven network architecture for video streaming (DTN4VS) to enable network virtualization and tailored network management. With the architecture, various types of DTs can characterize physical entities' status, separate the network management functions from the network controller, and empower the functions with emulated data and tailored strategies. To further enhance network management performance, three potential approaches are proposed, i.e., domain data exploitation, performance evaluation, and adaptive DT model update. We present a case study pertaining to DT-assisted network slicing for short video streaming, followed by some open research issues for DTN4VS.

Introduction

Video streaming services have evolved dramatically, transitioning from simple on-demand platforms to sophisticated and real-time interactive systems. A statistics report shows that the global video streaming market size was valued at $89 billion in 2022, and is expected to grow at a compound annual growth rate of 21.5% until 2030 [1]. Emerging video streaming services, including intelligent short video streaming, extreme immersive virtual reality (VR), and holographic video streaming, demand tailored network management to satisfy users' personalized requirements [2]. For instance, by adding multiple video branches and view angles, intelligent short video streaming emphasizes the interaction with users, which is triggered by the users' swipe and rotation behaviors. To satisfy smooth video playback while reducing bandwidth consumption, communication networks should accurately mine the preferences and behavior characteristics of users for better intelligent video caching. Furthermore, extreme immersive VR and holographic video streaming aim at providing high-fidelity three-dimension (3D) object display and immersive experience, which demand efficient coordination between sensing, video tile transmission, video rendering, and specialized video codecs. To satisfy these evolving requirements, efficient network management through advanced communication technologies becomes an imperative endeavor.

Advanced communication technologies, such as enhanced mobile broadband (eMBB)-Plus, native artificial intelligence (AI), sensing, network slicing, and digital twin (DT), are expected to satisfy the above requirements [3], [4]. For instance, eMBB-Plus can provide gigabit-level data rates and seamless connections, while native AI enables intelligent data processing and decision-making. Moreover, sensing-related techniques facilitate real-time 3D object modeling, and network slicing is used to isolate network resources for prescribed service requirements. To seamlessly integrate these technologies for video streaming services, a holistic network management architecture is essential. As a promising approach, DT can realize the holistic network virtualization for video streaming services by exploiting its real-time monitoring, analytics, and emulation capabilities. Specifically, DTs can characterize users' real-time status, quality of service (QoS), and quality of experience (QoE) through native AI and sensing, and provide an emulation environment for network management by implementing tailored network slicing and resource allocation policies on eMBB-Plus. By leveraging the capabilities of DTs, an efficient holistic network management architecture for video streaming services can be realized.

However, developing an efficient DT-driven network architecture for video streaming services faces many challenges, such as

- *Lack of Efficient Data Abstraction Mechanism:* An efficient data abstraction mechanism should be developed to facilitate the real-time and intricate interplay among DTs, slice domain, and physical domain, which includes the determination of data types, granularities, and features.
- *Lack of Comprehensive Performance Evaluation Framework:* Since the DT performance consists of two-fold aspects, i.e., the accuracy and cost of itself, and its impact on network performance, it is crucial to develop a new and comprehensive performance evaluation framework to evaluate the DT performance.
- *Lack of Adaptive DT Model Update:* Due to the distinct spatiotemporal dynamics that

Digital Object Identifier: 10.1109/MNET.2024.3386030
Date of Current Version: 18 November 2024
Date of Publication: 8 April 2024

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. Haojun Yang is corresponding author.

0890-8044/24©2024IEEE

exist in network conditions and user behaviors, as well as diversified service requirements, it is essential to fine-tune tailored DT models to adapt to the dynamics and diversity.

In this article, we propose a DT-driven network architecture for video streaming (DTN4VS) to enable network virtualization and tailored network management. Three kinds of DTs, i.e., user DT (UDT), infrastructure DT (IDT), and slice DT (SDT), are built to characterize the network from the user-level, operation-level, and service-level perspectives, respectively. DTs can provide distilled user information, emulated environment, and tailored network management strategies to realize efficient network management. To tackle the mentioned challenges, we first propose an efficient data collection, fusion, and abstraction mechanism. Secondly, we propose a comprehensive performance evaluation framework, which integrates DT data freshness, QoS/QoE gain, and DT operation cost. Thirdly, we propose an adaptive DT model update method that integrates distributed and transfer learning algorithms to realize computing load balance and computing overhead reduction. A case study pertaining to DT-assisted network slicing for short video streaming is presented, followed by a discussion on potential research issues.

The remainder of this article is organized as follows. Firstly, emerging video streaming services and the corresponding communication techniques are discussed, followed by the proposed DTN4VS. Then, we discuss the challenges for DTN4VS and some potential solutions. Next, a case study about DT-assisted network slicing for short video streaming is presented. Finally, the open research issues are identified, followed by the conclusion.

DT-Driven Network Architecture for Video Streaming

In this section, we first introduce emerging video streaming services and advanced communication technologies, and then the DTN4VS architecture is proposed.

Emerging Video Streaming

Various innovative video streaming is emerging, including intelligent short video streaming, extreme immersive VR streaming, and holographic video streaming, which poses enhanced requirements on communication networks, as shown in Table 1.

