Journal of the Brazilian Computer Society, 2023, 29:1, doi: 10.5753/jbcs.2023.2228

Dynamic video service migration in flying edge computing networks

Fábio Araújo [**Federal University of Pará** | ***fabioraraujo@ufpa.br***]
Felipe Araújo [**Federal University of Pará** | ***felipearaujo@ufpa.br***]
Derian Alencar [**Federal University of Pará** | ***derian.alencar@itec.ufpa.br***]
Stefany Pinheiro [**University Center of Pará** | ***stefpinheiro9@gmail.com***]
Helder Oliveira [**Federal University of Pará** | ***heldermay@ufpa.br***]
Denis Rosário [**Federal University of Pará** | ***denis@ufpa.br***]

Rua Augusto Corrêa, 01, Guamá, Belém/PA, CEP 66075-110, Brazil.

Received: 20 September 2021 • **Accepted:** 22 February 2023 • **Published:** 31 August 2023

Abstract Recently, Mobile Edge Computing (MEC) and service migration policies have shown promising results to improve the user experience and optimize infrastructure resources. In addition, Unmanned Aerial Vehicles (UAVs) appear as a promising solution to provide cloud service in collaboration with traditional MEC scenarios. However, in conjunction with the resources available on users' devices, contextual information has been ignored by most policies. In this article, we propose a service migration strategy based on contextual information and evaluate the influence of user mobility on migration strategies, called DVSM. Simulation results highlight the superior performance of the DVSM compared to state-of-the-art algorithms and a performance equivalent to the optimal solution when the collection and analysis of context information are carried out correctly.

Keywords: MEC, FEC, QoE Support, Service Migration, UAV

1 Introduction

Mobile communication technologies have experienced exponential growth over the last few years in the number of users, applications, and data traffic with services such as high-quality multimedia streaming, Virtual and Augmented Reality (VR/AR), and online gaming Xu et al. (2020); Alencar et al. (2021). However, it is challenging to deliver such content with Quality of Service (QoS) support (*i.e.,* low latency, mobility support, energy efficiency, and high bandwidth) over the current communication infrastructure Li et al. (2018); Araújo et al. (2019). Meanwhile, the distance between mobile users and geographically distributed cloud data centers substantially impacts latency.

According to recent surveys, multimedia services represented 82% of traffic by 2022 (Cisco, 2019). Most of the time, these services are requested by users in constant movement, and their continuity is one of the key factors for a seamless Quality of Experience (QoE). As a promising technology to accommodate the explosive growth of such service in mobile scenarios, Mobile Edge Computing (MEC) has attracted a great amount of research interest from global researchers and engineers Rosário et al. (2018). MEC deploys calculation, storage, processing, and other functions on the side of Base Station (BS) without uploading tasks to the cloud data center Santos et al. (2020). It has a number of benefits, such as reducing bandwidth consumption, reducing latency by locating resources at nearby users, and improving transmission efficiency Aggarwal and Kumar (2019).

However, decreasing transmission delay and maintaining good QoE and QoS is much more than simply deploying mobile applications and services at the network's edge. This is due to unpredictable user mobility, limited range of fixed BS, and scarcity of available resources. To overcome these challenges, Unmanned Aerial Vehicles (UAVs) are solid alternatives due to UAVs have been growing in computing and storage capabilities, as well as longer flight times leading to Flying Edge Computing (FEC) era Faraci et al. (2020); Yuan and Muntean (2020). Specifically, FEC provides computing and storage capabilities in UAVs, which can connect directly to users and provide services, which diverges from traditional MEC at its capabilities for dynamically positioning UAV nodes in 3D space to provide Line of Sight (LOS) links between servers and end-users Pacheco et al. (2021). In general, UAVs can be aggregated with traditional MEC networks and act as mobile base stations to improve QoS and QoE in very low latency and very high bandwidth services Zhao et al. (2022). In addition, a UAV can serve local content caching and forwarding Montero et al. (2021).

