



Multisite gaming streaming optimization over virtualized 5G environment using Deep Reinforcement Learning techniques

Alberto del Rio ^{a,*}, Javier Serrano ^a, David Jimenez ^a, Luis M. Contreras ^b, Federico Alvarez ^a

^a Universidad Politécnica de Madrid, GATV Research Group, Madrid, 28040, Spain

^b Telefónica I+D, Global CTIO Unit, Madrid, 28050, Spain

ARTICLE INFO

Keywords:

Deep Reinforcement Learning
Multiaccess edge computing
Software-defined networks
5G
Network function virtualization
Adaptive multimedia
A3C
Quality of Experience

ABSTRACT

The massive growth of live streaming, especially gaming-focused content, has led to an overall increase in global bandwidth consumption. Certain services see their quality diminished at times of peak consumption, degrading the quality of the content. This trend generates new research related to optimizing image quality according to network and service conditions. In this work we present a gaming streaming use case optimization on a real multisite 5G environment. The paper outlines the virtualized workflow of the use case and provides a detailed description of the applications and resources deployed for the simulation. This simulation tests the optimization of the service based on the addition of Artificial Intelligence (AI) algorithms, assuring the delivery of content with good Quality of Experience (QoE) under different working conditions. The AI introduced is based on Deep Reinforcement Learning (DRL) algorithms that can adapt, in a flexible way, to the different conditions that the multimedia workflow could face. That is, adapt, through corrective actions, the streaming bitrate, in order to optimize the QoE of the content on a real-time multisite scenario. The results of this work demonstrate how we have been able to minimize content losses, as well as the fact of obtaining high audiovisual multimedia quality results with higher bitrates, compared to a service without an optimizer integrated in the system. In a multi-site environment, we have achieved an improvement of 20 percentage points in terms of blockiness efficiency and also 15 percentage points in block loss.

1. Introduction

Gaming streaming services are emerging entertainment that is growing in terms of usage and resource demand. Their market is expected to register a CAGR of 9% during the forecast period 2023–2028 [1]. It is essential to anticipate this growing demand, bearing in mind that the end user must define the final perceived quality of application or services. In view of this, not only services must be developed to guarantee the Quality of Service (QoS) of the infrastructures, meeting the objectives of bandwidth and latency, but also guarantee a good Quality of Experience (QoE) perceived by the end-user. Generally, multimedia services operate at low bitrates, reflecting lower quality of content. This bitrate level is adjusted as a trade-off for serving multimedia to a large number of users. This kind of service is growing in popularity and their needs could only be satisfied with the combination of new-generation 5G networks and cloud technologies [2,3].

Traditional content, such as films, music, or video, is demanding in terms of bandwidth consumption [4], but a more challenging experience is video gaming in remote environments, where users want to see how a specific gamer plays a video game while commenting on

his experience through the game. This gaming streaming service has increased the number of daily users, and new platforms have appeared to provide this kind of service, such as Twitch [5]. This industry is constantly growing [6], absorbing a large number of viewers of traditional content, so research should focus on improving the efficiency of these services.

Some solutions appeared in recent years where gaming experiences are considered a key technology demonstrated in use cases and proof-of-concept (e.g., H2020 5G PPP 5G-MEDIA project [7]). These solutions are preliminary research to implement all the capabilities of 5G-based cloud infrastructures. The European Commission (EC) launched, under the H2020 framework, a call for the provision of real 5G infrastructures to test pilots in real 5G environments. One of those projects is 5G EVE [8].

As the latest services are demanding increasingly higher data transmission capacities, planning, and operating networks lead to more complex and critical issues. Adding intelligence to networks is one of the most attractive ways to make management easier and more effective in terms of OPEX and CAPEX for operators [9]. The addition

* Corresponding author.

E-mail address: arp@gatv.ssr.upm.es (A. del Rio).

<https://doi.org/10.1016/j.comnet.2024.110334>

Received 8 September 2023; Received in revised form 16 February 2024; Accepted 12 March 2024

1389-1286/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Nomenclature

A2C	Advantage Actor–Critic
A3C	Asynchronous Advantage Actor–Critic
AI	Artificial Intelligence
CDN	Content Delivery Network
DRL	Deep Reinforcement Learning
EC	European Commission
GOP	Group of Pictures
HFR	High Frame Rate
KPI	Key Performance Indicator
MAPE	Monitoring, Analysis, Policy, and Execution
MEC	Multi-access Edge Computing
ML	Machine Learning
NFV	Network Function Virtualization
PLR	Packet Loss Rate
PNF	Physical Network Function
QoE	Quality of Experience
QoS	Quality of Service
RL	Reinforcement Learning
SDN	Software Defined Networking
TG	Traffic Generator
TM	Traffic Manager
VNF	Virtual Network Function
VQA	Video Quality Assessment

of this kind of intelligence to manage the correct behavior of a service is aligned with the evolution of 5G networks [10]. Early research on Beyond 5G (B5G) and 6G topics is focusing on intelligent adaptation of the network to possible issues that create event-driven and intent-based smart networking, among many other improvements [11–13]. The concept of smart networks is the key to future generations of networks. This concept assumes a combination of Smart Connectivity, Data Analytics based on the application of Artificial Intelligence (AI) and Machine Learning (ML), high-performance distributed computing and Cybersecurity [14].

The workload for these services is done in large data centers. Typically, gaming streaming platforms use their own data centers to accommodate their service, although the demand to rent a third-party provider of virtual resources to support the workload has grown greatly [15,16]. This last solution brings new advantages, like the high flexibility of the business model that allows us to host different services in parallel with resource sharing between all of them. All the mentioned solutions require a complex infrastructure for their perfect operation, so the main challenge is to design a flexible configuration to obtain the best QoE.

Finally, to provide this service for users that could be placed in different regions of the globe, the service uses a multisite approach, where each site counts with Edge Cloud infrastructure, allowing performing an in-network adaptation of the content for each user.

In this research article, we present a significant advancement in the field of multimedia service optimization by harnessing the power of Deep Reinforcement Learning (DRL) algorithms in the context of a multisite scenario [17]. Our work introduces a novel approach that addresses a critical gap in the existing literature: the absence of real-time optimization strategies tailored specifically for a multisite environment.

At the heart of our study lies the application of DRL algorithms to optimize a multimedia service, with a specific focus on the challenging domain of video gaming. While previous research has explored video analysis and transmission optimization, the uniqueness of our approach lies in its multisite context and real-time capabilities. By considering the

intricacies of video gaming, we address the balance between network conditions, video quality, and transmission bitrate.

One of the key contributions of our work is the introduction of a parallel optimization paradigm. Unlike previous methods that focus primarily on optimizations on a single site [18], our approach takes into account multiple sites simultaneously. This innovation is a fundamental departure from the existing state of the art [19–21], where optimization strategies have predominantly been designed for isolated systems. By optimizing different sites in parallel, we enable a holistic enhancement of the multimedia service, effectively transforming how optimization challenges are addressed in a multisite setting.

Our proposed algorithm optimizes the transmission bitrate, accounting for the quality of the video content received. This approach involves the use of transcoders that encode and distribute content to end users. The central goal of our optimization strategy is to dynamically adjust the bitrate of the stream, in order to maximize QoE under varying conditions. Moreover, our approach is poised to handle extreme scenarios, such as network bottlenecks that may lead to content loss. By leveraging the intelligence provided by DRL algorithms, our system can assess real-time conditions and enact corrective measures to ensure optimal video quality-bitrate trade-offs.

The work presented here is organized as follows. Section 2 relates to similar work developed to establish the baseline of this research. In Section 3 there is a description of the architecture of the work to understand the different components at a high level. In Section 4 we analyze the tools used to demonstrate this article, specifically the content of the gaming media used to demonstrate a real-life environment. Also, in Section 4, the theoretical part of the article to explain the algorithm developed. Section 5 describes the analysis of the results obtained, presenting the approach obtained with graphical data. Finally, Section 6, closes the article with a conclusion about the results obtained and how to continue future work.

2. Related work

This section of the article covers various research areas related to Multi-access Edge Computing (MEC), general video quality assessment, gaming streaming, and AI solutions. This section highlights the importance of MEC in improving network performance and improving low-latency communication and transmission bandwidth. Additionally, it covers research on objective metrics for multimedia content quality assessment, innovative approaches for game streaming, and the use of AI solutions in various contexts. The objective of this review is to provide an overview of the current state of research in these areas and to highlight the most significant contributions to date.

2.1. MEC opportunities

A complete 5G platform enables several services for different purposes. These platforms are usually presented as the interconnection of MEC environments. In general, these MEC architectures allow resource optimization that is not possible in individual environments [22,23]. They are proliferating due to their speed when it comes to deploying new services and enabling new actors [24,25].

Our work focuses on QoE optimization on the viewer side, but there are other works that perform optimizations through coding techniques directly on the server hosted in the MEC. These techniques, such as layer-caching in [26] demonstrate different scopes of optimization in the multimedia scenario.

2.2. General quality assessment

The field of Video Quality Assessment (VQA) has generally focused on traditional content, such as television content. Generally, these analyses are subjectively produced by evaluating personal results to generate a quality scale [27–29]. Typically, these quality assessments are performed on a subjective MOS scale [30–32]. Others instead focus on image distortion-based assessment methods [33].

Nevertheless, many of the databases on which subjective algorithms are trained are out of date [34] and are generally based on traditional content. In other words, they have not considered the effects that new video encoders add and do not fully reflect the evaluation of the video. For this reason, certain works seek to find a model to predict the quality of non-reference video measurement [35], even for gaming streaming applications [36].