1) Intelligent Short Video Streaming: It has two main characteristics, i.e., rotation-based swipe and multi-branch. The former indicates that the short video streaming will be extended from the current two-dimension (2D) format to 3D format, thus users can watch different angles through rotation-based swipe behaviors. Video tiles are selectively transmitted to user terminals based on the analysis of users' swipe behaviors to reduce network traffic load. The latter means that more video branches will be added to the main video storyline to boost users' interaction. Part of critical video branches are preferentially cached in users' buffers to save bandwidth consumption.

DTs can provide distilled user information, emulated environment, and tailored network management strategies to realize efficient network management.

2) Extreme Immersive VR Streaming: It provides immersive and interactive experience by transmitting real-time 3D videos and audio to specialized headsets or mobile devices. Since future VR streaming is expected to provide a panoramic view and ultra-high-definition resolution, field of view (FoV) transmission is an effective method to reduce the transmission burden. Furthermore, the latency requirements of extreme immersive VR are very stringent, reaching just tens of milliseconds. It is essential to develop advanced sensing technologies to quickly capture users' dynamic behaviors, and optimize the computing process for video rendering.

3) Holographic Video Streaming: It refers to the 3D holographic content transmission, which usually requires the sensing-assisted communication technology to perceive users' behaviors and conduct 3D object modeling. Due to the ultra-low delay and strong interaction, high-performance computing nodes are needed to quickly compress and render the holographic video. For instance, to swiftly respond to users' movements in the six degrees of freedom (6-DoF), the end-to-end delay needs to be lower than 5 *ms* [5].

Advanced Communication Techniques for Video Streaming

In the rapidly evolving landscape of emerging video streaming, advanced communication techniques, such as eMBB-Plus, native AI, sensing, network slicing, and DT, can offer groundbreaking solutions to satisfy enhanced requirements, including personalized video buffering, ultra-low-latency interaction, and smooth 3D video transmission.

1) eMBB-Plus: As a cornerstone technology, eMBB-Plus is designed to provide higher bandwidth capacity, wider access coverage, and smarter caching and computing, which can

Video Type	Characteristics	Video Codec	Bandwidth Requirements	Latency Requirements	Component
Intelligent Short Video	• Swipe • Multi-Branch	H.264, MPEG	4K: 45 Mbps	Several Seconds	Segments, 3D Tiles
Extreme Immersive VR Video	• Interactive • Viewpoint • Rendering	X3D, MPEG-I [6]	• 8K: 80 ~ 100 Mbps • 30K: 800 ~ 1000 Mbps	• Strong Interaction Mode: 5 ~ 10 ms • Weak Interaction Mode: 10 ~ 20 ms	2D/3D Tiles
Holographic Video	• Interactive • 3D Modeling • Rendering	HEVC, AV1, VP9	4K: 100 Mbps	• 6-DoF Movement: 5 ms • 3-DoF Movement: 20 ms	3D Tiles

TABLE 1. Emerging video streaming services.

effectively improve network throughput and reduce rebuffering.

- For intelligent short video streaming, eMBB-Plus can provide efficient distributed video caching and collaborative transcoding strategies to ensure seamless delivery of ultra-high-definition (UHD) video segments.
- For extreme immersive VR streaming, eMBB-Plus can provide the increased frequency band and exploit more advanced modulation and coding techniques to ensure the smooth transmission for high-bitrate 3D videos.
- In holographic video streaming, since 3D modeling occupies plenty of computing time, eMBB-Plus will provide more advanced data offloading and collaborative computing mechanisms to help reduce computing delay.

2) Native AI: As a built-in component in next-generation communication networks, native AI is promised to provide more intelligent data processing and resource management strategies [7].

- For intelligent short video streaming, graph neural networks (GNNs) can model the complex relationship between users and contents to enable more personalized and context-aware video recommendations and buffering.
- For extreme immersive VR streaming, deep reinforcement learning (DRL) algorithms can accurately allocate network resources and distribute video tiles to adapt to users' dynamic behaviors for smooth video playback.
- In holographic video streaming, convolutional neural networks (CNNs) can be employed for efficient data compression and semantic extraction, significantly reducing bandwidth consumption while maintaining a high fidelity.

3) Sensing: It can capture users' real-time macro and micro behaviors to help tailor network management for individual users.

- For intelligent short video streaming, advanced facial recognition sensors can capture users' micro-expressions, which are further analyzed by machine learning algorithms for intelligent video recommendations and buffer control.
- In extreme immersive VR streaming, communication and sensing signals can be multiplexed in the time, frequency, and spatial domains to improve spectrum utilization. For instance, IEEE 802.11 working group proposed the Wi-Fi sensing technology to exploit the features of the physical layer and medium access control, which can measure users' motion in real time [8].
- For holographic video streaming, efficient radio spectrum allocation and advanced beamforming technologies can effectively improve both data transmission reliability and holographic video resolution [9].

