

Deep Reinforcement Learning-based Video-Haptic Radio Resource Slicing in Tactile Internet

Georgios Kokkinis, Alexandros Iosifidis, Qi Zhang

DIGIT and Department of Electrical and Computer Engineering, Aarhus University

Aarhus, Denmark

Email: {gkokkinis, ai, qz}@au.dk

Abstract—Enabling video-haptic radio resource slicing in the Tactile Internet requires a sophisticated strategy to meet the distinct requirements of video and haptic data, ensure their synchronized transmission, and address the stringent latency demands of haptic feedback. This paper introduces a Deep Reinforcement Learning-based radio resource slicing framework that addresses video-haptic teleoperation challenges by dynamically balancing radio resources between the video and haptic modalities. The proposed framework employs a refined reward function that considers latency, packet loss, data rate, and the synchronization requirements of both modalities to optimize resource allocation. By catering to the specific service requirements of video-haptic teleoperation, the proposed framework achieves up to a 25% increase in user satisfaction over existing methods, while maintaining effective resource slicing with execution intervals up to 50 ms.

Index Terms—Tactile Internet, Haptic Communications, Network Slicing, Deep Reinforcement Learning, and radio resource allocation

I. INTRODUCTION

The Tactile Internet is envisioned to include the communication of touch, allowing immersive interaction and manipulation of physical or virtual objects across distance. Tactile Internet is expected to create a new dimension for human-to-machine and inter-human interaction, enabling disruptive services and applications such as teleoperation, industrial control, and immersive virtual reality. However, current communication systems fall short of the stringent latency and reliability demands of these applications. Bridging this gap calls for novel solutions, which can integrate edge computing and AI-driven optimization techniques to meet these needs, particularly in transmitting the sense of touch [1]. In Tactile Internet applications, an operator sends a control signal to a remote teleoperator system, which returns force feedback, creating a closed-loop for real-time interaction. Maintaining the stability of this loop requires a high sampling rate of 1 kHz [2]. Additionally, the teleoperator must provide both haptic and video feedback, which need precise synchronization between them. Synchronized video-haptic data transmission necessitates a resource management system capable of meeting the distinct service requirements of each modality while ensuring synchronization.

In Radio Access Network (RAN), the co-existence of haptic and video streams can be achieved through RAN slicing, which logically partitions radio resources to allow independent configuration of each stream [3]. In the case of video-haptic

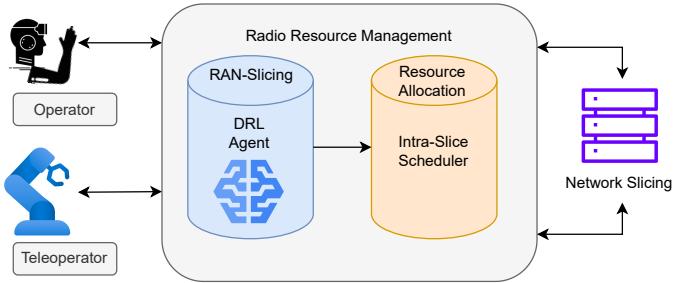


Fig. 1. Overall haptic teleoperation pipeline.

teleoperation, both the operator and teleoperator can transmit haptic feedback through an Ultra-Reliable Low-Latency Communications (URLLC) slice, while using an enhanced Mobile Broadband (eMBB) slice for high-bandwidth video feedback. Since the service requirements for video-haptic teleoperation are fixed, they are defined in the RAN-slicing algorithm before deployment and remain constant throughout operation. Resource slicing, however, must be dynamic as the environment of wireless networks is fast-changing. For dynamic RAN-slicing, common heuristics such as splitting resources to equal slices are inefficient as they do not account for slice diversity or complex environment dynamics [4]. Determining the optimal resource allocation at any given state is an NP-hard problem, rendering theoretical approaches impractical [5]. As a result, recent research has shifted towards data-driven methods, such as Deep Reinforcement Learning (DRL), where Deep Neural Network (DNN) agents are trained to infer efficient resource scheduling across diverse network conditions. In RAN-slicing, DRL enables the DNN agent to learn through reward feedback based on user satisfaction.

