EIC MESSAGE

Taking a "Deep" Look at Multimedia Streaming

Balakrishnan Prabhakaran, *The University of Texas at Dallas, Dallas, TX, 75080, USA*

Multimedia streaming refers to content such as audio and video that are delivered and consumed in a continuous manner. The source of the streamed multimedia content might be from one source as in "live" streaming or from a single streaming server. On-demand streaming could potentially involve multiple sources as in peer-to-peer streaming and content delivery networks. This multimedia streaming process goes through several steps during transmission.[1]

1) *Source coding*: This step compresses the data by removing temporal and spatial redundancies.
2) *Channel coding*: This step is for (wired/wireless) network transmission to handle communication errors (including packet losses) as well as to address security.
3) *Modulation*: This step is for improving the throughput while making the bitstream more resilient to errors and erasures.
4) *Scheduling*: Multimedia content, being large in size, is packetized for network delivery. Appropriate real-time scheduling of packets is needed to ensure continuous delivery of the streamed multimedia data.
5) *Routing*: Routing of the packets in an optimal manner ensures their delivery in time for the client to consume the multimedia content without interruptions.
6) *Caching/buffering*: As the streamed packets travel through the Internet, the information might be cached/buffered at intermediate nodes depending on both the type of content and the associated policies.

On the receiving side, the client follows the reverse procedure before playing it to the user: decoding the channel, sequencing the packets received, and decompressing/decoding the multimedia content. There is also feedback from the client to the source(s) regarding the client's status. This feedback information could involve not only network (TCP/IP)-level data such as congestion parameters but also user-level information such as quality of experience (QoE) in general and point of view (PoV) in cases such as mixed-reality streaming. QoE is a subjective indicator with influential factors comprising human perception, system (including displays and the Internet service), multimedia content, and context (such as purpose and task).

Deep learning algorithms have the ability to improve the above steps in the multimedia streaming process. For the latest versatile video coding (VVC) H.266 standard, a multistage early termination convolutional neural network (MET-CNN) model has been proposed for motion prediction of coding units (CUs) and sub-CUs while greatly reducing the execution time.[6] Experimental results show that MET-CNN can completely replace the complex rate-distortion optimization process and reduce the encoding time by more than 49%. For source coding, Tung and Gündüz[7] describe a deep learning-based strategy for end-to-end joint source-channel coding to directly map H.264/H.265 video to channel symbols, combining video compression, channel coding, and modulation steps into a single neural transform. Forward error correction (FEC) is a strategy used to recover the packets lost during multimedia streaming, but the additional bandwidth consumption is also a concern. Lu et al.[10] introduced RL-AFEC, an adaptive FEC scheme based on reinforcement learning (RL) to improve reconstructed video quality in the presence of packet loss while minimizing bandwidth consumption for different network conditions.

Routing strategies typically focused on leveraging deep learning techniques for ensuring network QoS. For instance, Al-Jawad et al.[3] proposed a LearnSDN framework that utilized RL to decide the most suitable routing algorithm when delivering video content over 5G networks to ensure QoS. To accomplish this, LearnSDN learns the most convenient routing algorithm to be deployed for the background traffic given the existing dynamic network conditions. Multimedia content delivery demands significant resources at various levels, and this resource allocation requires a

1070-986X © 2023 IEEE
Digital Object Identifier 10.1109/MMUL.2023.3308401
Date of current version 28 September 2023.

comprehensive view of the workload requirements. In Khudhur and Jeiad,[2] LSDStrategy is proposed to analyze the received multimedia stream based on its binary content using ML techniques with artificial and real datasets and derive the corresponding resource requirements. Mobile-edge computing (MEC) aims to reduce latency by reducing the need for end-to-end communication between client and remote cloud servers. This is done by facilitating storage and processing of video content (e.g., caching, transcoding, and preprocessing) at the base stations. Khan et al.[5] provide a detailed overview of the state-of-the-art in MEC including the use of artificial intelligence (i.e., ML, deep learning, and RL) in MEC-assisted video streaming services.

Understanding QoE in multimedia streaming and using the QoE for managing system resources (such as network QoS) is really important. Kougioumtzidis et al.[4] provide a detailed survey of significant quality metrics for QoE assessment and ML-based QoE prediction models for the goal of QoE management. Quality of live video is often limited by the streamer's uplink network bandwidth. Neural-enhanced live streaming is used to address this issue by enhancing the video quality using neural super-resolution at the server side.[11] Personalization of QoE is another important aspect that needs to be considered during multimedia streaming. A meta-learning (using the long-short term memory model as a building block) is used in Lu et al.[9] into 360° video streaming services to capture the heterogeneous preferences among viewers to quickly adapt to new viewer preferences that have proven to improve the QoE of viewers. Barakabitze and Hines[8] summarize multiple state-of-the-art approaches to QoE management using software-defined networks, including traffic rerouting to improve the received quality at the end user, using MultiPath Transport Control Protocol to improve the aggregate system throughput and QoE of the video streaming client, eliminating streaming performance problems in dynamic adaptive streaming (DASH) by utilizing a server and network-assisted DASH (SAND) approach, and personalized QoE-centric control mechanisms.