In addition to architecture, an adequate service migration mechanism that jointly considers context factors and device characteristics becomes an important issue in an FEC scenario. The mechanism guarantees the efficiency of the requested service migration, migrating it in a timely manner to be executed, avoiding new requests and an overload on the network. In this sense, to best serve user needs, services must be migrated to UAVs to keep the services as close as possible to the mobile users, considering the trade-off between minimal QoS requirements for mobile users and cost Pacheco et al. (2021). Hence, mechanisms for managing services are fundamental to maintaining the QoE.

This article introduces a dynamic video service migration in the FEC scenario called DVSM. It logically migrates services to serve as many user requests as possible and reduces latency and energy consumption. The dynamic video service migration mechanism relies on latency, storage, available bandwidth, and remaining energy. Simulation results show that DVSM maintains QoE close to the optimal solution and up to 74% higher compared to recent studies in the literature. For instance, DVSM can reduce the number of stalls up to 80% and the duration of stalls up to 83%. In addition, DVSM increases the average bitrate of adaptive video stream up to 50% better than analyzed service migration mechanisms

We extend the previous work Araújo et al. (2021) with an extensive review of related works, a more detailed description of the mechanism, and an extensive evaluation of the impact of contextual information in video service migration mechanisms under a different network scenario. In this context, the contributions of this work can be summarized as follows:

- We model the video service migration problem as Markov Decision Process (MDP), and the service migration reward function is constructed taking into account context features.
- As solving the optimal solution in a problem might be hard due to high complexity in time, we also design a dynamic video service migration algorithm, which reduces the search space, determining a cutoff threshold.
- We study the influence of predictive features on the quality of video services migration, evaluating different schemes of service migration. Results show that service quality can be degraded if context information is not adequately predicted, even if looking for the optimal solution.

The remainder of this article is organized as follows. Section 2 describes existing proposals in the literature and their main differences. Section 3 discusses the system model and the dynamic video service migration mechanism. Section 4 describes the simulation environment and discusses simulated results. Section 5 presents conclusions.

2 Related Work

Based on the user mobility model, Chen et al. (2017) designed a green and mobility-aware caching scheme by jointly optimizing user mobility and transmission power, which can improve the content service migration ratio and reduce energy consumption. Besides, Quer et al. (2018) proposed a proactive storage policy as a promising possibility to overcome the reactive paradigm. They provided a closed-form expression for the average cost of the system and derived an optimization framework for the storage service. Results show the performance of optimal and heuristic policies for a static (non-proactive) policy, showing that it can significantly reduce system cost for all levels of user mobility when users share a common interest in some files.

Alencar et al. (2021) proposed a QoE VR-based mechanism for allocating microservice dynamically in 5G architectures named Fog4VR. Moreover, they presented an integer linear programming model to find the optimal global solution for microservice allocation known as INFORMER. Their results demonstrate the efficiency of Fog4VR compared to existing mechanisms in terms of cost, migration time, fairness index, and QoE.

Vo et al. (2020) proposed a multi-layer storage and resource-sharing solution that allows mobile users to receive video versions through communications carried out between devices, micro BSs, and macro BSs in ultra-dense 5G networks. Their work considers downlink resources and storage resources in macro cells, micro, and mobile nodes in order to satisfy the mobile user requesting a specific video.

As a promising solution, many studies have proposed Markov-based techniques to migrate services. Li et al. (2021) proposed a dynamic service migration based on deep Q learning in a Software Defined Network (SDN)-based MEC architecture to provide users with seamless service migration to ensure service continuity and high-quality services. Firstly, the service migration problem was expressed as a Markov decision process. Secondly, the service migration process was analyzed, and a service migration reward function was constructed. Finally, deep Q learning was used to obtain the optimal service migration strategy and to effectively improve the cache hit rate, to reduce the backhaul traffic load, and to control the average access delay and energy cost.

Ouyang et al. (2018) proposed a mobility-aware dynamic placement for MEC, based on Lyapunov optimization and Markov approximation technique. Their proposed algorithm aims to dynamically migrate services among multiple edge nodes to maintain the service performance. Simulation results demonstrated the effectiveness of the mobility-aware algorithm compared to some baseline strategies such as the Always migration strategy, No migration strategy, and a Greedy strategy.