Several DRL-based schedulers have been proposed for radio resource management, primarily targeting general applications and services. In [4], a deep Q-network (DQN) agent collects network observations to allocate resources across video, voice, and URLLC slices, training a DNN to approximate the objective function. Y. Abiko et al. [6] improve this approach by using multiple Dueling DQNs to provide resource scheduling for a variable number of slices. In [7], a high-level DRL scheduler handles inter-slice allocation, while a lower-level scheduler manages intra-slice scheduling at 1 ms intervals [8]. However, the use of multiple DRL methods can increase complexity and resource demands.

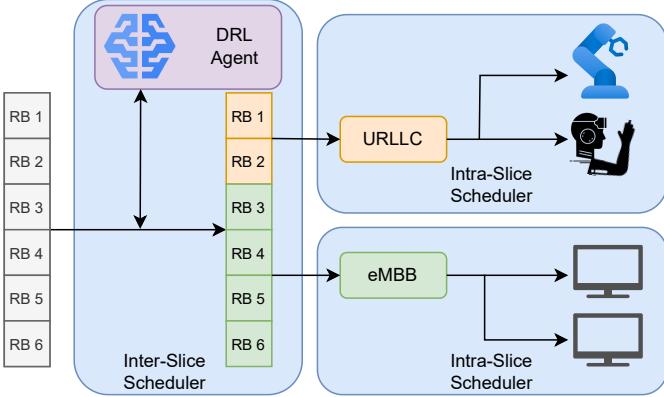


Fig. 2. Diagram of radio resource management with RAN-slicing.

To address the specific Service Level Agreements (SLAs) of network slices, the paper [9] prioritizes packet loss and latency for URLLC, and throughput for eMBB using weighted metrics. This approach shows promise for managing the diverse requirements of haptic and video modalities in teleoperation. A traditional reinforcement learning method for haptic data transmission was explored in [10], but its Q-learning approach relies on memory-intensive Q-tables. In contrast, DRL uses neural networks, making it more suitable for complex network environments. Moreover, advances in haptic technology, including improved delay tolerance and data reduction through novel encoders, have introduced new service requirements. This paper proposes a framework of DRL-based radio resource slicing that considers video-haptic teleoperation requirements. Simulations demonstrate that the proposed method significantly improves the satisfaction rate for haptic teleoperation services compared to conventional methods. The main contributions of the study are:

- Specifying and modeling the video-haptic requirements and synchronization for RAN-slicing frameworks. We use real-world haptic data traces in the performance evaluation.
- Implementing a DRL reward function system with enhanced reliability for teleoperation services, which increases the number of satisfied users by up to 25% under time-varying channel quality, when compared to conventional DRL methods.
- Developing reliable RAN-slicing with inter-slice resource allocation scheduled every 50 ms instead of every 1 ms, which reduces computing costs and is a feasible solution for resource-constrained edge computing.

For the remainder of the paper, background technologies related to RAN-slicing and DRL frameworks are outlined in Section II. In Section III, the proposed DRL-based video-haptic RAN-slicing framework is detailed, including its mathematical formulations. A comparative analysis with a baseline DRL method is presented in Section IV, followed by the conclusions in Section V.

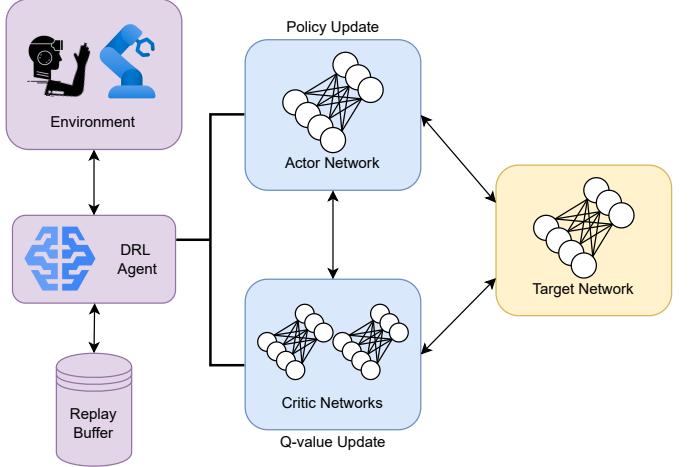


Fig. 3. Soft Actor-Critic diagram

II. BACKGROUND

This Section provides insight to the key network requirements for haptic teleoperation, as well as an overview of the Deep Reinforcement Learning framework.