Some works combine the characteristics of 5G networks with the assessment of QoE. In [37] they proposed to model QoE in a 5G environment in streaming video. This work is similar in concept to us in terms of what our work uses a QoE probe by estimating the quality of the video in the streaming itself as an indicator in real-time.

2.3. Gaming streaming

Gaming content requires a more in-depth analysis given its High Frame Rate (HFR) nature (stable 60 FPS). There is some work like [38] in which a specific dataset for gaming known as GamingVideoSET was developed. It can serve as an approximation to the ideas of analysis in the field of gaming. However, the sequences analyzed do not meet the requirements of a HFR.

Generally, video analysis is performed on the basis of certain metrics by developing a quality scale. These metrics allow to estimate new video sequences on this scale [39,40]. However, there are cases where they based the scale on an analysis of the Packet Loss Rate (PLR) [41].

Another work related to architecture, and, above all, the gaming field applied in that publication, can be seen in [42]. They analyzed the need and use of GPUs for content that requires graphics acceleration, but the most applicable to the work is the multimedia analysis of gaming content compared to traditional ones such as browsing or video.

A final study [43] as a comparison between traditional TV services, streaming services, and live video platforms. Interesting in this study analysis is the inclusion of gaming content, which evaluates the effects on QoE.

2.4. AI solutions

The use of intelligence algorithms applied thanks to an MEC environment is the premise of [44], specifically the use of DRL algorithms to optimize computing tasks and improve general load. The consumption of bandwidth can be one of the critical points in these services. In [45] they proposed cloud gaming techniques to compress the streaming bitrate with few perceivable differences.

The future of networks with innovations beyond 5G will allow services with high bandwidth consumption. 360-degree immersive content, such as the one featured in [46], where they proposed to further push 2D content for fully immersive content. There are cases that seek to optimize the video based on a previously generated dataset. This case would be an AI application based on supervised training [47]. Although it showed optimal results, since they considered additional data from the network, these models were not redundant in the face of possible changes in the network or additional unforeseen events that may arise. That is, they based the training on previously collected data, but in the face of new inconveniences, they did not have enough information to correctly control the bitrate and so, maximize the QoE.

The applications discussed previously require high bandwidth consumption. The idea is to be able to anticipate possible inconveniences in the networks, with the utilization of the DRL algorithms necessary

for an active participation in the service. In [21], they tackled this problem by addressing the concept of dynamic media adaptation using RL algorithms with Markov decision processes. This can be seen as an initial approach of dynamic optimization by using the first DRL techniques. Works like [19], used DRL techniques with QoE, but in this case approaching the multimedia traffic control. They considered directly network characteristics such as jitter, bandwidth, and latency to decide the optimal path for the transmission.

More recent work [48] integrated the QoE of streaming in the algorithm, seeking to reflect this service improvement in shorter buffer times. Finally, another point of view is observed in the work [49], developing an adaptive video framework, both at the bitrate and the encoding level.

Both works did not directly address the analysis of QoE, but rather an estimation of it by measuring stalls in the video experience and the necessary rebuffering processes. Instead, the presented article continues the work started in [18], with an optimization of the QoE on a multimedia scenario. The great difference and advance is the inclusion of different distant sites processed in real time, allowed due to the asynchronous behavior of the algorithm applied in this work.

3. Methodology

The methodology section of this research article encompasses various aspects, including the design of a virtualized platform that applies several 5G network concepts; the utilization of specific components at the application layer; the simulation of workflows to enable the training; and the adoption of the DRL algorithm to optimize the entire service.

3.1. Service virtualization platform architecture

The increasing number of users of gaming streaming services is becoming a challenge for today's networks to provide the required bandwidth and latency, in terms of quality and costs for operators and service providers [50]. In addition, traditional network services need dedicated hardware with specific network functions. This approach is very hardware-centric and lacks flexibility, increasing the operation complexity and interfering with the innovation and business model of the gaming streaming service.

The 5G networks bring highly increased bandwidth and reduced latency, all together with higher reliability, while built upon Software Defined Networking (SDN) and Network Function Virtualization (NFV), they support the resource assignment and improve the overall QoS. These two fast-moving technologies allow the location of changing network infrastructures to be located and where they are most appropriate on the network. Therefore, the proposal idea is an integration of microservices in a general multisite infrastructure, the final service being consumed in different parts of the planet.

Service requirements fluctuate depending on demand. High bandwidth is not always required, or the resources required may vary depending on the load on the other servers in the system. To balance the dimensioning of resource usage and the required performance, Content Delivery Networks (CDNs) are commonly used. CDNs offer dynamic services in a virtual way, with the concept like that exposed in the idea of gaming streaming. In this way, all possible hardware and adaptations to user applications will be virtualized and deployed to provide service anywhere the user is connected. The main expected benefits are reduced costs, improved flexibility, time, and complexity, and the ability to transmit high-quality streaming anywhere, consume content anytime, and adapt the quality to any device.

Powerful CDNs enable game streaming from anywhere to anywhere without the need for dedicated infrastructure or data centers. Users connect via a 5G network to virtual gaming streaming applications deployed at the edge to ensure high performance levels in terms of bandwidth and low latency, so the media application functions will

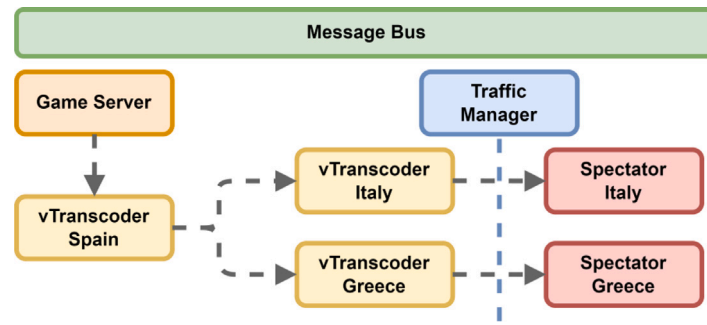


Fig. 1. Multisite workflow architecture.

be deployed close to traffic sources and sinks. Multimedia signals are transported between different sites and are available to all users.

The proposed architecture applies the SDN and NFV concepts to embed workflow applications flexibly and dynamically as Virtual Network Functions (VNFs) over 5G networks and cloud infrastructures. To ensure high performance levels in terms of quality, bandwidth, and low latency, a smart optimizer component takes advantage of the flexibility of a cloud infrastructure, to provide optimization policies both at the network and application levels.

3.2. Application layer

The components developed for the game streaming scenario are open source. They are flexible, scalable, and evolve more easily than traditional networks. Applications will be composed of different Physical Network Functions (PNFs) and VNFs. This deployment in different services allows for better integration in the edge cloud and continuous improvement of each service. That is, each of them is independent and allows for scalability and ease of use without overinvesting. A specific workflow is created that facilitates the automation of tasks, allowing flexibility where and when it is needed, thanks to NFV. In Fig. 1 the architecture developed for the investigation is represented, which included the following components:

- **Game Server:** Core component of the service, acting as a PNF. It supports the greatest load of resources, being a powerful physical machine capable of handling a resource-demanding game, the real-time multimedia transmission process, and the AI-based optimization service. This component hosts a third-person game to explore a 3D environment, based on Unity [51]. This engine is one of the most used thanks to its support in practically all platforms, from video games to mobile platforms. Inside the PNF hardware, it also hosts the Optimizer, which is a smart component that optimizes the behavior of the service in a 4-step loop: Monitoring, Analysis, Policy, and Execution (MAPE). It needs a connection to all components of the service to collect all monitoring data from applications and the network. The optimizer uses DRL algorithms to analyze and provide corrective actions with the aim of obtaining the best QoE. Finally, an instance of the OBS studio [52] will capture the screen of the game server and create the stream through the network with the desired multimedia characteristics. This PNF is hosted on a custom workstation that runs with 64 CPUs, 32 GB of RAM memory, and is equipped with a NVIDIA 2080 Ti GPU.
- **vTranscoder:** VNF service for multimedia handling, capable of adapting input content and transmitting it over the network according to the desired parameters. It is based on open-source encoding techniques and libraries included in the FFmpeg [53] tool suite. The encoder used for image processing was H.264.

Although there are new codecs that improve some features, it is also true that they are not as widespread as H.264. Our scenario required several integrated components under a 5G architecture, so a fundamental requirement was to use a fully integrated codec in all layers of the network. However, this encoding is one of the most efficient in terms of resource consumption with respect to the desired objective QoE and the encoding speed. This step is necessary to reduce the bandwidth of multimedia signal transmission and can be critical in some cases, so it should be considered. Apart from codec features, deploying this VNF on the edge gave flexibility to adapt the stream depending on the user bandwidth or a specific device. This flexibility means that it was not connected to the central graphics card and was processed according to the central resources where the VNF is deployed.

- **Spectator Client:** Component responsible for simulating the behavior of an end user and observing the received content, also deployed as a VNF. Therefore, we included a multiplatform framework system, such as VLC to watch the content and a quality probe to assess the received multimedia. This probe was a video quality indicator service based on work by [54–56], which performed an analysis to determine image characteristics. This probe took the stream flow as input, managing to analyze different indicators of the image based on QoE. Some of these indicators, being used later as optimizer states, were blockiness, block loss, temporal activity, spatial activity, and blur.