4) Network Slicing: Empowered by technologies such as software-defined networking (SDN) and network function virtualization (NFV), network slicing can provide isolated resources for emerging video streaming services to satisfy the differentiated requirements [7].

- For intelligent short video streaming, network resources can be sliced to support real-time analysis of video content and user behaviors, thereby enabling intelligent video recommendation and adaptive video delivery.
- In the realm of extreme immersive VR, specialized slices can be constructed for high-performance sensors and computing nodes to satisfy ultra-low-latency requirements.
- For holographic video streaming, dedicated slices can guarantee high bandwidth and computing requirements, facilitating real-time and high-fidelity 3D interactions.

5) Digital Twin: It is defined as a digital representation of a physical object or a process and real-time synchronization between the physical object or process [10]. DT is usually classified into three types, i.e., UDT, IDT, and SDT, deployed at the core network and network edge nodes.

- UDT corresponds to an end user that reflects its fine-grained information, such as network conditions, playback status, and interaction behaviors, etc. UDTs can emulate user status to help the network controller (NC) make tailored resource management strategies.
- IDT is a digital mirror of network infrastructure, such as a base station (BS) or an edge server, which reflects its operation status, traffic load, resource utilization, etc. IDTs can separate the resource operation function from the NC and empower the function with emulated data and tailored strategies.
- SDT is constructed by aggregating UDTs and IDTs to obtain coarse-grained distilled information, such as spatiotemporal service demand distribution, resource utilization, etc. By separating the resource planning function from the NC, SDTs can strengthen the function's capability with emulated data and tailored strategies.

Based on fine- and coarse-grained information and tailored network management strategies via DT, customized network resources are allocated to users to enhance user experience.

DTN4VS

To seamlessly integrate these emerging technologies, as shown in Fig. 1, we develop a DTN4VS framework to enhance video streaming service performance. The physical domain includes real-world video streaming infrastructures, while the slice domain leverages network slicing for prescribed QoS. The DT domain provides real-time digital replicas for data analytics, emulation, and network management decision-making, where an orchestration among UDTs, IDTs, and SDTs is intelligently coordinated to facilitate efficient network management.

1) Physical Domain: In the physical domain, the proposed DTN4VS meticulously integrates user terminals, radio access networks (RANs), edge networks, and cloud networks to build a holistic video streaming ecosystem. User terminals are not merely endpoints for video delivery but are also equipped with advanced codec technologies and adaptive bitrate (ABR) algorithms for efficient data exchange with RANs. The RAN layer consists of small BSs (SBSs) and macro BSs (MBSs) that employ advanced