A. Teleoperation service slicing

Bilateral teleoperation consists of at least an operator and a teleoperator. The users (human or machines) are able to constantly exchange haptic data. To achieve immersive teleoperation, the teleoperator also transmits video feedback, either through video streaming or by rendering and transmitting graphics of a virtual environment. In this paper, we consider video transmitted from the remote side. For the haptic and video modalities, two radio resource slices are required. The objective of the RAN-slicing algorithm is to dynamically allocate radio resources to these slices and maximize user satisfaction.

B. Service Requirements

This study focuses on three primary network requirements for haptic teleoperation: the latency, packet loss, and data rate.

1) *Latency*: Low latency is a necessary requirement for haptic signals, as it affects the stability of a closed-loop system. The recommended delay constraint for haptic teleoperation is set according to URLLC at 1 ms [11]. However, advances of control systems in teleoperation [12], [13] and human compensation for minor oscillations allow one-way latency tolerance of 10 ms without degradation of task precision or failure [2]. If latency exceeds 10 ms, humans will perceive instability and delay in movements, leading to significant degradation in teleoperation performance [14]. For video feedback, the latency constraint is more relaxed due to lack of stability constraints. However, physiological experimentation indicates that humans detect desynchronization with delays over 50 ms between haptic and video signals [15], justifying a stricter video latency threshold. Since we focus on radio resource optimization, we consider buffer latency in the end-to-end (E2E) delay loop.

TABLE I
NOTATION

Notation	Meaning	Notation	Meaning
h	haptic slice	τ	latency
v	video slice	r	data rate
\mathcal{B}	system bandwidth	ρ	packet loss
\mathcal{N}_{rb}	resource blocks	β_0	buffer occupancy
\mathcal{R}	DRL reward function	U	Teleoperation user pairs
Q	DRL long-term reward	η	mean spectrum efficiency
π	DRL action policy	σ	spectrum efficiency fluctuation
s	DRL state	t	time slot
a	DRL action	\mathbb{E}	mean value operator

2) *Packet Loss:* For packet loss, URLLC standards set a threshold of 10^{-5} . In modern teleoperation, many redundant haptic packets are omitted due to the use of Perceptual Deadband (PD) coding, which allows the operator to replay the same signal when changes are imperceptible to humans [16]. In contrast, video streaming can tolerate packet loss rates of up to 10^{-1} .

3) *Data Rate:* For haptic signals, the baseline requirement for data rate is quite low compared to video streaming, due to the small packet size of haptic signals. With the use of PD coding, we can achieve up to a 90% reduction in transmitted haptic packets [17]. For video data rate, haptic teleoperation sets a minimum recommendation of 30 Frames Per Second (FPS) with a high definition (HD) camera [18]. For modern video codecs like AV1 [19], 30 FPS can be achieved at a data rate of 4 Mbps.

C. Deep Reinforcement Learning

In reinforcement learning, an agent selects actions based on the observed state of the environment and refines its policy iteratively through a reward function. For complex scenarios like RAN-slicing with time-varying network conditions, a Deep Neural Network (DNN) approximates the policy function, mapping states to actions. We employ the Soft Actor-Critic (SAC) method, which is particularly effective in dynamic environments. SAC optimizes a balance between the expected return and policy entropy, using an entropy coefficient to manage the trade-off between exploration and exploitation. The agent's experiences, comprising state, action, reward, and next state, are stored in a replay buffer, allowing the SAC agent to sample mini-batches for policy updates [20]. Since SAC operates in continuous action spaces, its outputs are mapped to discrete values for resource allocation per slice, as described in [9].

III. DRL-BASED VIDEO-HAPTIC RAN-SLICING

In this Section, we provide an overview of the proposed DRL-based framework for video-haptic RAN-slicing in teleoperation.

A. DRL key components

The DRL method has the following key components:

- **Agent:** The agent is a Deep Neural Network (DNN) trained to optimize the actions taken based on a given

TABLE II
AGENT HYPERPARAMETERS (SOFT ACTOR-CRITIC)

Hyperparameter	Value	Hyperparameter	Value
Hidden Layers	3	Batch Size	1024
Hidden Layer Size	512, 512, 256	Learning Rate	0.0001
Discount Factor γ	0.99	Target Update Rate	0.005
Replay Buffer Size	1×10^6	Entropy Coefficient λ	Auto-tuned
Optimizer	Adam	Target Smoothing	0.995

observable state of the environment. In SAC, it consists of the actor and critic DNNs, which are responsible for stabilizing and refining the policy updates during training.