The optimizer included within the Game Server required information in real time to assess the state of the environment. For communication between applications, an Apache Kafka system was included at the platform level [57]. This message bus allowed the publication of monitoring metrics from all components of the service. The optimizer was subscribed to the bus of both the vTranscoder and the spectator, these metrics being essential for the optimizer to analyze the current state of the service.

3.3. Workflow simulation

The environment characterization was established by leveraging diverse locations accessible through the 5G EVE multisite platform [58, 59]. The study involved remote settings apart from Spain, like Italy and Greece. In Madrid (Spain), the central site hosted the PNF server, capable with graphic power capabilities to maintain a consistent 60 fps in the game and facilitate video transmission. Streaming began from this site to the central vTranscoder server, which, in turn, relayed the content to distant vTranscoder sites situated in Torino (Italy) and Athens (Greece).

The orchestration of the multisite environment was automated using the 5G EVE platform [60]. This enabled deploying and initiating components across various transcoders and client-side elements

for streaming analysis. Multimedia metric generation began when the gamer started generating content, allowing the algorithm to execute bitrate corrective actions.

Training occurred with a focus on two vTranscoder locations (Italy and Greece), aiming to streamline resource assimilation and foster seamless integration between these sites. This integration played a crucial role, particularly in the context of the 5G nature of the use case, demanding a consistent bandwidth flow during transmission. By comprehensively characterizing the scenario, we ensured that, following the completion of training, the algorithm could seamlessly extend its reach to any additional sites, if they met the minimum requirements for bandwidth and resources. The developed algorithm is parallelizable, so as long as the additional site covered the minimum requirements, considering the actions of the model (maximum bitrate of the profiles), this new site could be included.

A robust model requires the inclusion of network impairments within the training plan to preemptively address potential challenges once the model is deployed. Executing this task involved the development of the Traffic Manager (TM), which operated on the spectator's network interfaces, fine-tuning network parameters. This component served as the gateway to the spectator's received stream. Concurrently, alongside the TM, another component, the Traffic Generator (TG), was introduced to impose contingencies on the video stream. Essentially, we aimed to simulate additional users or services connected to the network, consuming the available bandwidth. In this instance, it involved another transcoder transmitting video with randomly adjusted bitrate, introducing disruptions to the transmission.

The primary objective revolved around delivering an optimization service within a multisite gaming scenario. The multisite environment specifically entailed two distinct locations: the initial site responsible for content generation and transmission, and the second site designated for reception and evaluation. Both sites were positioned at considerable distances, with content transmission occurring over a 5G cloud telecommunication network (refer to Fig. 1).

3.4. Performance evaluation metrics

In our assessment of the effectiveness of the proposed multisite gaming streaming optimization, we employ a set of key performance indicators (KPIs). These values are obtained to determine the state of different characteristics of the service, such as evaluating the quality of the video especially. Additionally, useful parameters are the streaming configuration values such as the assigned GOP, and the video bitrate, as well as the resources used on the infrastructure that can help the algorithm determine possible transmission losses. The subsequent itemized list elaborates on these performance indicators.

Video Quality Metrics:

- Blockiness [0, 1.1]. Computes the presence of block-like artifacts in a video, often caused by compression or lower bitrates, negatively impacting visual quality. A blockiness value between 0.9 and 1 indicates good quality without distortion. Acceptable quality is maintained from 0.65.
- Block loss [0, 200]. Represents the video blocks lost during transmission or compression, affecting the overall integrity of the video stream. Higher values imply more visible distortion, with good quality maintained in the range of 0 to 5.
- Blur [0, 70]. Measures the degree of blurriness or lack of sharpness in video frames, influenced by compression artifacts, focus issues, or motion. Higher values correspond to more visible distortion, while good quality is upheld within the range of 0 to 5.
- Temporal Activity [0, 255]. Describes the level of motion or activity across consecutive frames. Higher values signify greater temporal activity, with good quality within the range of 0 to 20.

- Spatial Activity [0, 270]. Reflects the complexity of spatial details within a video frame. Higher values indicate greater spatial activity, with good quality maintained within the range of 0 to 60.

Streaming Parameters:

- GOP configuration [0.5, 4]. Involves structuring GOP in a video stream, impacting compression efficiency and overall quality.
- Bitrate [5, 50]. Represents the amount of data processed per unit of time in a video stream, directly influencing compression levels and video quality.
- Encoding quality. Indicates the efficiency and effectiveness of the video encoding process, with higher quality yielding superior visual fidelity, albeit at a potential increase in computational demands.

Infrastructure Resource Utilization:

- CPU consumption. Measures the processing power utilized by the system to handle specifically video streaming tasks.
- RAM usage. Quantifies the amount of memory utilized by the streaming infrastructure.

3.5. Asynchronous advantage actor-critic (A3C)

RL algorithms learn through trial and error against a defined environment. This concept of trial and error is evaluated through rewards, either positive if the outputs are optimal or negative if the outcome penalizes the sequence of actions. The goal is to maximize the total reward in the long run. That is, at the cost of obtaining some minor rewards for a certain option, we must choose to make decisions at certain times so that we can find new avenues of rewards that increase the total accumulated [61]. This means that despite seemingly very positive decisions, there may be others that can offer us greater total reward with successive actions.

All RL algorithms are based on the Bellman equations [62], which by following some functions look to maximize the optimal policy (π) function. They are made up of two differentiated equations. One is value-based, through which we try to find the best decision through an approximation of the reward values according to the values of the shares. This formula computes the state value function (V_π) as the expected return over all possible actions for a specific state. In contrast, it is the policy-based type, which considers the set of states and actions, to decide the action, computing the state-action value function (Q_π), as the expected return over all possible actions in some state. Formulas are expressed in Eq. (1). The equations involve several parameters, including the expected value of the policy distribution (\mathbb{E}_π), the current state (s) and the set of possible states (S), the current action (a) and the set of possible actions (A), the accumulated reward (R_t), the returned reward for a specific trajectory (G_t), and the discount factor (γ).

$$\begin{aligned} V_\pi(s) &= \mathbb{E}_\pi[R_t + \gamma G_{t+1} | S_t = s] \\ &= \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[r + \gamma v_\pi(s')] \end{aligned} \quad (1)$$

$$\begin{aligned} Q_\pi(s, a) &= \mathbb{E}_\pi[R_t + \gamma G_{t+1} | S_t = s, A_t = a] \\ &= \sum_{s',r} p(s',r|s,a)[r + \gamma \sum_a \pi(a|s) Q_\pi(s', a')] \end{aligned}$$

Each of the RL types mentioned above has advantages and disadvantages. From the speed of training to the stability and convergence of the model. One of the ideas that emerged was to be able to merge both ideas, having the best characteristics of each one and alleviating the problems. In this case, one of the algorithms that emerged was based on the concept of Actor-Critic [63].

These algorithms divided RL problems and capabilities into two parts. The actor oversees determining the best possible action considering the states, looking for the best action policy (policy-based). The

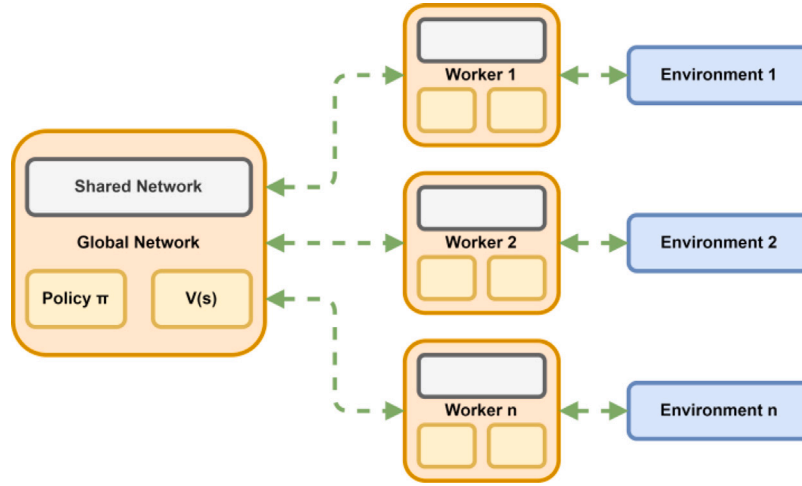


Fig. 2. A3C general overview.

critic evaluates the action taken by the actor, computing the value function (value-based). As iterations progress, these two components learn and improve their approximations.

Among all the action values, the agent must choose the best one to obtain the most optimal consecutive actions, possibly causing high variability in the results. The advantage function was introduced in the A3C algorithms to estimate how much better a particular action is compared to the expected average reward for that state, given the current policy (see Eq. (2)). It was defined as the difference between the Q value of taking a specific action (a) in a particular state (s) and the average Q value of that state ($V(s)$).

$$\begin{aligned} A(s, a) &= Q(s, a) - V(s) \\ &= r + \gamma * V(s') - V(s) \end{aligned} \quad (2)$$

This improvement over previous algorithms helped the agent differentiate between the effectiveness of different actions in the same state, allowing it to take more optimal actions. In A3C algorithms, the advantage function is used as a baseline for computing the policy gradient, which is then used to update the value and policy function.

The A3C algorithm has many advantages over conventional RL algorithms, due to either the combination of techniques or the use of neural networks. The use of separate networks (actor and critic) for the training also provides high stability, avoiding problems related to the high dynamic gradient range. However, the specific choice of this algorithm for the use case was due to its parallel training nature. Its previous version, called Advantage Actor–Critic (A2C), marked the beginning of parallelizable algorithms. However, the training occurred synchronously. In other words, the agent feedback to the central server was carried out in a coordinated manner, so that if there were any interruptions in an agent or possible delays, general interruptions in training could be experienced. The next version, called Asynchronous Advantage Actor–Critic (A3C) (Fig. 2 and explained pseudocode in Algorithm 1), solved this problem with the asynchronous sharing of network weights. This asynchronous version allows for an AI model with multiple sites in parallel. In other words, our multisite environment can be fully developed without the need for synchronization between different sites, each being independent. The limitation of parallel sites depends on the number of resources deployed.