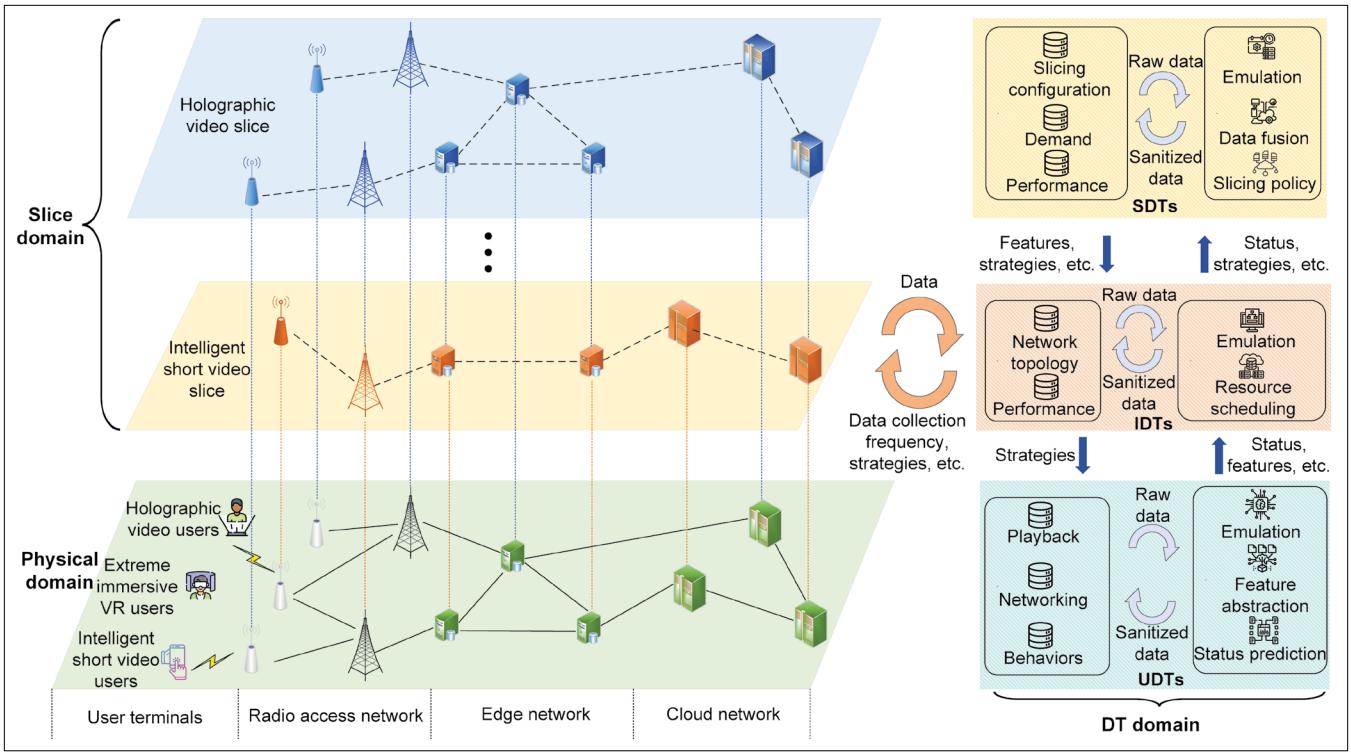

FIGURE 1. DTN4VS framework.

communication technologies, such as eMBB-Plus and sensing, to support high-throughput and high-fidelity video transmission. Edge networks provide localized computing and caching capabilities for latency-sensitive video streaming, while cloud networks handle large-scale video processing and storage. These elements collectively form the backbone of our framework, and through real-time resource scheduling and optimization in the DT domain, we can achieve holistic network management.

2) Slice Domain: Network slicing emerges as a pivotal technology for guaranteeing QoS requirements. A single physical network can be partitioned into multiple isolated slices to satisfy differentiated service requirements. For instance, one slice could be optimized for low-latency and interactive VR video streaming, while another might target high-throughput holographic video streaming. Each slice consists of a unique set of network resources, policies, and protocols, which can enable tailored control over network resources. Through intelligent orchestration in the DT domain, these slices can be dynamically adjusted to meet varying resource demands, thereby achieving a harmonious balance between resource utilization and service quality.

3) DT Domain: In the DT domain, each kind of DTs consists of a finite database and a model pool. For instance, the database of UDTs includes users' playback-related, network-related, and behavior-related data. These data can reflect users' actual watching process, and be analyzed to fine-tune native AI models, such as long short-term memory (LSTM), recurrent neural network (RNN), and CNN, etc., to emulate and predict user status, and abstract distilled features. As a crucial hub connecting UDTs and SDTs, IDTs can further aggregate UDTs' data to obtain some global information that can be provided to SDTs for slice adjustment. Furthermore, IDTs are responsible for interacting with the physical domain and slice domain, such as adaptive data collection frequency and resource management strategies, etc.

A Processing Procedure Example for Short Video

Take the short video slice as an example, the processing procedure is shown in Fig. 2. UDTs store users' historical status information, including channel conditions, locations, swipe timestamps, and preferences, etc. The data stored in UDTs are analyzed by embedded models to emulate user status, such as swipe behaviors, and abstract some essential user features, such as swipe probability distribution. In the small timescale, the emulated user status and abstracted user features are transferred to IDTs, integrated with network topology and performance metrics to emulate network operation status and design tailored resource allocation algorithms. The resource allocation policy will be delivered to the NC and then implemented on the access points to facilitate real-time video transmission. In the large timescale, the emulated network operation status and system performance are further transferred to SDTs to abstract global network information for the slicing policy adjustment. The slicing policy will be delivered to the NC and then implemented on the access points to reserve resources.

Research Challenges and Solutions

To realize the proposed DTN4VS, some research challenges need to be addressed.

Efficient Data Abstraction From Physical Domain

1) Challenge: Intuitively, addressing the complex interplay among the physical domain, slice

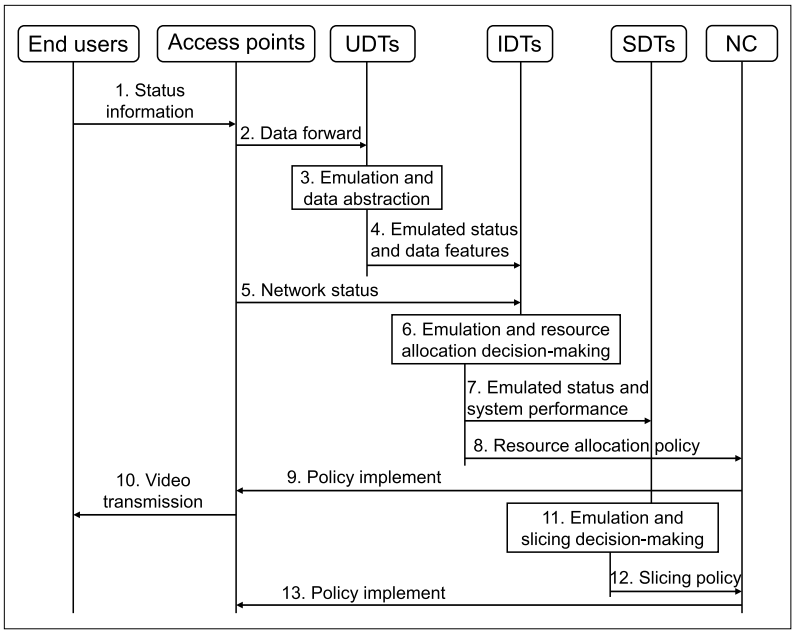

FIGURE 2. DT-assisted short video streaming processing procedure.

domain, and DT domain requires an efficient data abstraction mechanism. Since DTs need to be updated to guarantee accuracy, how to autonomously identify the types and granularities of collected data is crucial. To enhance the network management level, it is imperative to clarify what specific insights and optimizations that DTs can provide. Furthermore, the development of efficient algorithms for real-time data collection is essential, especially in ultra-low latency extreme immersive VR scenarios.