- **State:** The network state represents the observable space of the agent and consists of key communication properties of the users, which are the current spectrum efficiency η , buffer occupancy β_0 , and the latency τ , packet loss ρ and data rate r . To reduce the state dimensions, only the average of each component is observed. This can influence the optimization of the DRL's action policy. However, as detailed in subsection III-C, the latency reward function is determined based on the user with the lowest spectrum efficiency, thereby guiding the resource slicing strategy.
- **Action:** In RAN-slicing, the action is the allocation ratio of the total frequency bandwidth per slice. The smallest frequency resource unit in 5G New Radio and 6G is the Resource Block (RB). Therefore, the action of the DRL dictates the number of RBs allocated for each slice, leading to a discrete action space.
- **Reward:** The reward feedback is a function that evaluates a given action. In the case of RAN-slicing, the feedback is positive when the actions lead to achieving service requirements for all users. DRL agents are trained to choose actions based on maximizing the long-term reward.

B. VH-SAC formulation

In RAN-slicing with the soft actor-critic DRL method, the actor-network chooses a slicing action a_t at every time slot t , given a network state s_t . The policy π with which the actor chooses an action is stochastic, meaning that a_t is selected from a distribution with a probability ϵ . The policy update aims to maximize a combination of the expected rewards and policy entropy:

$$\pi^* = \arg \max_{\pi} \mathbb{E} \left[\sum_t \gamma^t (\mathcal{R}(s_t, a_t, s_{t+1}) - \lambda \log \pi_{\epsilon}(a_t | s_t)) \right], \quad (1)$$

where γ is the reward discount factor, λ is the entropy coefficient. The Q-value, the expected long-term reward, is given by the following equation:

$$Q(s_t, a_t) = \mathbb{E} \left[\mathcal{R}(s_t, a_t, s_{t+1}) + \gamma \left(\min_{i=1,2} Q_{\theta_i}(s_{t+1}, a_{t+1}) - \lambda \log \pi(a_{t+1} | s_{t+1}) \right) \right], \quad (2)$$

where Q_{θ_1} and Q_{θ_2} are the Q-values of the two critic-networks.

TABLE III
DATA TRAFFIC AND NETWORK CONSTRAINTS

Traffic characteristic	Haptic Data	Video Data
Packet Size (bytes)	8	16667
Packet Arrival Rate (ms)	1	33.33
Jitter (ms)	0	[min,max] = [-4,4]
Slice requirement	Haptic Slice	Video Slice
Latency (ms)	≤ 10	≤ 50
Packet Loss (%)	≤ 0.001	≤ 10
Data Rate (Mbps)	≥ 0.2	≥ 4

TABLE IV
SIMULATION PARAMETERS

Experiment	U	η (bits / Hz)	σ (%)
Over U	10 - 25	5	25
Over η	20	4.5 - 5.5	25
Over σ	20	5	0 - 50

C. Reward function for teleoperation

For the reward function \mathcal{R} , the key concept is to avoid exceeding the thresholds set by the service requirements for each data modality. In many studies, the average of each requirement for all users inside each slice is considered. We set $\bar{r}^k, \bar{\rho}^k, \bar{\tau}^k$ to be the average data rate, packet loss, and latency for each slice $k \in \{h, v\}$, where h denotes the haptic slice and v the video slice. For both the haptic and video slices, the data rate reward is calculated as follows:

$$\mathcal{R}_r^k = \begin{cases} -\frac{r_0^k - \bar{r}^k}{r_0^k}, & \text{if } \bar{r}^k < r_0^k \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where r_0^k is the data rate requirement of the slice. We use a negative reward function in the DRL to encourage minimization of the objective, with the reward reaching zero once the target is reached. Similarly we formulate the reward function for the packet loss:

$$\mathcal{R}_\rho^k = \begin{cases} -\frac{\bar{\rho}^k - \rho_0^k}{c\rho_0^k}, & \text{if } \bar{\rho}^k > \rho_0^k \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where we set the constant c at the denominator during the initial stages of training to stabilize the reward contribution of packet loss as it can reach much higher values than the requirement.