Deep ML algorithms and techniques are being widely deployed to improve the performance and efficiency of various steps comprising multimedia streaming pipelines. As the demands increase for variety and versatility of multimedia streaming addressing new content capture and display technologies, continuous research will be needed to adopt the advances in deep learning to further the advances in multimedia streaming. For instance, Barakabitze and Hines[12] identify possible approaches such as quantum ML (quantum computing + ML) to accelerate the learning of all transactions occurring at the 6G network's edge and new paradigms of internetworking including new human–machine interaction network and blockchain resource management.

REFERENCES

1. N. Thomos, T. Maugey, and L. Toni, "Machine learning for multimedia communications," *Sensors*, vol. 22, no. 3, Jan. 2022, Art. no. 819, doi: 10.3390/s22030819.
2. S. D. Khudhur and J. A. Jeiad, "LSDStrategy: A lightweight software-driven strategy for addressing big data variety of multimedia streaming," *IEEE Access*, vol. 10, pp. 111,794–111,810, Oct. 2022, doi: 10.1109/ACCESS.2022.3215531.
3. A. Al-Jawad, I.-S. Comşa, P. Shah, and R. Trestian, "LearnSDN: Optimizing routing over multimedia-based 5G-SDN using machine learning," in *Proc. 14th Int. Conf. Commun. (COMM)*, Bucharest, Romania, 2022, pp. 1–6, doi: 10.1109/COMM54429.2022.9817277.
4. G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, "A survey on multimedia services QoE assessment and machine learning-based prediction," *IEEE Access*, vol. 10, pp. 19,507–19,538, Feb. 2022, doi: 10.1109/ACCESS.2022.3149592.
5. M. A. Khan et al., "A survey on mobile edge computing for video streaming: Opportunities and challenges," *IEEE Access*, vol. 10, pp. 120,514–120,550, Nov. 2022, doi: 10.1109/ACCESS.2022.3220694.
6. T. Li et al., "DeepQTMT: A deep learning approach for fast QTMT-based CU partition of intra-mode VVC," *IEEE Trans. Image Process.*, vol. 30, pp. 5377–5390, May 2021, doi: 10.1109/TIP.2021.3083447.
7. T.-Y. Tung and D. Gündüz, "DeepWiVe: Deep-learning-aided wireless video transmission," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2570–2583, Sep. 2022, doi: 10.1109/JSAC.2022.3191354.
8. A. Barakabitze and A. Hines, "QoE management of multimedia services using machine learning in SDN/NFV 5G networks," in *Proc. Multimedia Streaming SDN/NFV 5G Netw., Mach. Learn. Manag. Big Data Streaming*, Piscataway, NJ, USA: IEEE Press, 2023, pp. 73–97, doi: 10.1002/9781119800828.ch5.
9. Y. Lu, Y. Zhu, and Z. Wang, "Personalized 360-degree video streaming: A meta-learning approach," in *Proc. 30th ACM Int. Conf. Multimedia (MM)*, New York, NY, USA: Association for Computing Machinery, 2022, pp. 3143–3151, doi: 10.1145/3503161.3548047.
10. K. Chen, H. Wang, S. Fang, X. Li, M. Ye, and H. J. Chao, "RL-AFEC: Adaptive forward error correction for real-time video communication based on reinforcement learning," in *Proc. 13th ACM Multimedia Syst. Conf.*

(MMSys), New York, NY, USA: Association for Computing Machinery, 2022, pp. 96–108, doi: 10.1145/3524273.3528184.

11. H. Yeo, H. Lim, J. Kim, Y. Jung, J. Ye, and D. Han, "NeuroScaler: Neural video enhancement at scale," in *Proc. ACM SIGCOMM Conf. (SIGCOMM)*, New York, NY, USA: Association for Computing Machinery, 2022, pp. 795–811, doi: 10.1145/3544216.3544218.
12. A. Barakabitze and A. Hines, "Multimedia streaming services delivery in 2030 and beyond networks," in *Proc. Multimedia Streaming SDN/NFV 5G Netw., Mach. Learn. Manag. Big Data Streaming*, Piscataway, NJ, USA: IEEE Press, 2023, pp. 203–220, doi: 10.1002/9781119800828.ch12.

BALAKRISHNAN PRABHAKARAN is a professor with the Computer Science Department, The University of Texas at Dallas, Dallas, TX, 75080, USA. Contact him at bprabhakaran@utdallas.edu.