Algorithm 1 Asynchronous Advantage Actor–Critic (A3C).

Input: Number of parallel workers $N = 2$, learning rate $\alpha = 0.001$, discount factor $\gamma = 0.99$, policy parameter θ , value function v , episodes per step $T = 10$

Initialize global shared neural network with parameters θ

for each site worker i in N

 Initialize the policy network π_i with parameters θ

 Initialize the value network V_i

 Initialize the random state

s_0 =[video quality, streaming, infrastructure]

for each episode t in T

 Execute action a_t (set bitrate) following policy π_i

 Observe discounted reward R_t

 Compute the advantage function $A_t = R_t - V_i(s_t)$

 Update actor network parameters θ_i

end for

 Update the global shared neural network (critic network)

end for

Update shared parameters θ

Output: Updated policy parameter θ , value function v

4. Experiments

4.1. Experimental setup

The optimizer component, placed on the Game Server PNF, was designed to consume all the monitoring data from the Kafka message bus, analyze it, and provide an optimization policy for actions to the Site Transcoder VNF. In the proposed case, optimization was based on an adaptation of the bitrate of the content in the network to solve potential quality losses due to possible network bottlenecks.

A real-time scenario requires a service capable of responding with actions with the available information. Generally, AI models capable of predicting values need a previous training dataset to find internal patterns in the data. In these cases, these data were transmitted in real time by both the vTranscoder and the vProbe, so the optimizer depended exclusively on the data in the instantaneous time to be carried out.



Fig. 3. Frame extracted from the video sequence in original quality.

Given the problem of lack of a dataset in a very heterogeneous environment and the need for a very flexible solution capable of adapting to changes in the topology status, the choice for the optimization algorithm was a combination of DL and RL called DRL.

The full integration of the use case covers the connection between PNF - VNF - VNF (Game Server - vTranscoder - vProbe) so that the validation of the complete system consists of a simultaneous execution of all the components to enable the optimization task, with the consumption of results on the enabled Kafka bus.

4.2. Gaming streaming content

In this use case, the gamer transmitted the game content through the network, so that the end user can enjoy the video in the best possible quality, regardless of the network conditions. The video game on which the transmission was to be made consisted of a 3D 1080p (1920×1080) adventure game (see Fig. 3 for reference image). The concrete aspect is cartoon, but it contained a diversity of scenarios, whether it was an open world with contrasts of lights or caves in which the content was complex in cases of low bitrate.

One of the mantras in the gaming world is to guarantee a minimum of 60 stable FPS, in contrast to the 24 or 30 FPS of the traditional TV industry. On the game server, the game runs at a higher frame rate (variable +100 FPS); however, to ensure optimal transmission, it was limited to 60 FPS to meet the minimum threshold for gaming transmission. Given the increase in the rate of frames per second compared to traditional content, an increase in bitrate was expected. Through a previous analysis, we determined that the bitrate of the gameplay was around 25 Mbps, with some frequent spikes at 30 Mbps. This content was already encoded according to the H.264 codec.

The video used for the use case consisted of a video sequence from a game developed under the Unity Engine. A 10-minute sequence was recorded to make an estimate on the video analysis, considering statistics to develop an optimization. Therefore, the duration of the video (even so, the video is played in a loop) is not so important for the validation but the overall effectiveness of the components for generating information.

Within this analysis, five characteristics were extracted: blockiness, block loss, blur, temporal activity, and spatial activity. Among the five features, blockiness and block loss stood out as better indicators of image behavior. In particular, a blockiness value of 0.65 or higher is indicative of a desirable high QoE. On the other hand, a block loss value greater than 5.0 signifies significant image degradation and noticeable losses in quality. The duration of the recording was not fixed in advance, but rather consisted of the complete flow of the video game, venturing the character through the different areas of the scenarios to be able to analyze most of the graphics in the game. In other words, the 10-minute sequence contained a complete round for the completion

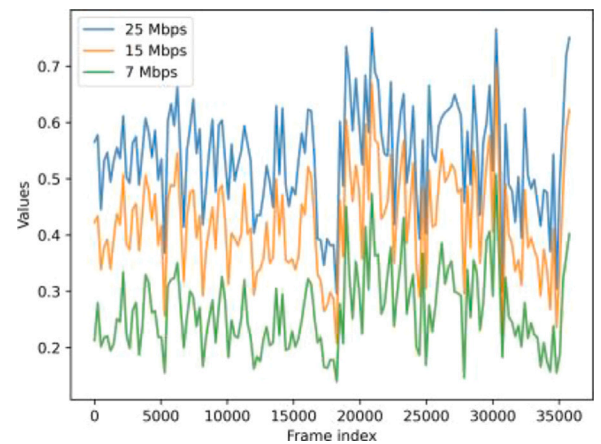


Fig. 4. Blockiness performance throughout the entire video sequence.

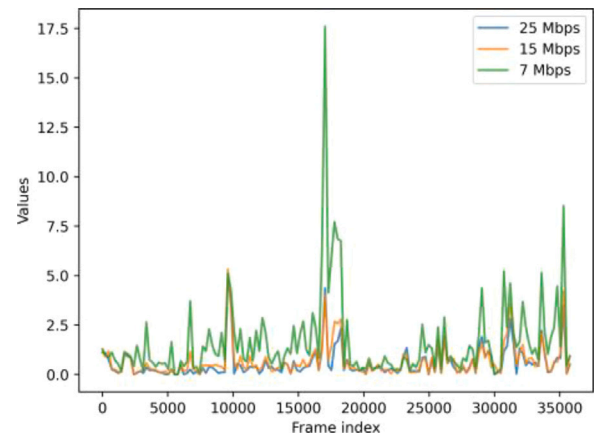


Fig. 5. Block loss performance throughout the entire video sequence.

of the game, from start to finish. In Figs. 4 and 5 the two main characteristics, blockiness and block loss, are examined throughout the training video sequence. These figures showed the oscillation values for different bitrate profiles, both for blockiness and block loss.

Such high transmission rates may not be supported in certain environments, so a previous content analysis was performed to evaluate the appropriate network profiles. These network profiles were the target bitrate at which the transcoder had to adjust the streaming bitrate. Analyzing the content (Fig. 6), it was observed that by limiting the gameplay to 25 Mbps, there were hardly any distortions in the image,

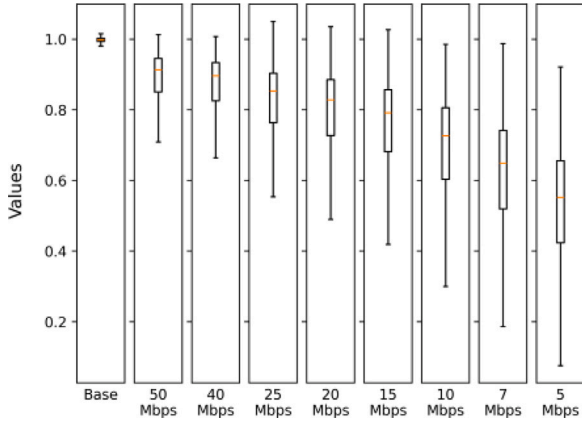


Fig. 6. Blockiness analysis performed over different bitrates.

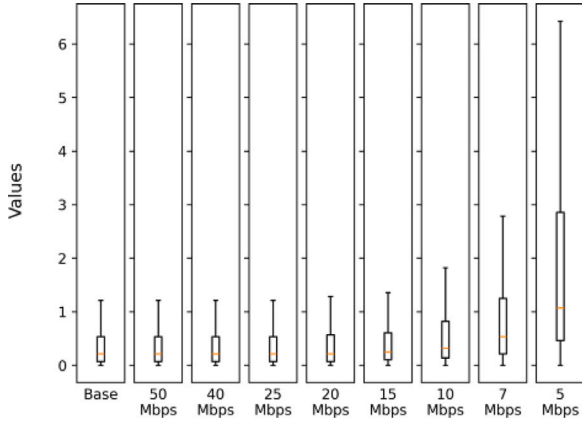


Fig. 7. Block loss analysis performed over different bitrates.

so this was considered the upper limit of what we would consider a high-quality profile. On the other hand (Fig. 7), progressively reducing the bitrate we observed that from 10 Mbps downward, we began to observe artifacts in the image. At 7 Mbps the lower profile was set, where we considered that going further, we could not guarantee decent quality in the image. Finally, an intermediate point was established for the medium profile, establishing at 15 Mbps.

Another aspect addressed in the image analysis before training was the size of the Group of Pictures (GOP). The GOP is the distance between two main frames, counting as distance the number of frames that separate both. In the first instance, we analyzed four GOP sizes, so due to our 60 FPS approach we had a GOP size of 30-60-120-240 frames. At the time of deployment, we observed that the transcoder, when placed at a point where it did not have a graphics card, was not able to maintain a smooth transmission with the largest GOP size (4). Reducing GOP sizes, we obtained somewhat higher quality measures, but totally negligible compared to a reduction in bitrate (Fig. 8). It was expected that a reduction in the GOP could affect the compression quality, so we saw a slight increase in the video bitrate but that we adjusted according to the network profiles mentioned above. Finally, the transmission parameters were established.