Taking user behavior data as an example, we encounter several challenges. Initially, user behavior data, such as swipe, likes, subscribe, favorites, comments, etc., constitutes a complex user profile. It is challenging to determine the importance of different user behavior data and the data update frequency for maintaining UDTs. Additionally, since user behavior data is multi-dimensional, it is challenging to discern which data dimensions can be effectively fused, and which user behavior patterns can be extracted to assist network management. Finally, user movement in VR scenarios spans from macro gestures like head and hand motions to micro shifts in viewpoint, which requires a well-designed data abstraction algorithm to realize data synchronization with low communication overhead.

2) Solution: To address the complex interplay among different domains, an efficient data collection mechanism can be first developed, which relies on data importance and distribution to adjust the collected data type and granularity. Then, since network performance is usually closely related to a part of network status, a meticulously designed data fusion mechanism can be developed to abstract some distilled features from network status to facilitate network management. Finally, a lightweight semantic abstraction algorithm can be developed to capture users' real-time behaviors and network conditions, which can help reduce transmission overhead and guarantee real-time responsiveness.

Similarly, taking user behavior data abstraction as an example, we can first employ the principal component analysis technique to analyze which data types have a strong relationship with network performance, and make an adaptive data collection granularity based on data distribution variation. For data having strong statistical characteristics, DT can generate them through the emulation module with high fidelity. Then, meticulously designed machine learning algorithms, such as autoencoders and transformers, can be leveraged to abstract and predict user behavior patterns, such as swipe probability distributions in short video streaming and region of interest (RoI) in VR streaming. Finally, some lightweight feature abstraction algorithms, such as shallow visual geometry group (VGG) and vision transformers (ViT), can be embedded into Internet of Things (IoT) sensors to abstract user behavior information for transmission in real time with low computing overhead.

Comprehensive Performance Evaluation Framework

1) Challenge: The development of a comprehensive performance evaluation framework is crucial for effectively assessing DT performance in network management. Such a framework should integrate a diverse and adaptable set of key performance indicators (KPIs) to analyze DT performance. Traditional KPIs such as latency, throughput, and buffering time may not be sufficient to capture the multi-faceted nature of DT performance, which includes not only network metrics but also user behavior and QoE. The integration of machine learning algorithms for predictive analytics, real-time data synchronization between the physical network and its DT, and edge computing for low-latency data processing adds layers of complexity to performance evaluation. Additionally, the dynamic nature of video streaming, characterized by fluctuating bit rates, stochastic video requests, and user interactivity, requires KPIs to be robust and sensitive to these fluctuations. Furthermore, the KPIs must be adaptable to different network architectures and technologies, ranging from traditional content delivery network (CDN) to next-generation communication infrastructures. Therefore, the challenge lies in developing a comprehensive performance evaluation framework that can holistically evaluate DT performance, taking into account the intricate interplay among the physical domain, slice domain, and DT domain.

2) Solution: To address this gap, a novel KPI, termed *holistic DT value*, is introduced, denoted by V. From the perspective of DT itself, data freshness [11] and model operation cost are very important metrics to optimize the synchronization between DTs and physical entities. From the perspective of DT impact, resource management gains, such as QoS and QoE, directly reflect the impact of DTs on service quality and watching experience. We refer to the utility function construction method, where service delay, energy consumption, and revenue constitute the comprehensive performance metric to optimize the communication, caching, and computing resource management in vehicular networks [12]. Therefore, the new KPI should integrate the key

metrics to provide a holistic view of network performance, which is expressed as:

$$V = \alpha \cdot \left(\frac{f}{\mathcal{F}} \right) + \beta \cdot Q(C, A, P, S, L) - \gamma \cdot R(L) \quad (1)$$

where α, β, γ are weighting factors. Here, f and $\mathcal{F}$ represent the data collection frequency and data freshness, respectively. Function, Q, represents the traditional KPIs, such as QoS and QoE, which is related to communication resource scheduling matrix **C**, caching resource scheduling matrix **A**, computing resource scheduling matrix **P**, sensing resource scheduling matrix **S**, and data abstraction level $\mathcal{L}$. In the actual network optimization, we can select a part of resource scheduling decisions as the joint optimization variables to avoid dimension curse. Here, function R reflects the DT model operation cost related to its data abstraction level $\mathcal{L}$, which can be quantified based on the model structure analysis.