Due to FEC being new in service migration, we have addressed some works like Ouyang et al. (2021) and Costanzo et al. (2020), where UAVs act as base stations in migration strategies. Ouyang et al. (2021) established a model of the Markov decision process with unknown rewards (MDPUR) based on the traditional MDP, which comprehensively considers the three aspects of the migration distance, the residual energy status of the UAVs, and the load status of the UAVs. Based on the MDPUR model, they proposed an advantage-based value iteration (ABVI) algorithm to obtain an effective task migration strategy, which can help the UAV group to achieve load balancing and reduce the total energy consumption of the UAV group under the premise of ensuring user service quality. Results of simulation experiments showed that the ABVI algorithm is effective compared to the traditional value iterative algorithm.

Costanzo et al. (2020) proposed a dynamic optimization strategy to allocate communication and computation resources in an MEC scenario, where UAVs act as flying base station platforms endowed with computation capabilities to provide edge cloud services on demand. The method exploits stochastic optimization tools to optimize the resources and the KW algorithm to update the altitude of the vehicle in order to find the optimal position concerning the overall system energy saving under average latency constraints.

Gao et al. (2017) proposed a storage-aware scheme for distributed surveillance video processing with hybrid storage architecture, which considers the heterogeneity of the node pro-

Table 1. Summary of related work.

Study	Features					Solution
	Location	Storage	Net. Band.	Energy	Delay	
Chen et al. (2017)	✓	-	-	✓	-	Proposed an optimization of user mobility and transmission power.
Quer et al. (2018)	✓	-	-	-	✓	Proposed an integer linear programming and an optimization heuristic
Alencar et al. (2021)	✓	✓	✓	-	✓	Proposed an QoE-VR-based mechanism and a Linear Programming
Vo et al. (2020)	-	✓	-	✓	-	Proposed a mixed integer linear programming
Li et al. (2021)	-	-	✓	✓	✓	Proposed a dynamic service migration based on deep Q learning
Ouyang et al. (2018)	✓	-	-	-	-	Proposed a mobility-aware dynamic placement for MEC, based on Lyapunov optimization and Markov approximation technique
Ouyang et al. (2021)	✓	✓	-	✓	-	Proposed an ABVI algorithm for effective task migration.
Costanzo et al. (2020)	✓	-	-	✓	✓	Proposed a dynamic optimization strategy to allocate communication and computation resources in a Multi-access Edge Computing scenario
Gao et al. (2017)	-	✓	-	-	-	Proposed a storage-aware scheme for distributed surveillance video processing with hybrid storage architecture
Wang et al. (2019)	✓	-	-	-	✓	Designed an MDP and new algorithm and numerical technique for computing the optimal solution
Hao et al. (2020)	✓	✓	✓	-	✓	Proposed a cognitive caching scheme
Zhu et al. (2019)	✓	✓	✓	-	✓	Proposed an architecture and a scheme for optimizing video service migration
Proposal	✓	✓	✓	✓	✓	Designed a MDP and an dynamic service migration mechanism

cessing and storage capabilities. Their results showed that the proposed scheme could significantly improve the efficiency of surveillance video processing and data migration.

In this way, Wang et al. (2019) formulated a service migration problem as an MDP. Their formulation captures general cost models and provides a mathematical framework to design optimal service migration policies. In order to overcome the complexity associated with computing the optimal policy, they approximated the underlying state space by the distance between the user and service locations. Moreover, the authors also proposed a new algorithm and a numerical technique for computing the optimal solution, which is significantly faster than traditional methods based on the standard value or policy iteration.

Migrating services or contents to edge servers reduce the transmission and latency to access the service. Based on this, Hao et al. (2020) proposed a cognitive caching scheme architecture and scheme that includes what to do in cognitive caching and how to do cognitive caching. Their experimental results show that the proposed cognitive caching scheme is superior to other caching schemes in terms of learning regret and caching cost. Further, Zhu et al. (2019) proposed an architecture and a scheme with solutions for optimizing video service migration called PMVCS. The authors exploited network node prediction, caching, and D2D communication capabilities while considering the adaptive characteristics of multi-bitrate videos.