4.3. Optimization

The optimization of the streaming service was achieved by training the A3C algorithm on a target bitrate, representing the desired flow of streaming content. RL algorithms, such as A3C, operate through an action policy, defining a set of possible movements or decisions. In our case, these actions corresponded to network profiles associated with

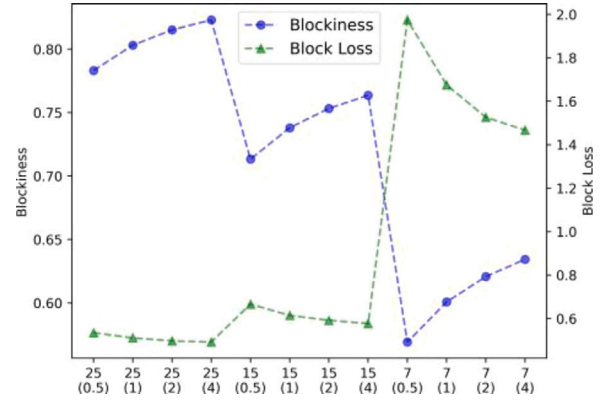


Fig. 8. Comparison metrics for different GOP sizes in network profiles.



Fig. 9. Neural Network architecture.

target bitrates that the transcoder aimed to reflect in the streaming. This set of profiles were defined for high (25 Mbps), medium (15 Mbps), and low (7 Mbps) quality streaming.

$$\text{Value Loss} = (A(s, a))^2$$

$$\text{Policy loss} = -\log(\pi(a|s)) * A(s, a)$$

(3)

In the A3C algorithm, a combination of RL and DL was employed to learn the optimal policy for the streaming environment. We established a standard architecture for the two neural networks within the A3C algorithm (see Fig. 9). The Actor neural network, responsible for selecting actions, comprised several fully connected layers (dense layers) followed by a SoftMax function, generating a probability distribution over possible actions. This distribution informed the action selection process, facilitating the decision-making in the streaming optimization task. The value loss was determined by the squared value of the advantage function.

Simultaneously, the Critic neural network, sharing a similar architecture to the Actor network, computed the policy loss. The policy loss served as a measure of the disparity between predicted and actual action distributions. The total loss produced was used to update the parameters and weights of the shared network, in order to improve the general policy. Both policies are presented on Eq. (3).

In detail, the optimizer needed values from the current state of the environment. On the one hand, we had the states from the vTranscoder, which offered the transmitted bitrate values that were approximated to the target bitrate, the encoding quality, CPU consumption, and RAM usage. On the other hand, in the component considered as a spectator, video analysis was carried out through the video quality probe, which extracted measures of blockiness, block loss, blur, temporal activity, and spatial activity. The values were consumed by the optimizer via the REST API in the vTranscoder and via Kafka in the probe. All set of states are summarized in Table 1.

Finally, the algorithm evaluated each action using reward functions. These functions, positively or negatively, represented the extent to which the actions had been taken. The main reward function was based on the analysis of the content received in the spectator, since the objective was to ensure that the user experience was maximized. However, a reward function was additionally included to emphasize the highest network profiles. This was because on certain occasions the RL

Table 1
Set of states and actions.

	Features
States	Actual bitrate; Target bitrate; Encoding Quality; CPU usage; RAM usage; Blockiness; Spatial Activity; Block Loss; Blur; Temporal Activity
Actions	25 (2); 15 (1); 7 (0) Mbps (Profile)

algorithms could become conservative, and despite looking for the best quality, we also wanted to be able to transmit at the highest possible bitrate, keeping close the bitrate to the original content.

Even though the RL algorithm generates actions quickly, effective choices require a deeper understanding of their consequences. We prioritized analysis over immediate reactions by implementing a 10-second delay. This time was adopted based on the video quality probe which required around 4–8 s to analyze the sequences. In addition, we included additional time to ensure that the process was completed and the environment was adapted to the new action, offering accurate data.

5. Results

The transmission of the video was carried out over UDP, since we guarantee fluid streaming with the aim that an optimization service guarantees the QoE of the video. Being a gaming environment, the frame rate we guarantee was 60 FPS, so the initial analysis of the video was carried out for different GOP sizes. We looked for standard GOP sizes in the broadcast industry, so efforts were focused on analyzing image measurements based on these parameters to decide the best suitable GOP.

The conclusions drawn are that the GOP of size 2 (and, therefore, 120 frames) was the one that met our requirements. Size 4 was also considered, but with such a large frame size (240), transmission problems were observed, as it was a long time, causing continuous disconnections in streaming. The rest of the GOPs also met the requirements, but since there were no great differences between them, we decided to maintain the standard in the industry.

The UDP transmission configuration was set under ultra-fast speed encoding. Sending via UDP could have some negative effects on the quality of the image, but, since we have a high binary rate, we did not perceive impacts to highlight them. However, the effects produced by the speed of stream encoding were more notable. The important part of the transmission was to be able to maintain a stable flow of 60 FPS. As the output video was generated at a high bitrate, it was decided to select a higher preset to get a fluid transmission.

The original video used had a bitrate of 376 Mbps (28.2 GB in total size), so for a correct transmission, the bitrate was reduced by means of coding techniques. This bitrate must not only be optimal so that the components had support for this amount, but we had also considered the limits imposed between network links so as not to saturate the link with a single service. Using an ultrafast preset over UDP, the bitrate was increased to 473 Mbps (35.4 GB). We can observe that when transmitting with the settings mentioned above, the total size was increased while still preserving a good level of quality. However, in terms of quality metrics, the effects can be observed in Figs. 10 and 11. Although these effects are permissive in terms of quality, they created differences for each network profile. They allowed us to create a baseline for different profiles for the use case optimization.

Once the stream characteristics were defined, we focused on the optimizer task to maximize the spectator's QoE. One of the main inconveniences that can occur in a transmission is loss of packets or seeing bad blocks in the image. An example of them can be seen in Fig. 12, where its quality has been seriously reduced by both a reduction in bitrate and external agents, obtaining a notable degradation in QoE. It

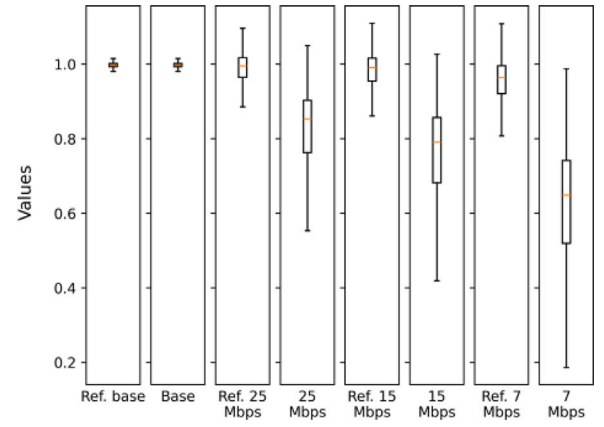


Fig. 10. Comparison of blockiness between video reference and transmission.

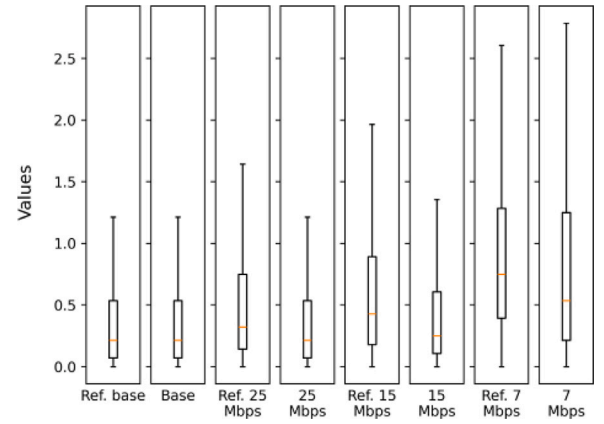


Fig. 11. Comparison of block loss between video reference and transmission.

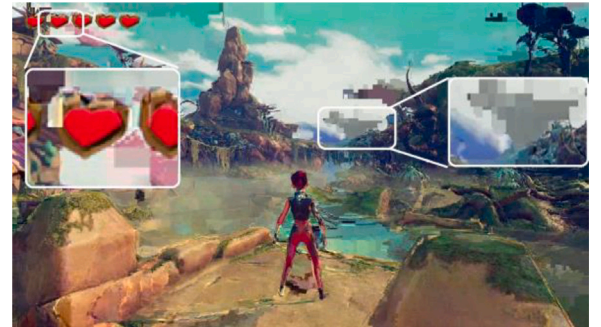


Fig. 12. Frame extracted from the video sequence with impairments.

is presented as an example of a bad QoE on the transmission, which the optimizer must avoid for a proper training.

The evaluation of the training was carried out using the set of functions developed. This set was a combination of small rewards which defined specific parts of the use case evaluation. Due to the previous analysis, we observed that blockiness and block loss were the two fundamental aspects that allowed us to differentiate the reception of high-quality content from low-quality or even lossy content. The first two equations represented the evaluation of the video quality based on the obtained values of blockiness and block loss. These developed rewards were based on activations of popular neural networks functions such as \tanh and inverted $ReLU$. In Eq. (4) the rewards used for training

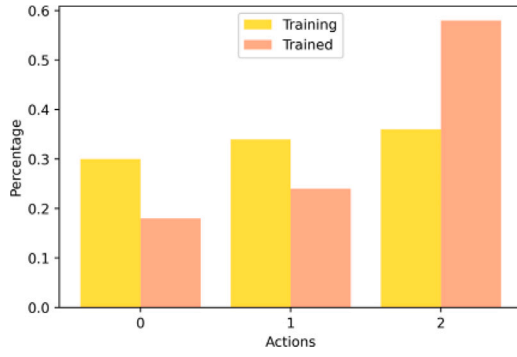


Fig. 13. Optimizer action distribution.