Adaptive DT Model Update

1) Challenge: Unlike static models, DTs must continuously evolve to reflect actual changes in both network conditions and user behaviors. This requires sophisticated machine learning algorithms capable of processing large data volumes in real time, possibly leveraging collaborative cloud-edge computing mechanism embedded with distributed learning for the low-latency DT model update. Moreover, the dynamics of video streaming, characterized by fluctuated bit rates, diverse content types, and user interactivity, also add the complexity of the DT model update. Techniques such as DRL or generative adversarial networks (GAN) can facilitate efficient decision-making and network emulation [13], but come with their own challenges, such as the requirement of extensive training data and computing resources, and the risk of model overfitting. Furthermore, since users' network conditions, behaviors, and preferences own certain similarities, DTs can exchange part of data and learn models from each other to evolve together. This demands an efficient learning algorithm design with low computational overhead. Therefore, the challenge lies in developing adaptive DTs that can adapt to highly dynamic and complicated video streaming services.

2) Solution: To effectively tackle the intricate challenge of developing adaptive DT models for emerging video streaming services, a multi-layered solution that integrates distributed learning and transfer learning is put forth. In a hybrid cloud-edge computing architecture, distributed learning algorithms are employed to facilitate model split and parallel computing. The architecture allows DTs to parallelly process seamless network-related data and user-specific behavior data, which can effectively reduce service latency. Additionally, transfer learning can be particularly effective when exploiting similarities between different DTs. For instance, a model trained on one DT that has successfully been adapted to certain network conditions and user behaviors can be fine-tuned for another DT with similar conditions. This approach capitalizes on the inherent similarities between different DTs to quickly adapt to new scenarios, eliminating the requirement for extensive retraining and thereby reducing computing overhead. Moreover, a feedback loop mechanism can also be used to refine DT models, where real-time performance data are used to continuously refine the data processing algorithms and ensure that DT models remain accurate and effective in the face of evolving network dynamics and service diversity.

A case study is provided on DT-assisted network slicing, aimed at improving the system utility consisting of user satisfaction and resource consumption.

Case Study: DT-Assisted Network Slicing for Short Video Streaming

In this section, a case study is provided on DT-assisted network slicing, aimed at improving the system utility consisting of user satisfaction and resource consumption.

Considered Scenario

We consider a DT-assisted multicast short video streaming (MSVS) network, which consists of two BSs, an edge server (ES), and sixty UDTs. Bandwidth and computing resources are sliced (or reserved) for each multicast group to guarantee QoS requirement. Each UDT corresponds to an individual user consisting of a finite data pool and a data analysis function. Specifically, in each UDT data pool, we first simulate the user's trajectory within the University of Waterloo (UW) campus with differentiated speed, and the user's real-time channel condition is generated based on `propagationModel` at Matlab. Then, we employ the real-world dataset[1] to simulate the user's swipe timestamps and preference on the sampled YouTube 8M dataset.[2] The data analysis function investigates the user's swipe timestamps to obtain a swipe probability distribution for each video type. After the construction of UDTs, a DRL-based user clustering algorithm is implemented to cluster UDTs into different multicast groups. UDTs' swipe timestamps and preferences are used to abstract the swipe probability distribution and recommended video list for accurate bandwidth and computing resource demand prediction in each multicast group. Based on the predicted information, the NC can make appropriate bandwidth and computing resource reservation strategy for each multicast group to enhance the system utility.

We propose a hybrid data-model-driven solution, where UDTs' data are analyzed by the DRL-based user clustering algorithm to update multicast groups, and the resource reservation problem is transformed into a convex problem to obtain the optimal solution [14]. For performance comparison, we adopt two schemes, i.e., 1) heuristic solution, where multicast groups are updated based on users' preferences and locations, and bandwidth and computing resource reservation is based on historical video traffic distribution; 2) optimization-based solution, where multicast groups are updated based on the density-based spatial clustering of applications with noise (DBSCAN) algorithm, and resource reservation is based on the branch- and bound-based scheduling algorithm.

[1] ACM MM Grand Challenges: https://github.com/AItransCompetition/Short-Video-Streaming-Challenge/tree/main/data.

[2] YouTube 8M dataset: https://research.google.com/youtube8m/index.html.

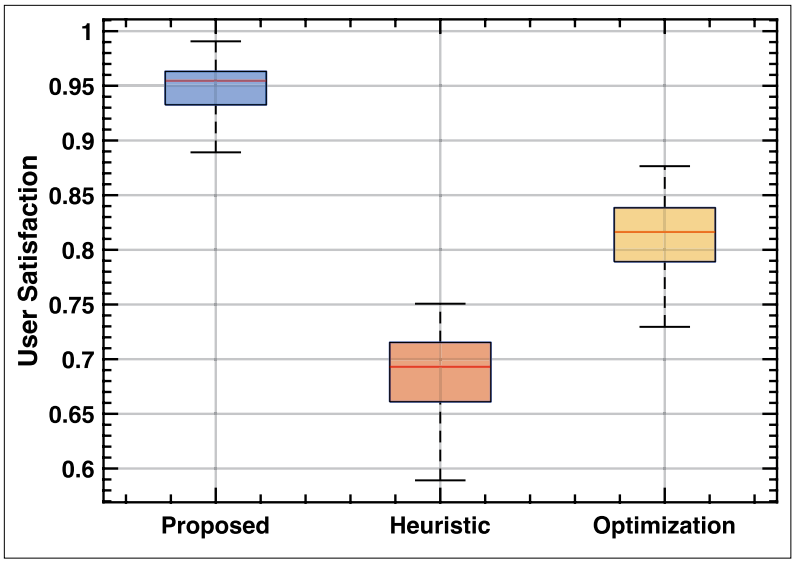

FIGURE 3. User satisfaction comparison.