The above works have studied the problem of service migration. Even though most work has considered user mobility and network delay, the factors they consider need to be more comprehensive. Hence, a dynamic service migration mechanism based on delay, user mobility, available storage, and devices' remaining energy in FEC networks is proposed. The dynamic service migration mechanism considers several contextual factors to make mobile edge caching more effective. Compared with previous research works, the differences of the proposed mechanism in this paper are shown in **Table 1**.

3 Dynamic Video Service Migration in Flying Edge Computing Scenario

In this section, we present the service migration mechanism named DVSM by considering the remaining energy of users' devices, user mobility, and network conditions as contextual information. We consider adaptive video distribution in an FEC architecture, where server nodes are deployed anywhere in the network, such as macro-BSs, fixed micro-BSs, UAVs, edge nodes, etc. Micro BSs nodes are placed between mobile devices (at the bottom) and macro BSs, as shown in Figure 1. In this way, both cloud, BSs, and edge nodes work collaboratively to provide video services with QoE support. Moreover, we model the whole system and migration mechanism in an MDP model in order to find the optimal solution and show how close DVSM is to that.

3.1 System Model

Figure 1 illustrates an FEC Architecture formed of end-users, mobile Micro-BSs (UAV-BSs), fixed Micro-BSs, Macro-BSs, and cloud, all connected directly or indirectly to each other using wireless technologies. We define a set of micro BS as $b \in \mathcal{B} \equiv \{b_1, \ldots, b_{SBS}\}$ and macro BSs as $m \in \mathcal{M} \equiv \{m_1, \ldots, m_{MBS}\}$. Micro BSs are divided into two categories: fixed micro BS with a low transmission power; and mobile micro BS (*i.e.,* UAV as BS), which can move in three dimensions and connect directly to the end-user when there is a LOS link. Further, we assume that a user cannot simultaneously connect to multiple micro BSs. Each user $u \in \mathcal{U} \equiv \{u_1, \ldots, u_U\}$ is mobile and is always connected to a macro BS m or a micro BS b closer.

Time has been divided into time intervals that are labeled with a discrete index $t \in \mathbb{N}$. Users can request a file at each time interval, and the server proactively migrates the services to a specific area (*i.e.,* mobile nodes, micro BSs, or macro BSs). Also, the requested file can be transmitted over a Device to Device (D2D) link, if it is available in the local cache of another user in the same area. Otherwise, it is delivered over an mmWave link by the micro or macro BS.

3.2 Markov Decision Process Model

According to Chen et al. (2017), user mobility impacts service migration performance and wasted resources on cellular networks. Hence, it is essential to consider user mobility and node distribution to ensure high system quality and provide a service migration without wasting resources. In addition, inaccurate information can also limit service performance. Overestimating remaining energy, storage, and bandwidth values can cause disruptions in video service migration, and underestimating these values can limit migration, such as the quality of an adaptive video.

In this way, we model each request of video service migration as an MDP as shown in Figure 2. Let n_t be a state at time t, and $\mathcal{N}$ denote the state space that contains all states.