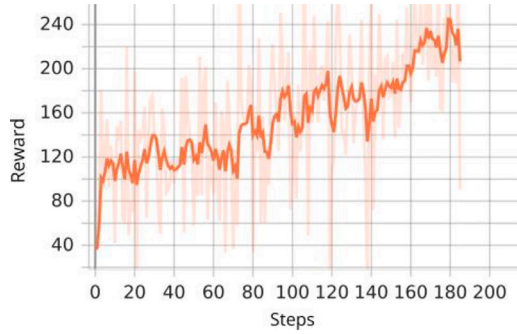


Fig. 14. Reward values during training.

are presented.

$$\begin{aligned}
 R_{\text{blockiness}} &= 110 \cdot \tanh(\text{blockiness} - 0.55) && \in [-46, 46] \\
 R_{\text{bl}} &= -\min(\max(0, \text{bl} - 4), 20) && \in [-20, 0] \\
 R_{\text{profile}} &= 6 \cdot \text{action} && \in [0, 12] \\
 R_{\text{total}} &= R_{\text{blockiness}} + R_{\text{bl}} + R_{\text{profile}} && \in [-66, 58]
 \end{aligned} \quad (4)$$

The third rewards equation corresponds to balance the training. In other words, Reinforcement Learning algorithms, despite having exploration patterns beyond the action considered most sensible, can become conservative over time. Therefore, an additional reward attached to the selected profiles was developed, with the highest value corresponding to the profile with the highest bitrate, as opposed to no reward returned for the lowest profile.

This exploration ability can be seen in Fig. 13, where during training the actions were balanced throughout the set. However, with the optimizer enabled, there was an emphasis in the action for high bitrate. The rest of actions were also selected in cases of inadequate network situations or if the image presented some distortions.

Rewards were defined to evaluate each of the optimizer episodes. They represented not only the quality of audiovisual multimedia but also a reference to balance training. In Fig. 14 we show the evolution of the reward during 185 steps. Each included step involves an accumulation of the rewards obtained over several episodes, that is, obtaining the current state of the environment and acting based on the bitrate profiles. The rewards obtained during the initial training phases were around 40. In the final steps, this reward was increased to 240, obtaining a 500% outstanding difference increase. The total duration of training was approximately 5:30 h, the point at which the optimizer could start overfitting the training data and incur future errors predictions.

A crucial aspect of training with this algorithm involved proactively addressing potential challenges, such as network bottlenecks. To simulate real-world scenarios, the background bitrate generation was randomized throughout the training process. Table 2 represents

Table 2

Comparison of Mean Optimizer results.

Features	Training	Trained
Blockiness (% ≥ 0.65)	0.68 (56%)	0.81 (76%)
Block loss (% ≤ 5.0)	15.5 (80%)	2.3 (95%)
Bitrate	12	19

the results obtained, comparing on average the value of blockiness, block loss and bitrate obtained, in the training phase compared to the model trained and applied to the environment. As explained in previous sections, for a suitable high QoE, the appropriate blockiness values are 0.65 and above. Furthermore, block loss values below 5.0 are considered optimal, while values exceeding this threshold indicate significant losses in the image. Upon comparison of the results, there is a noticeable enhancement of 20 percentage points in blockiness and a concurrent improvement of 15 percentage points in block loss, accompanied by a gain of approximately 7 Mbps in bitrate.

6. Conclusions

The result of this publication exposes the ability to simultaneously deploy a multimedia optimization service. This deployment achieves a great goal in cooperation for the instantiation of services in distant sites. The application of AI and, more specifically, the A3C algorithm has allowed this parallelization of environments to perform their services independently. Not only that, but with the results obtained, we demonstrated how we were able to minimize content losses (15% reduction), as well as the fact of obtaining higher image quality results (20% increase in blockiness) with higher bitrates (+7 Mbps), compared to a service without an optimizer integrated in the system.

An objective to achieve in future work would be to analyze new DRL algorithms, such as Soft Actor-Critic algorithms [64], which adds the use of Experience Replay to the combination of algorithms, with which to stabilize training even more. This would open doors to training in shorter periods, but it would be necessary to test its effectiveness in scalable environments.

Finally, the ability to optimize in parallel different streams opens a new research field, where we can consider different privileges per user (premium user, standard user, etc.). This consideration also opens the door to research from a business perspective, where we can apply different billing models depending on the consideration of the user and the optimization of the workflow.

Another consideration is that the optimization carried out in this work is focused on the application layer of a virtualized service. The use of topology discovery techniques for including lower level parameters and actions in the optimization procedures could lead to a more real and complete scenario, where several abstraction layers are considered for a fully optimization, not only of the service end QoE, but also other domains such as resource consumption and energy efficiency.

All of this future work is planned to be integrated with research and innovation within topics of B5G and 6G by adding intelligence to the management and optimization of the network.

CRedit authorship contribution statement

Alberto del Rio: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Javier Serrano:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **David Jimenez:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Formal analysis. **Luis M. Contreras:** Writing – review & editing, Resources, Investigation, Conceptualization. **Federico Alvarez:** Writing – review & editing, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Alberto del Rio reports financial support was provided by Horizon Europe CODECO and NEMO projects, and UNICO-5G I+D B5GEMINI-AIUC project.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgments

This work has been partially funded by the following projects: Horizon Europe CODECO project, grant number 101092696; Horizon Europe NEMO project, grant number 101070118; and by UNICO-5G I+D TSI063000-2021-79 (B5GEMINI-AIUC) project funded by the Ministry of Economic Affairs and Digital Transformation of the Spanish Government and the NextGenerationEU (Recovery, Transformation and Resilience Plan - PRTR).