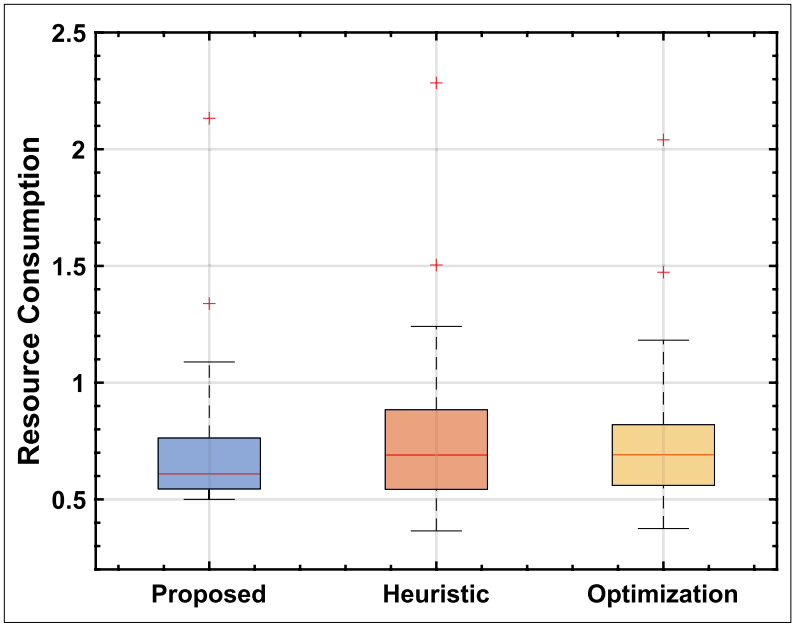

FIGURE 4. Resource consumption comparison.

Hence, constructing a closed-loop network management system requires efficient data exchange, network emulation, data synchronization, and model update.

Simulation Results

1) Simulation Settings: We emulate two BSs at the UW campus and users' initial positions are randomly and uniformly generated around two BSs. Each user moves along a prescribed path within the UW campus at a speed of 2~5 km/h. The transmission power and noise power are set to 27 *dBm* and -174 *dBm*, respectively. We sample 1000 short videos from the YouTube 8M dataset, which includes 8 video types, i.e., Entertainment, Games, Food, Sports, Science, Dance, Travel, and News. Each video has a duration of 15 *sec* and is encoded into four versions by the H. 265 encoder. The detailed simulation setting can be found at [14].

2) Performance Analysis: As shown in Fig. 3, we present user satisfaction comparison with the help of box plot. It can be observed the proposed scheme can achieve the highest user satisfaction with the lowest fluctuation, because the DT-based user clustering algorithm can well mine users' intrinsic correlation to accurately update multicast groups, while the convex optimization algorithm can make the optimal resource reservation strategy based on the updated multicast groups to enhance user satisfaction. The reason why user satisfaction fluctuates is due to users' time-varying resource demands and limited network resources. Finally, we present the resource consumption comparison in Fig. 4. It can be observed that the proposed solution can achieve lower median, third-quartile, and maximum values compared with other schemes. Our proposed scheme demonstrates superior performance with relatively minimal variations, while the heuristic scheme exhibits a larger fluctuation.

Open Research Issues

Efficient Coordination of DT Modules

As an efficient data management platform for video streaming services, the coordination of DT modules directly influences network performance. Specifically, the development of real-time synchronization mechanisms among different DT modules is crucial for ensuring that all components are operated based on the latest data, thereby enhancing network responsiveness and DT accuracy. Furthermore, the establishment of standardized communication protocols is essential for seamless interoperability among different modules, facilitating a unified and effective system architecture. Lastly, the design of adaptive algorithms that can dynamically allocate resources among modules based on real-time performance analysis and emulation is vital for optimizing resource utilization. Therefore, it is crucial to develop an efficient coordination mechanism of DT modules to improve network performance.

Closed-Loop Network Management

To realize a sustainable and continuously evolving network for video streaming services, it is essential to construct an internal and external closed-loop network management system. Internally, UDTs, IDTs, and SDTs are mainly responsible for user status analysis, network emulation, and network slicing, respectively. For instance, UDTs analyzing users' high-frequency interaction behaviors could prompt the SDTs to reserve more resources, validated by the IDTs, which can create a self-regulating loop for optimal performance. Externally, DTs continuously monitor the physical network's status and provide useful information for network management. The physical network feeds back its actual performance data to update DT data and models. Hence, constructing a closed-loop network management system requires efficient data exchange, network emulation, data synchronization, and model update.