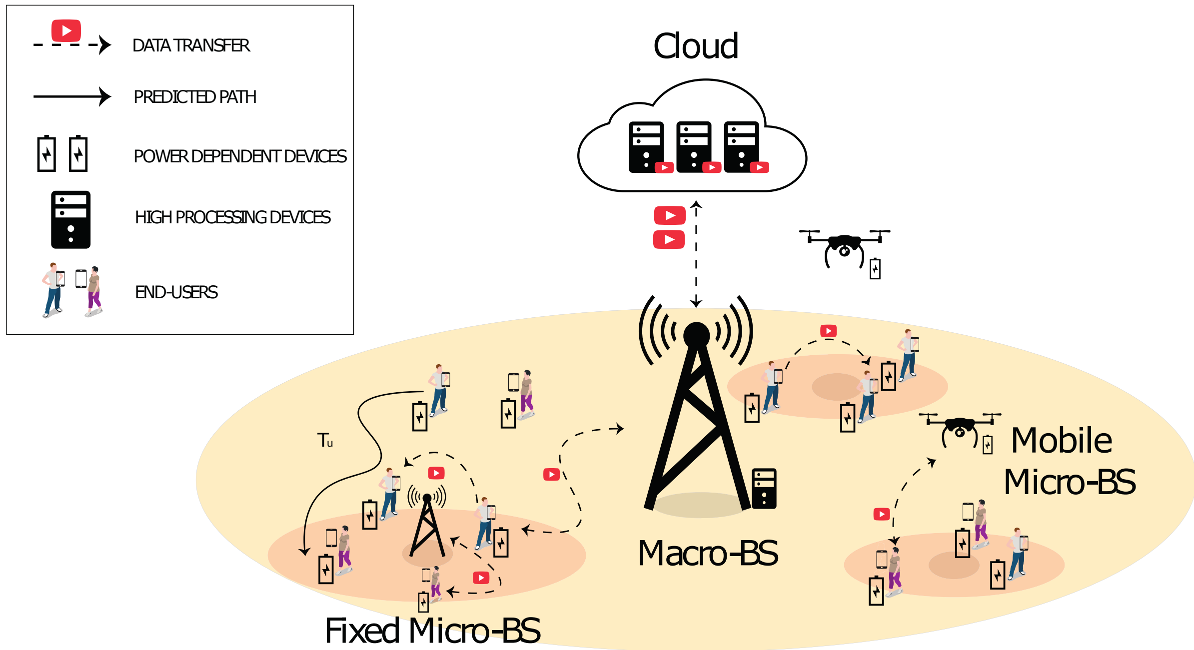


Figure 1. Flying Edge Computing Scenario.

It is assumed that UAVs and users have predefined locations that change over time, as well as limited energy and storage. The state space $\mathcal{N}$ is defined as $\mathcal{N} = \{\mathcal{C} \cup \mathcal{M} \cup \mathcal{B} \cup \mathcal{U}\}$, where $\mathcal{C}$ is a set of distributed cloud server. Same as Wang et al. (2019), we use the concept of action set. For example, $A_n = (a_1, a_2)$ can denote the action set available at states, where action a_1 means that the service is migrated to another server, while action a_2 means the same server still serves the user.

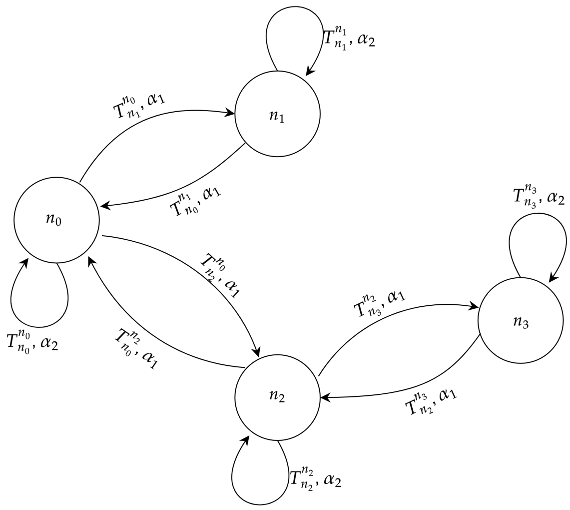


Figure 2. Markov Decision Process model of service migration.

For a given action a, there will be a state transition from state to another state n', with which there is also a reward $r(n, n', a)$ given by:

$$r(n, n', a) = (\alpha S_{n'} - \beta d_{u,n'})\, \delta^{S_{n'},net}_{f,u,n}\, Z^f_u \tag{1}$$

where n' is the state at time $t + 1$; α and β are weighted factors; $S_{n'}$ is the available storage on destination node at state n'; and $d_{u,n'}$ is the network delay between node u and destination node at state n'. Z^f_u is the probability of a user u requesting a file $f \in \mathcal{F}$, and $\delta^{S_{n'},net}_{f,u,n}$ is a binary state variable that determines whether the minimum network, storage, and energy requirements were met. Let user mobility be known through the probability matrix T, where $T^n_{n'}$ denotes the probability of a user moving from state n to state n', which is equal to the transition probability $p(n'|n, a)$. Given that, a long-term expected reward generated by policy π is denoted as follows:

$$v^\pi = \lim_{k\to\infty} E\{\sum_{t=0}^{k} r(n_t = n, n_{t+1} = n', a_t = a_\pi)\} \tag{2}$$