References

- [1] Mordor intelligence, gaming market - Growth, trends, covid-19 impact, and forecasts, 2023 - 2028), 2020, <https://www.mordorintelligence.com/industry-reports/global-gaming-market>. (Accessed 22 February 2023).
- [2] Ericsson, 5G to Top One Billion Subscriptions in 2022 and 4.4 Billion in 2027, Ericsson Mobility Report, 2022, <https://www.ericsson.com/en/press-releases/2022/6/ericsson-mobility-report-5g-to-top-one-billion-subscriptions-in-2022-and-4.4-billion-in-2027>. (Accessed 22 February 2023).
- [3] Ericsson, Global 5G Growth Amid Macroeconomic Challenges, Ericsson Mobility Report, 2022, <https://www.ericsson.com/en/press-releases/2022/11/ericsson-mobility-report-global-5g-growth-amid-macroeconomic-challenges>. (Accessed 22 February 2023).
- [4] O. Wallach, The world's most used apps, by downstream traffic, 2021, <https://www.visualcapitalist.com/the-worlds-most-used-apps-by-downstream-traffic/>. (Accessed 22 February 2023).
- [5] A. Peša, D. Čičin-Šain, New business model in the growing E-sports industry. Poslovna izvrsnost, Bus. Excell. 11 (2017) 121–131, <http://dx.doi.org/10.22598/pi-be/2017.11.2.121>.
- [6] M. Iqbal, Twitch revenue and usage statistics, 2021, <https://www.businessofapps.com/data/twitch-statistics/>. (Accessed 22 February 2023).
- [7] F. Alvarez, D. Breitgand, D. Griffin, P. Andriani, S. Rizou, N. Zioulis, F. Moscatelli, J. Serrano Romero, M. Keltsch, T. Panagiotis, U. Acar, O. Prieto, F. Iadanza, G. Carozzo, H. Koumaras, D. Zarpalas, D. Jimenez Bermejo, An edge-to-cloud virtualized multimedia service platform for 5G networks, IEEE Trans. Broadcast. 65 (2019) 369–380, <http://dx.doi.org/10.1109/TBC.2019.2901400>.
- [8] J. Garcia-Reinoso, M. Gupta, M. Rosello, E. Kosmatos, G. Landi, G. Bernini, R. Legouable, L. Contreras, M. Lorenzo, K. Trichias, The 5G EVE multi-site experimental architecture and experimentation workflow, in: IEEE 2nd 5G World Forum, 5GWF, 2019, pp. 335–340, <http://dx.doi.org/10.1109/5GWF.2019.8911624>.
- [9] M. Ananth, R. Sharma, Cloud management using network function virtualization to reduce capex and OPEX, in: 8th International Conference on Computational Intelligence and Communication Networks, CICN, 2016, pp. 43–47, <http://dx.doi.org/10.1109/CICN.2016.17>.
- [10] S. Anshel, 5G and AI: Complementary technologies now and into the future, 2020, Forbes. <https://www.forbes.com/sites/moorinsights/2020/11/25/5g-and-ai-complementary-technologies-now-and-into-the-future/>. (Accessed 22 February 2023).
- [11] Y. Xiao, G. Shi, Y. Li, W. Saad, H.V. Poor, Toward self-learning edge intelligence in 6G, IEEE Commun. Mag. 58 (2020) 34–40, <http://dx.doi.org/10.1109/MCOM.001.2000388>.
- [12] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, K. Wu, Artificial intelligence-enabled intelligent 6G networks, IEEE Netw. 34 (2020) 272–280, <http://dx.doi.org/10.1109/MNET.011.2000195>.
- [13] Y. Zhou, L. Ling, L. Wang, N. Hui, X. Cui, J. Wu, Y. Peng, Y. Qi, C. Xing, Service-aware 6G: An intelligent and open network based on the convergence of communication, computing and caching, Digit. Commun. Netw. (2020) 253–260, <http://dx.doi.org/10.1016/j.dcan.2020.05.003>.
- [14] NetWorld2020, European Technology Platform, Smart Networks In the Context of NGI, 2018, <https://www.networld2020.eu/wp-content/uploads/2018/11/networld2020-5gia-sria-version-2.0.pdf>. (Accessed 22 February 2023).
- [15] Why is the demand for cloud services growing rapidly?, 2021, <https://www.syntax.com/en-eu/blog/accelerated-growth-in-the-demand-for-cloud-services/>. (Accessed 22 February 2023).
- [16] Cloud computing market to hit USD 791.48 billion by 2028, 2021, <https://www.globenewswire.com/news-release/2021/08/11/2278451/0/en/Cloud-Computing-Market-to-Hit-USD-791-48-Billion-by-2028-Rising-Demand-for-Improved-Virtual-Access-to-Information-among-Industries-to-Foster-Steady-Growth-Fortune-Business-Insights.html>. (Accessed 22 February 2023).
- [17] J. Serrano, A. del Río, W. Nakimuli, D. Jiménez, J. Garcia-Reinoso, L.M. Contreras, F. Alvarez, Design, implementation, and validation of a multi-site gaming streaming service over a 5G-enabled platform, IEEE Trans. Broadcast. 68 (2) (2022) 464–474, <http://dx.doi.org/10.1109/TBC.2022.3141615>.
- [18] A. del Río, J. Serrano, D. Jimenez, L.M. Contreras, F. Alvarez, A deep reinforcement learning quality optimization framework for multimedia streaming over 5G networks, J. Appl. Sci. 12 (20) (2022) 2076–3417, <http://dx.doi.org/10.3390/app122010343>.
- [19] X. Huang, T. Yuan, G. Qiao, Y. Ren, Deep reinforcement learning for multimedia traffic control in software defined networking, IEEE Netw. 32 (6) (2018) 35–41, <http://dx.doi.org/10.1109/MNET.2018.1800097>.
- [20] A. Al-Jawad, I.-S. Comşa, P. Shah, O. Gemikonakli, R. Trestian, An innovative reinforcement learning-based framework for quality of service provisioning over multimedia-based SDN environments, IEEE Trans. Broadcast. 67 (4) (2021) 851–867, <http://dx.doi.org/10.1109/TBC.2021.3099728>.
- [21] V. Charvillat, R. Grigoraş, Reinforcement learning for dynamic multimedia adaptation, J. Netw. Comput. Appl. 30 (3) (2007) 1034–1058, <http://dx.doi.org/10.1016/j.jnca.2005.12.010>.
- [22] A. Kherani, G. Shukla, S. Sanadhy, N. Vasudev, M. Ahmed, A. Patel, R. Mehrotra, B. Lall, H. Saran, M. Vutukuru, A. Singh, S. Seshasayee, V. Viswakumar, K. Loganathan, Development of MEC system for indigenous 5G test-bed, in: International Conference on COMmunication Systems & NETworkS, COMSNETS, 2021, pp. 131–133, <http://dx.doi.org/10.1109/COMSNETS51098.2021.9352907>.
- [23] N. Slamnik-Kriještorac, G.M. Yilma, F. Yousaf, M. Liebsch, J. Marquez-Barja, Multi-domain MEC orchestration platform for enhanced back situation awareness, in: IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops, No. PS51825.2021.9484632, INFOCOM WKSHPS, 2021, pp. 1–2, <http://dx.doi.org/10.1109/INFOCOMWKSHPS51825.2021.9484632>.
- [24] S. Zhou, P. Netalka, Y. Chang, The MEC-Based Architecture Design for Low-Latency and Fast Hand-Off Vehicular Networking, in: 2018 IEEE 88th Vehicular Technology Conference, VTC-Fall, 2018, pp. 1–7, <http://dx.doi.org/10.1109/VTCFall.2018.8690790>.
- [25] M. Liu, G. Feng, Y. Sun, N. Chen, W. Tan, A network function parallelism-enabled MEC framework for supporting low-latency services, IEEE Trans. Serv. Comput. <http://dx.doi.org/10.1109/tsc.2021.3130247>.
- [26] D. Barboza, D. Muchaluat-Saade, E. Clua, A real-time game streaming optimization technique based on layer caching, in: 12th Annual IEEE Consumer Communications and Networking Conference, CCNC, 2015, pp. 714–719, <http://dx.doi.org/10.1109/CCNC.2015.7158066>.
- [27] International Telecommunication Union (ITU), ITU-R BT.500: Methodologies for the subjective assessment of the quality of television images, 2019, <https://www.itu.int/rec/R-REC-BT.500>. (Accessed 25 May 2023).
- [28] P. Joveluro, H. Malekmohamadi, W. Fernando, A. Kondoz, Perceptual video quality metric for 3D video quality assessment, in: Transmission and Display of 3D Video, 2010, pp. 1–4, <http://dx.doi.org/10.1109/3DTV.2010.5506331>.
- [29] Z. Wang, L. Lu, A. Bovik, Video quality assessment based on structural distortion measurement, Signal Process., Image Commun. 19 (2) (2004) 121–132, <http://dx.doi.org/10.1109/ICIP.2002.1038904>.
- [30] J. Lopez Velasco, D. Martín Gutiérrez, D. Jimenez Bermejo, J.M. Menéndez, Prediction and modeling for no-reference video quality assessment based on machine learning, in: 14th International Conference on Signal Image Technology & Internet Based Systems, SITIS, 2018, pp. 56–63, <http://dx.doi.org/10.1109/SITIS.2018.00019>.
- [31] L. Tiotsop, E. Masala, A. Aldahdooh, G. Wallendaal, M. Barkowsky, Computing quality-of-experience ranges for video quality estimation, in: Eleventh International Conference on Quality of Multimedia Experience, QoMEX, 2019, pp. 1–3, <http://dx.doi.org/10.1109/QoMEX.2019.8743303>.
- [32] T. Minhas, M. Shahid, A. Rossholm, B. Löfström, H. Zepernick, M. Fiedler, Assessment of the rating performance of ITU-t recommended video quality metrics in the context of video freezes, in: 2013 Australasian Telecommunication Networks and Applications Conference, ATNAC, 2013, pp. 207–212, <http://dx.doi.org/10.1109/ATNAC.2013.6705382>.
- [33] U. Reiter, J. Korhonen, Comparing apples and oranges: Subjective quality assessment of streamed video with different types of distortion, in: 2009 International Workshop on Quality of Multimedia Experience, 2009, pp. 127–132, <http://dx.doi.org/10.1109/QoMEX.2009.5246963>.
- [34] A. Moorthy, K. Seshadrinathan, R. Soundararajan, A. Bovik, Wireless video quality assessment: A study of subjective scores and objective algorithms, IEEE Trans. Circuits Syst. Video Technol. 20 (2010) 587–599, <http://dx.doi.org/10.1109/TCSVT.2010.2041829>.
- [35] C. Lee, J. Ok, G. Seo, Objective video quality measurement using embedded VQMs, in: 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, 2014, pp. 129–130, <http://dx.doi.org/10.1109/QSHINE.2014.6928671>.