Security and Privacy of DT

While much of the current research primarily emphasizes the constructions of UDTs, IDTs, and

SDTs to enhance video streaming services, there exists a notable oversight in addressing the issues of DT security and privacy protection. From the perspective of users, the urgency of data privacy escalates within the DTN4VS framework. Particularly, not only content service providers but also NCs need to gather sensitive user data, such as video preferences and locations, and the unique data collection model intensifies the complexity of effective data privacy regulation. Furthermore, the creation of DTs mandates collaboration amongst various stakeholders, which requires them to contribute their own data and analytic models [15]. Thus, establishing trust in such a distributed environment and protecting both data and AI model security poses significant challenges. Although current privacy-preserving techniques such as differential privacy, secure multi-party computation, and homomorphic encryption, offer potential solutions, they mandate further exploration for efficiency enhancements and tailored strategies.

Conclusion

We have proposed the DTN4VS to realize holistic network virtualization for emerging video streaming services. Specifically, DTN4VS aims to seamlessly integrate eMBB-Plus, native AI, sensing, and network slicing through DTs to achieve efficient network management. It can further separate the resource management functions from NCs and empower the functions with emulated data and tailored strategies, which can reduce the centralized computation burden and enhance network robustness. To supplement the DTN4VS's functionality, we have proposed a data importance-based abstraction mechanism, a holistic DT performance evaluation metric, and a distributed transfer learning algorithm, respectively. A case study has been presented, and some open research issues have been provided for accelerating the pace of DTN4VS development.

References

[1] Grand View Research. (2022). *Video Streaming Market Size, Share & Trends Analysis Report by Streaming Type, by Solution, by Platform, by Service, by Revenue Model, by Deployment Type, by User, by Region, and Segment Forecasts, 2023–2030*. Accessed: Oct. 10, 2023. [Online]. Available: https://www.grandviewresearch

[2] Z. Nadir et al., "Immersive services over 5G and beyond mobile systems," *IEEE Netw.*, vol. 35, no. 6, pp. 299–306, 2021.

[3] S. Dang et al., "What should 6G be?," *Nat. Electron.*, vol. 3, no. 1, pp. 20–29, 2020.

[4] O. El Marai, T. Taleb, and J. Song, "Roads infrastructure digital twin: A step toward smarter cities realization," *IEEE Netw.*, vol. 35, no. 2, pp. 136–143, Mar./Apr. 2021.

[5] Y. Huang et al., "Toward holographic video communications: A promising AI-driven solution," *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 82–88, Nov. 2022.

[6] C. Timmerer, "Immersive media delivery: Overview of ongoing standardization activities," *IEEE Commun. Standards Mag.*, vol. 1, no. 4, pp. 71–74, Dec. 2017.

[7] W. Wu et al., "AI-native network slicing for 6G networks," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 96–103, Feb. 2022.

[8] C. Chen et al., "Wi-Fi sensing based on IEEE 802.11bf," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 121–127, Jan. 2023.

[9] C. B. Barneto et al., "Full duplex radio/radar technology: The enabler for advanced joint communication and sensing," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 82–88, Feb. 2021.

[10] X. Shen et al., "Holistic network virtualization and pervasive network intelligence for 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 1–30, 1st Quart., 2022.

[11] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, Apr. 2016.

[12] L. Xu et al., "Socially driven joint optimization of communication, caching, and computing resources in vehicular networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 461–476, Jan. 2022.

[13] Y. Hua et al., "GAN-powered deep distributional reinforcement learning for resource management in network slicing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 2, pp. 334–349, Feb. 2020.

[14] X. Huang, et al., "Digital twin based user-centric resource management for multicast short video streaming," *IEEE J. Sel. Topics Signal Process.*, early access, Dec. 18, 2023, doi: 10.1109/JSTSP.2023.3343626.

[15] X. S. Shen et al., "Data management for future wireless networks: Architecture, privacy preservation, and regulation," *IEEE Netw.*, vol. 35, no. 1, pp. 8–15, Jan. 2021.

Biographies

Xinyu Huang (Student Member, IEEE) (x357huan@uwaterloo.ca) received the B.E. degree from Xidian University in 2018 and the M.S. degree from Xi'an Jiaotong University, Xi'an, China, in 2021. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Waterloo, Waterloo, ON, Canada. His research interests include digital twins, generative AI, and network resource management.

Haojun Yang (Member, IEEE) (haojun.yang@uwaterloo.ca) received the B.S. degree in communication engineering and the Ph.D. degree in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014 and 2020, respectively. He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include ultra-reliable and low-latency communications, resource management, and vehicular networks.

Shisheng Hu (Student Member, IEEE) (s97hu@uwaterloo.ca) received the B.Eng. and M.A.Sc. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include AI for wireless networks and networking for AI.

Xuemin (Sherman) Shen (Fellow, IEEE) (sshen@uwaterloo.ca) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include network resource management, wireless network security, the Internet of Things, AI for networks, and vehicular networks. Dr. Shen is a Registered Professional Engineer in ON, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Member, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.