In most scenarios, v^π is limited by the user mobility, reaching a terminal state in MDP. In order to limit the maximum value of v^π regardless of user mobility, we denote a discount factor $0 \le \gamma \le 1$, and the long-term expected discounted reward is given by Eq. 3.

$$v^\pi_\gamma = \lim_{k\to\infty} E\{\sum_{t=0}^{k} \gamma^t r(n_t = n, n_{t+1} = n', a_t = a_\pi)\} \tag{3}$$

We denote $v^*(n)$ the maximum long-term expected discounted reward as $v^*(n) = max_\pi v(n)$, given initial state n. Using the predefined denotations, we solve the maximization problem by the Bellman Expectation equation $v^*(n)$ (Puterman, 2014), which can be expressed in Eq.4, and we find the optimal solution. **Table 2** summarizes the most used notations in this work.

$$v^*(n) = max_\pi\{r(n, n', a) + \sum_{n'\in\mathcal{N}} \gamma p(n'|n, a) v(n')\} \tag{4}$$

Table 2. Most used notation.

Symbol	Definition
$\mathcal{U}, \mathcal{U}^b, u$	User set, user index
$\mathcal{B}, b$	Micro BS set, micro BS index
$\mathcal{M}, m$	Macro BS set, macro BS index
$\mathcal{N}, n, n'$	State set, state at the current step, state at time t+1
$\mathcal{F}, f$	File set, file index
$\mathcal{J}, j$	Destination node set, destination node index
S_n	Available storage of destination node at state n
$d_{u,n}$	Network delay from user u to destination node at state n
T_u^n	Probability of a user node u being in the range of a destination node at state n
Z_u^f	Probability of a user u request a file f
α, β	Weight variables
$\delta_{f,u,n}^{S,net,en}$	Binary variable of network, storage and energy requirements
γ	Discount factor
$r(n, n', a)$	Instant reward
$NET_{u,j}^{t+1}$	Instant network available bandwidth at time $t+1$
S_j^{t+1}	Instant available storage at time $t+1$
EN_j^{t+1}	Instant remaining energy at time $t+1$
τ	Minimum threshold
v^π	Long-term expected reward
v_γ^π	Long-term expected discounted reward
$v^*(n)$	Maximum Long-term expected discounted reward/Bellman Expectation equation

3.3 Dynamic Video Service Migration

The key principle of our proposed strategy is to ensure that the requirements for service migration are met and delays are mitigated to provide a seamless service migration in terms of end-user QoE, under the conditions of that context. It means that video quality and location are limited by context factors such as user mobility, node storage capacity, or power supply.

We define the optimal solution as the MDP solution using Bellman's expectation equation (Eq 4) in the search for the highest reward, which relies on the amount of available storage, as well as remaining energy in the devices and good connection quality. In addition, assuming the optimal solution of the MDP model is a context-aware solution, we developed a Dynamic Video Service Migration algorithm (DVSM) that uses cutoff values and prediction methods in search of a satisfactory solution for various scenarios. Let j be a destination node and $\mathcal{J}$ a set of destination nodes. Assuming $1\{W\}$ the indicator function of a requirement W, we denote with $\delta_{f,u,j}^{NET}$, $\delta_{f,j}^{S}$, and $\delta_{f,j}^{EN}$ binary state variables, provided by:

$$\delta_{f,u,j}^{NET} = 1\{NET_{u,j}^{t+1}\} \geq \tau_{f,u,j}^{net}, \forall f \in \mathcal{F}; \forall u \in \mathcal{U}; \forall j \in \mathcal{J} \tag{5}$$

$$\delta_{f,j}^{S} = 1\{S_j^{t+1} \geq \tau_{f,j}^{S}\}, \forall f \in \mathcal{F}; \forall j \in \mathcal{J} \tag{6}$$

$$\delta_{f,j}^{EN} = 1\{EN_n^{t+1} \geq \tau_{f,j}^{en}\}, \forall f \in \mathcal{F}; \forall j \in \mathcal{J} \tag{7}$$