The latency reward functions for the haptic and video slices $\mathcal{R}_\tau^h, \mathcal{R}_\tau^v$ follow separate calculation approaches. For the video slice, we consider the synchronization threshold, which is based on the relative delay between haptic and video packets:

$$\mathcal{R}_\tau^v = \begin{cases} -\frac{|\bar{\tau}^v - \bar{\tau}^h| - \tau_{\text{sync}}}{\tau_{\text{sync}}}, & \text{if } |\bar{\tau}^v - \bar{\tau}^h| > \tau_{\text{sync}} \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

where τ_{sync} represents the maximum tolerable synchronization delay between video and haptic packets, ensuring that the

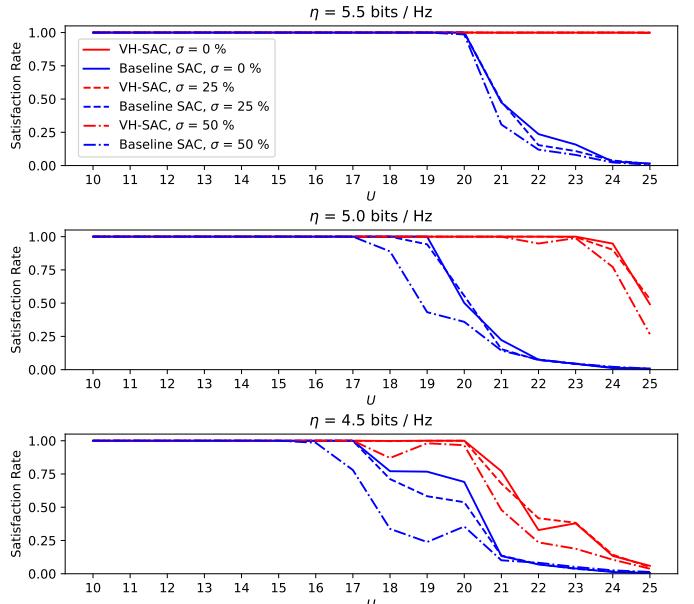


Fig. 4. Satisfaction Rate over the number of operator-teleoperator user pairs.

relative delay remains within acceptable limits to maintain a seamless user experience. For the haptic slice, we consider the average values of the data rate and packet loss, but the strict latency requirement must be satisfied for all users. For latency, we calculate the reward according to the user with the worst spectrum efficiency:

$$\mathcal{R}_\tau^h = \begin{cases} -\frac{\hat{\tau} - \tau_0^k}{\tau_{\max}^k}, & \text{if } \hat{\tau} > \tau_0^k \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

where $\hat{\tau}$ is the latency of the user with the worst spectrum efficiency, and τ_{\max}^k is the maximum time that a packet is allowed to stay in the buffer without being discarded. By controlling the haptic latency according to the worst user, we aim for a reliable system for the users who have poor channel quality while maintaining satisfaction of all users. The total reward function is calculated as the sum of all the components mentioned above:

$$\mathcal{R} = \sum_{k \in \{h, v\}} (\mathcal{R}_r^k + \mathcal{R}_\rho^k + \mathcal{R}_\tau^k), \quad (7)$$

For intra-slice scheduling, we distribute the RBs of each slice \mathcal{N}_{rb}^k to the user RBs \mathcal{N}_{rb}^u proportionally to the user buffer occupancy β_o^u :

$$\mathcal{N}_{rb}^u = \left\lfloor \frac{\beta_o^u}{\beta_o} \mathcal{N}_{rb}^k \right\rfloor, \quad (8)$$

where β_o is the cumulative buffer occupancy for the whole slice. Unlike conventional resource schedulers, which often use round-robin to achieve uniform resource distribution, the proportional scheduler derived from Eq. 8 prioritizes users with higher buffer occupancy, allocating more resources to accommodate higher packet loads.

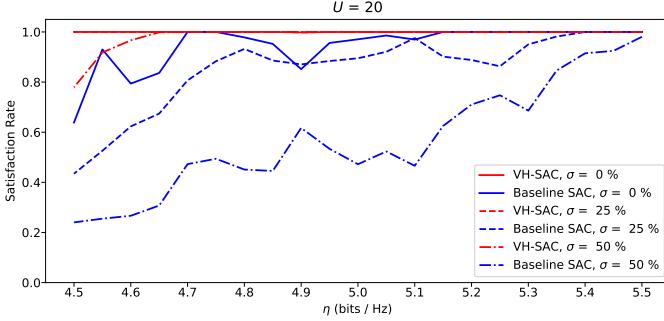


Fig. 5. Satisfaction Rate over the mean spectrum efficiency.