- [36] N. Barman, E. Jammeh, S. Ghorashi, M. Martini, No-reference video quality estimation based on machine learning for passive gaming video streaming applications, *IEEE Access* 7, 74511–74527, <http://dx.doi.org/10.1109/ACCESS.2019.2920477>.
- [37] J. Nightingale, P. Salva-Garcia, J. Calero, Q. Wang, 5G-QoE: QoE modelling for ultra-HD video streaming in 5G networks, *IEEE Trans.* 64 (2018) 621–634, <http://dx.doi.org/10.1109/TBC.2018.2816786>.
- [38] N. Barman, S. Zadtootaghaj, S. Schmidt, M. Martini, S. Möller, GamingVideoSET: A dataset for gaming video streaming applications, in: 16th Annual Workshop on Network and Systems Support for Games, Vol. 10, No. 1109/NetGames.2018.8463362, NetGames, 2018, pp. 1–6, <http://dx.doi.org/10.1109/NetGames.2018.8463362>.
- [39] International Telecommunication Union (ITU), P.914: Display requirements for 3D video quality assessment, 2016, <https://www.itu.int/rec/T-REC-P.914-201603-I>. (Accessed 25 May 2023).
- [40] International Telecommunication Union (ITU), P.910: Subjective video quality assessment methods for multimedia applications, 2022, <https://www.itu.int/rec/T-REC-P.910-202207-1>. (Accessed 25 May 2023).
- [41] A. Wahab, N. Ahmad, J. Schormans, Variation in QoE of passive gaming video streaming for different packet loss ratios, in: 2020 Twelfth International Conference on Quality of Multimedia Experience, QoMEX, 2020, pp. 1–4, <http://dx.doi.org/10.1109/QoMEX48832.2020.9123071>.
- [42] D. Winter, P. Simoens, L. Deboosere, F. Turck, J. Moreau, B. Dhoedt, P. Demeester, A hybrid thin-client protocol for multimedia streaming and interactive gaming applications, 2006, p. 15, <http://dx.doi.org/10.1145/1378191.1378210>.
- [43] N. Barman, M. Martini, Z. Zadtootaghaj, S. Möller, S. Lee, A comparative quality assessment study for gaming and non-gaming videos, in: 2018 Tenth International Conference on Quality of Multimedia Experience, QoMEX, 2018, pp. 1–6, <http://dx.doi.org/10.1109/QoMEX.2018.8463403>.
- [44] C. Chen, Y. Zhang, Z. Wang, S. Wan, Q. Pei, Distributed computation offloading method based on deep reinforcement learning in ICV, *Appl. Soft Comput.* <http://dx.doi.org/10.1016/j.asoc.2021.107108>.
- [45] L. Lin, X. Liao, G. Tan, H. Jin, X. Yang, W. Zhang, B. Li, LiveRender: A cloud gaming system based on compressed graphics streaming, *IEEE/ACM Trans. Netw.* 24, 2128–2139, <http://dx.doi.org/10.1145/2647868.2654943>.
- [46] W. Xuekai, M. Zhou, S. Kwong, H. Yuan, W. Jia, A hybrid control scheme for 360-degree dynamic adaptive video streaming over mobile devices, *IEEE Trans. Mob. Comput.* <http://dx.doi.org/10.1109/TMC.2021.3058099>.
- [47] Y. Sani, D. Raca, J. Quinlan, J. Sreenan, SMASH: A supervised machine learning approach to adaptive video streaming over HTTP, in: 2020 Twelfth International Conference on Quality of Multimedia Experience, QoMEX, 2020, pp. 1–6, <http://dx.doi.org/10.1109/QoMEX48832.2020.9123139>.
- [48] L. Cui, D. Su, S. Yang, Z. Wang, Z. Ming, TCLiVi: Transmission control in live video streaming based on deep reinforcement learning, *IEEE Trans. Multimed.* 23 (2021) 651–663, <http://dx.doi.org/10.1109/TMM.2020.2985631>.
- [49] H. Mao, S. Chen, D. Dimmery, S. Singh, D. Blaisdell, Y. Tian, M. Alizadeh, E. Bakshy, Real-world video adaptation with reinforcement learning, 2020, <http://dx.doi.org/10.48550/arXiv.2008.12858>.
- [50] Operators should act quickly to secure a key position in the cloud-gaming value chain, 2020, <https://www.analysismason.com/research/content/articles/cloud-gaming-value-chain-rdmb0-rdmm0-rdvs0/>. (Accessed 22 February 2023).
- [51] Unity, Technologies unity real-time development platform | 3D, 2D VR & ar engine, 2023, <https://unity.com/>. (Accessed 22 February 2023).
- [52] Open broadcaster software OBS project, 2023, <https://obsproject.com/>. (Accessed 22 February 2023).
- [53] FFmpeg: A complete, cross-platform solution to record, convert, and stream audio and video, 2023, <https://ffmpeg.org/>. (Accessed 22 February 2023).
- [54] J. Nawala, L. Janowski, M. Leszczuk, Modeling of quality of experience in no-reference model, *J. Telecommun. Inf. Technol.* (2017) <http://dx.doi.org/10.26636/jtit.2017.114517>.
- [55] P. Romaniak, M. Mui, A. Mauthe, S. D'Antonio, M. Leszczuk, Framework for the integrated video quality assessment, *Multimedia Tools Appl.* 61 (2011) <http://dx.doi.org/10.1007/s11042-011-0946-3>.
- [56] P. Romaniak, L. Janowski, M. Leszczuk, Z. Papir, Perceptual quality assessment for H.264/AVC compression, in: 2012 IEEE Consumer Communications and Networking Conference, CCNC'2012, 2012, <http://dx.doi.org/10.1109/CCNC.2012.6181021>.
- [57] Foundation, A.S., Apache kafka, 2023, <https://kafka.apache.org/>. (Accessed 3 March 2023).
- [58] M. Gupta, R. Legouable, M.M. Rosello, M. Cecchi, J.R. Alonso, M. Lorenzo, E. Kosmatos, M.R. Boldi, G. Carrozzo, The 5G EVE end-to-end 5G facility for extensive trials, in: 2019 IEEE International Conference on Communications Workshops, ICC Workshops, 2019, pp. 1–5, <http://dx.doi.org/10.1109/ICCW.2019.8757139>.
- [59] J. Garcia-Reinoso, M.M. Roselló, E. Kosmatos, G. Landi, G. Bernini, R. Legouable, L.M. Contreras, M. Lorenzo, K. Trichias, M. Gupta, The 5G EVE multi-site experimental architecture and experimentation workflow, in: 2019 IEEE 2nd 5G World Forum, 5GWF, 2019, pp. 335–340, <http://dx.doi.org/10.1109/5GWF.2019.8911624>.
- [60] W. Nakimuli, G. Landi, R. Perez, M. Pergolesi, M. Molla, C. Ntogkas, G. Garcia-Aviles, J. Garcia-Reinoso, M. Femminella, P. Serrano, F. Lombardo, J. Rodriguez, G. Reali, S. Salsano, Automatic deployment, execution and analysis of 5G experiments using the 5G EVE platform, in: 2020 IEEE 3rd 5G World Forum, 5GWF, 2020, pp. 372–377, <http://dx.doi.org/10.1109/5GWF49715.2020.9221060>.
- [61] M. Deepanshu, State-of-the-art reinforcement learning algorithms, *Int. J. Eng. Technol. Res. V8* (12) (2020) <http://dx.doi.org/10.17577/IJERTV8IS120332>.
- [62] H. Dave, Bellman optimality equation in reinforcement learning, 2023, <https://www.analyticsvidhya.com/blog/2021/02/understanding-the-bellman-optimality-equation-in-reinforcement-learning/>. (Accessed 23 February 2023).
- [63] V. Mnih, A. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, K. Kavukcuoglu, Asynchronous methods for deep reinforcement learning, 2016, <http://dx.doi.org/10.48550/arXiv.1602.01783>.
- [64] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, S. Levine, Soft actor-critic algorithms and applications, 2018, arXiv 1812 (05905) URL <http://arxiv.org/abs/1812.05905>.



Alberto del Río was graduated in Mobile and Space Communications Engineering from the Universidad Carlos III of Madrid (UC3M), and has a Master's Degree in Signal Processing and Machine Learning for Big Data at the Universidad Politécnica de Madrid (UPM). He worked at Deutsche Telekom (Berlin) in the specification of the standard of 5G telecommunications system, Release 16. Specifically in the development of a system framework concept focused on cloud services. Currently, he is Ph.D. candidate at UPM and working within the Grupo de Aplicación de Telecomunicaciones Visuales (GATV) of the UPM in dedicated projects on 5G communications networks with the help of use cases focused on AI.



Javier Serrano was graduated on Sound and Image Engineering by Universidad Politécnica de Madrid (UPM) and did a Research Master on Signal, Image, Speech and Telecommunications at Institut Polytechnique (INP) de Grenoble (France).

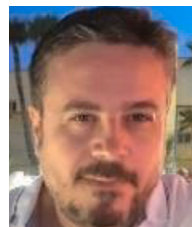
He is working now at Grupo de Aplicación de Telecomunicaciones Visuales (GATV) of UPM on testing and validation of new UHD and immersive content delivery models in next-generation mobile networks. He is a Ph.D. candidate at UPM focusing on the intelligent optimization of virtualized services over new generation networks.



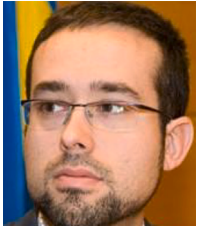
David Jiménez received the Telecom Engineer degree in 2004 by Universidad Politécnica de Madrid (UPM) where he also obtained his telecom Ph.D. degree ("cum laude") in 2012. He joined GATV as a researcher in 2004 and he is assistant professor within Department of Physical Electronics, Electrical Engineering and Applied Physics.

His main research activities are focused on 5G communications, media quality & Quality of Experience (QoE) assessment, smart energy, and electric vehicles.

He is part of the Executive Board of the Chair of RTVE at UPM. He is the author and co-author of several scientific publications and reviewer for several entities both for R&D certification and scientific publications.



Luis M. Contreras is Telecom Engineer (M.Sc.) at the Universidad Politécnica de Madrid (1997), and holds an M. Sc. on Telematics jointly from the Universidad Carlos III of Madrid (UC3M) and the Universitat Politècnica de Catalunya (2010). He is Ph.D. on Telematics from the UC3M (2021). Since August 2011 he is part of Telefónica I+D, working on 5G, SDN, virtualization, transport networks and their interaction with cloud and distributed services. He is also part-time lecturer at the UC3M. He is actively involved in research and innovation activities, with 70+ papers published in relevant journals, magazines and conferences, being regular speaker at reputed academic and industrial events. He has participated on several projects funded by the European Commission and the ESA. He is active contributor to several standardization bodies.



Federico Alvarez is Telecom Engineer with honors (2003) and Ph.D. (2009), both by Universidad Politécnica de Madrid (UPM).

He is working as Professor lecturing in UPM. He is head of the research group in the Visual Telecommunications Applications group (GATV) of the UPM since 2019. He has been in the last 10 years leading the UPM participation in several EU-funded projects, coordinating several of them. He worked as expert for the European Institute for Prospective

Technological Studies for mobile search. He had taken part in several standardization bodies and is author and co-author of (70+) papers in journals, congresses, and books in the field of ICT technologies. He is serving in the Programme Committee of several congresses and as reviewer of scientific journals.