These equations include instant values of available network bandwidth, available storage, and remaining energy at time $t+1$ ($NET_{u,j}^{t+1}$, S_j^{t+1}, and EN_j^{t+1}) and a set of minimum thresholds $\tau_{u*,f,j}$ that may vary according to each user, file and state.

We assume a node controller receives mobility matrix T and is also topology-aware. For each file $f \in F$ where $Z_u^f > 0$, the algorithm randomly selects an edge node as the destination node. Moreover, the algorithm computes a probability of the user moving to the same location as the destination node at time $t+1$ and estimates a network bandwidth according to exponential moving average (EMA) algorithm, where coefficient ϕ represents the degree of weighting decrease, a constant smoothing factor between 0 and 1, as shown in Eq. 8 (line 5).

$$\hat{NET} = \begin{cases} \phi \hat{NET}(i-1) + (1-\phi) NET(t)\ ,\ if\ t > 1 \\ NET(1)\ ,\ if\ t = 1 \end{cases} \tag{8}$$

Algorithm 1: Dynamic video service migration algorithm.

Input : $\mathcal{J}, T, Z_u^f$
Output : Destination Node

1 Randomly set destination node as $j \in \mathcal{J}_{edge}$
2 **foreach** $j \in \{\mathcal{J}_{edge}, \mathcal{J}_{micro}, \mathcal{J}_{macro}, \mathcal{J}_{cloud}\}$ **do**
3   **foreach** $T_{u,j} > 0 \in T$ **do**
4   **foreach** $Z_u^f > 0 \in Z$ **do**
5   Calculate $NET_{f,u,n}^{t+1}$ with Eq. 8;
6   Calculate S_j^{t+1} with Eq. 9;
7   **if** $j \in \mathcal{J}_{edge} || j \in$ *UAV)* **then**
8   Calculate EN_j^{t+1} with Eq. 10;
9   Initialize minimum thresholds τ;
10   Calculate binary variable $\delta_{f,u,n}^{S,net,en}$;
11   **if** $\delta_{f,u,n}^{S,net,en}$ **then**
12   Calculate instant reward with Eq. 1;
13   **if** *Instant reward* $> \epsilon$ **then**
14   Set destination node as j;
15   return destination node;

Algorithm 1 presents the main dynamic video service migration proposed. It starts the search at a node $j \in \mathcal{J}_{edge}$ and thus checks whether there is a probability of an encounter between the user and the destination node and the probability that node j contains the requested file (lines 2-4). After that, the proposed algorithm estimates network conditions, available storage, and remaining energy in destination node j at time $t+1$ as shown in Eq. 8, Eq. 9 and Eq. 10 (lines 5-11), where $NETt$ is the instant network bandwidth at time t, and c_1 and e_1 are constants value between 0 and 1. In addition, $EN(\text{size}(\text{index}(Z_u^f)))$ returns the energy required to transmit the most requested file. Finally, it computes the instant reward using Eq. 1 and sets the destination node as j if the reward value is higher than the cutoff threshold (lines 12-15). Otherwise, it evaluates the next candidate node in a set of the same category. When there are no nodes with minimum requirements in that category, the service migration strategy shifts the search space to a subsequent node category (line 2).

$$S_j^{t+1} = S_j^t - c_1 NETt \tag{9}$$

$$EN^{t+1} = EN^t - e_1 EN(\text{size}(\text{index}(Z_u^f))) \tag{10}$$

The MDP solution using Bellman's expectation equation may require a large number of iterations before converging to the optimal solution, which results in an unfeasible solution for real scenarios. However, the average complexity of the proposed algorithm depends directly on the probability of stopping the search. Eq. 11 describes the probability of stopping the search as the sum of the probabilities of the user finding the possible destination nodes multiplied by the probability of these nodes meeting the minimum requirements of a viable solution. As a result (shown in Figure 3), as the number of nodes in the network increases, the probability of stopping the search increases, reducing the time to find a viable solution to the problem.