IV. PERFORMANCE EVALUATION

A. Baseline

To the best of our knowledge, no DRL methods have been proposed that are specifically designed to address the latency, packet loss, and data rate requirements associated with video-haptic teleoperation. Hence, we use the SAC method from [14] as a baseline, which performs slicing for general URLLC and eMBB services, and adapt it to provide resource allocation for the haptic and video slice of the teleoperation service. This baseline is reasonable for comparison as it utilizes the SAC DRL method, while it also inherits service requirements from URLLC and eMBB standards, yielding better results than common heuristics.

B. Simulation environment

1) *Data Traffic*: We consider two types of users for the teleoperation service: the operator and the teleoperator. The operator generates haptic data for position and velocity, while the teleoperator provides haptic force feedback and video feedback. Each teleoperator uses one haptic and one video device, while the operator uses a single haptic device. Haptic data traces were captured from real-world activities using the Novit Falkon haptic device in a virtual environment, including free movements, object tapping, pressing, pushing, and interacting with rigid objects. The link of the database is provided in [21]. Each haptic packet is 8 bytes (6 bytes for 3D data and 2 bytes for control overhead). Video feedback uses a statistical model considering an HD camera at 30 FPS. The traffic characteristics and requirements per data modality are summarized in Table III.

2) *RAN configuration*: We utilize the channel modeling from [9], where Spectrum Efficiency (SE) files are extracted by using Quadriga simulation tool for channel profiling, at Frequency Range 1 with a center frequency of 2 GHz. The files were extracted from users being uniformly distributed during the trials, and the SE is measured every 1 ms. The 1 ms time slot is also used as the Transmit Time Interval (TTI). The provided SE files are time-varying and insert high dynamics to the simulation. In our simulated environment, we consider a bandwidth \mathcal{B} of 20 MHz, divided among $\mathcal{N}_{rb} = 100$ RBs.

3) *Simulation Parameters*: For the experiments we iterate over three simulation parameters, the number of operator-teleoperator user pairs U (i.e. for $U = 20$, we have 20 opera-

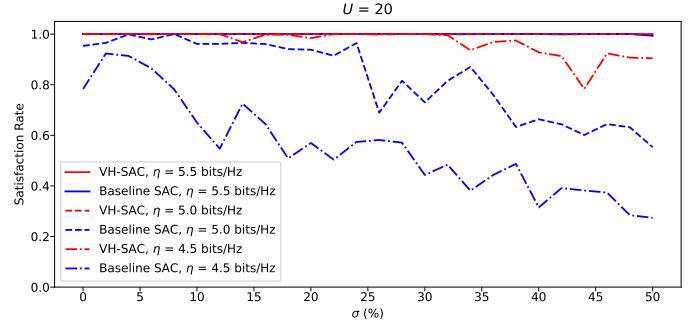


Fig. 6. Satisfaction Rate over SE fluctuation.

tors and 20 teleoperators), the average spectrum efficiency η , and the fluctuation of the spectrum efficiency σ . As shown in Table IV, we iterate over $U = 10-25$, $\eta = 4.5-5.5$ bits / Hz, and $\sigma = 0-50$ %. For each iterated value we keep the other two fixed.

C. Satisfaction Rate

1) *VH-SAC vs. baseline SAC*: We evaluate the performance of each method by the satisfaction rate (SR) of the teleoperation users. We define SR as the time that the latency, packet loss and data rate requirements of both operator and teleoperator are satisfied, divided by the total runtime of a simulation trial. Each trial lasts for 1×10^4 time slots or 10 seconds. The performance of the proposed video-haptic-aware SAC are shown in Fig. 4, Fig. 5, and Fig. 6.

In Fig. 4, we measure the SR over an increasing number of pairs of teleoperation users. At average SE of 5.5 bits/Hz, the proposed method can preserve satisfaction even with fluctuation of 50%, while for the baseline SAC the SR drops dramatically when the number of user pairs exceeds 20 (i.e., 20 teleoperation services). While the overall satisfaction rate decreases as the channel quality for all pairs decreases, the proposed method shows much more graceful degradation compared to the baseline SAC, achieving a higher overall SR. Similar results are observed in Fig. 5 and Fig. 6, where we perform simulations over a range of average SE and fluctuation range. Fig. 5 shows that for a system of $U = 20$, the proposed method can ensure good SR, while η is varying between 4.5 bits/Hz and 5.5 bits/Hz, however, the baseline SAC cannot sufficiently support a system of 20 pairs under the same channel conditions, especially during average SE below 5 bits/Hz plus fluctuations. Similarly, Fig. 6 shows that in a system of 20 pairs of users, the proposed method can provide stable and high SR in comparison with the baseline SAC under SE fluctuations.

2) *Multi-step prediction*: Dynamic RAN-slicing can place substantial demands on computing resources, which becomes especially critical in edge computing scenarios where computational capacity is limited. Frequent updates, such as those occurring every millisecond, exacerbate end-to-end latency as each iteration requires resource-intensive processing. To tackle this challenge, we are interested in measuring the performance of the proposed method when it performs radio resource

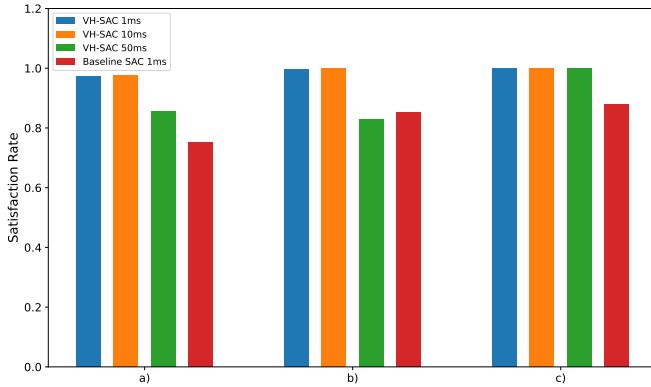


Fig. 7. Satisfaction Rate over multi-step prediction and baseline SAC for iterating over a) Teleoperation user pairs U b) mean spectrum efficiency η c) SE fluctuation σ

slicing every 10 or 50 ms, instead of every 1 ms. In Fig. 7 we showcase the performance of such scenarios compared to the baseline. While the satisfaction shows minor drop due to decreased scheduling granularity, it is still more robust compared to other methods, even when it runs at every 50 ms. This experiment shows that the proposed method can reduce computation cost in DRL-based resource allocation.

V. CONCLUSION

In this paper, we developed a DRL-based RAN-slicing framework for video-haptic teleoperation, addressing the unique challenge of managing both video and haptic data streams with a delay synchronization constraint between the two modalities. We utilized real-world haptic data traces for realistic data traffic in the experiments. Our proposed DRL reward function enhances the reliability of video-haptic teleoperation services, significantly improving user satisfaction. Furthermore, our framework remains robust even with extended resource allocation intervals, thus reducing computational overhead and increasing the feasibility of resource slicing deployment at the edge. Our findings therefore contribute to efficient and adaptive resource management solutions, supporting the realization of immersive teleoperation services.

ACKNOWLEDGMENT

This research was supported by the TOAST project, funded by the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie Actions Doctoral Network (Grant Agreement No. 101073465), the Danish Council for Independent Research project eTouch (Grant No. 1127- 00339B) and NordForsk Nordic University Cooperation on Edge Intelligence (Grant No. 168043).

REFERENCES

- [1] J. Sachs, L. A. A. Andersson, J. Araújo, C. Curescu, J. Lundsjö, G. Rune, E. Steinbach, and G. Wikström, “Adaptive 5g low-latency communication for tactile internet services,” *Proceedings of the IEEE*, vol. 107, no. 2, pp. 325–349, 2019.
- [2] K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi, and M. Dohler, “Toward haptic communications over the 5g tactile internet,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3034–3059, 2018.
- [3] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5g wireless network slicing for embb, urlc, and mmtc: A communication-theoretic view,” *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.
- [4] R. Li, Z. Zhao, Q. Sun, C.-L. I, C. Yang, X. Chen, M. Zhao, and H. Zhang, “Deep reinforcement learning for resource management in network slicing,” *IEEE Access*, vol. 6, pp. 74 429–74 441, 2018.
- [5] J. Zhao, Q. Li, Y. Gong, and K. Zhang, “Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7944–7956, 2019.
- [6] Y. Abiko, T. Saito, D. Ikeda, K. Ohta, T. Mizuno, and H. Mineno, “Flexible resource block allocation to multiple slices for radio access network slicing using deep reinforcement learning,” *IEEE Access*, vol. 8, pp. 68 183–68 198, 2020.
- [7] J. Mei, X. Wang, K. Zheng, G. Boudreau, A. B. Sediq, and H. Abou-Zeid, “Intelligent radio access network slicing for service provisioning in 6g: A hierarchical deep reinforcement learning approach,” *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 6063–6078, 2021.
- [8] M. Chen, L. Zhao, J. Chen, X. Wei, and M. Guizani, “Modal-aware resource allocation for cross-modal collaborative communication in iiot,” *IEEE Internet of Things Journal*, vol. 10, no. 17, pp. 14 952–14 964, 2023.
- [9] C. V. Nahum, V. H. L. Lopes, R. M. Dreifuerst, P. Batista, I. Correa, K. V. Cardoso, A. Klautau, and R. W. Heath, “Intent-aware radio resource scheduling in a ran slicing scenario using reinforcement learning,” *IEEE Transactions on Wireless Communications*, vol. 23, no. 3, pp. 2253–2267, 2024.
- [10] A. Ajiaz, “Hap – SliceR: A radio resource slicing framework for 5g networks with haptic communications,” *IEEE Systems Journal*, vol. 12, no. 3, pp. 2285–2296, 2018.
- [11] C. She, C. Sun, Z. Gu, Y. Li, C. Yang, H. V. Poor, and B. Vucetic, “A tutorial on ultrareliable and low-latency communications in 6g: Integrating domain knowledge into deep learning,” *Proceedings of the IEEE*, vol. 109, no. 3, pp. 204–246, 2021.
- [12] X. Xu, C. Schuwerk, B. Cizmeci, and E. Steinbach, “Energy prediction for teleoperation systems that combine the time domain passivity approach with perceptual deadband-based haptic data reduction,” *IEEE Transactions on Haptics*, vol. 9, no. 4, pp. 560–573, 2016.
- [13] D. Sun, F. Naghdly, and H. Du, “Wave-variable-based passivity control of four-channel nonlinear bilateral teleoperation system under time delays,” *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 1, pp. 238–253, 2016.
- [14] R. Chaudhari, C. Schuwerk, V. Nitsch, E. Steinbach, and B. Färber, “Opening the haptic loop: Network degradation limits for haptic task performance,” in *2011 IEEE International Workshop on Haptic Audio Visual Environments and Games*, 2011, pp. 56–61.
- [15] M. Di Luca and A. Mahnan, “Perceptual limits of visual-haptic simultaneity in virtual reality interactions,” in *2019 IEEE World Haptics Conference (WHC)*, 2019, pp. 67–72.
- [16] X. Xu, C. Schuwerk, B. Cizmeci, and E. Steinbach, “Energy prediction for teleoperation systems that combine the time domain passivity approach with perceptual deadband-based haptic data reduction,” *IEEE Transactions on Haptics*, vol. 9, no. 4, pp. 560–573, 2016.
- [17] S. Hirche, P. Hinterseer, E. Steinbach, and M. Buss, “Transparent data reduction in networked telepresence and teleaction systems. part i: Communication without time delay,” *Presence*, vol. 16, no. 5, pp. 523–531, 2007.
- [18] P. Stotko, S. Krumpen, M. Schwarz, C. Lenz, S. Behnke, R. Klein, and M. Weinmann, “A vr system for immersive teleoperation and live exploration with a mobile robot,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 3630–3637.
- [19] Y. Chen, D. Murherjee, J. Han, A. Grange, Y. Xu, Z. Liu, S. Parker, C. Chen, H. Su, U. Joshi, C.-H. Chiang, Y. Wang, P. Wilkins, J. Bankski, L. Trudeau, N. Egge, J.-M. Valin, T. Davies, S. Midtskogen, A. Norkin, and P. de Rivaz, “An overview of core coding tools in the av1 video codec,” in *2018 Picture Coding Symposium (PCS)*, 2018, pp. 41–45.
- [20] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” *CoRR*, vol. abs/1801.01290, 2018.
- [21] D. Rodriguez-Guevara, “Kinesthetic data traces,” Dataset, available: <https://cloud.lkn.ei.tum.de/s/M7xWrCecdYYZJsw>.