$$P_{stop} = \sum_{j=1}^{\mathcal{J}} T_{u,j} P_j(instantreward > \epsilon) \quad (11)$$

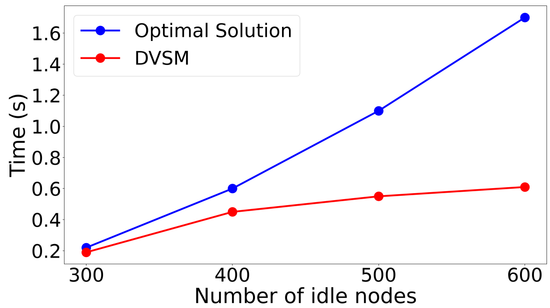


Figure 3. Response time per number of idle nodes

4 Performance Evaluation

This section presents an evaluation of the DVSM mechanism using various simulation tests and performance metrics. We describe the simulation environment and the performance metrics used. Specifically, we analyzed the performance of DVSM and four other mechanisms.

4.1 Simulation Setup

We implemented the proposed service migration strategy and the other evaluation strategies in the NS-3[1] simulator. NS-3 implements the LTE and mmWave protocol stack for communication. It is expected that 5G networks will be composed of ultra-dense heterogeneous radio networks compared to 4G systems, increasing the data rate at the network edge Costa et al. (2020). In this sense, we considered a scenario with an area of 2 km ×2 km, with 8 macro-BSs covering the entire scenario and 90 micro-BSs distributed across the scenario. Specifically, we deployed 78 fixed Micro-BS and 12 mobile Micro-BS (*i.e.,* 6 active mobile Micro-BS and 6 backup mobile Micro-BS). We consider the backup micro-BS to replace active BSs running out of battery. Macro cells have a transmit power of [46] dBm, while micro cells have a transmit power of [23] dBm, including UAV-BSs.

Further, we implemented different scenarios, varying the number of active nodes (nodes that request content/service) from 60 to 120, and idle/relay nodes from 300 to 600. The simulation considers the Nakagami and LOS path loss model, which can be very suitable for urban scenarios and four videos encoded in seven qualities, following a video service request distribution, as shown in **Table 3**. It is important to mention that the choice of video quality depends on the context conditions, which is like a part of the output of the proposed algorithm. Therefore, it is possible to have all 7 video qualities present in the scenarios, or it could be situations with only 1 video quality present at the moment. Hence, the choice depends on the conditions.

Table 3. Resolution and Bitrate.

Resolution	Bitrate
240p	300 kbps
360p	500 kbps
HD 480p	1000 kbps
HD 720p	2250 kbps
Full HD 1080p	4500 kbps
Quad HD 1440p	9000 kbps
4K UHD 2160p	13000 kbps

We conducted a total of 231 simulations distributed uniformly (33) for each parameter set: 1) 300 idle nodes and 60 active nodes; 2) 300 idle nodes and 80 active nodes; 3) 300 idle nodes and 90 active nodes; 4) 300 idle nodes and 120 active nodes; 5) 400 idle nodes and 60 active nodes; 6) 500 idle nodes and 60 active nodes; 7) 600 idles nodes and 60 active nodes. It is important to state that there are seven different scenarios, but they were aggregated into two different setups. The first setup unites the four first scenarios (*i.e.,* scenarios 1-4), while the second setup unites scenarios 1 and 5-7. All 231 simulations were executed in a single batch of the script, uninterruptedly, for a few days; however, they are presented here as seven distinct scenarios and then split among two different simulation setups for a better presentation and discussion of the results. Details of simulation parameters are shown in **Table 4**.

Table 4. Simulation Parameters.

Parameter	Value
Maximum number of active nodes	60, 80, 100, 120
Number of inactive nodes	300, 400, 500, 600
Storage Size (Edge, Micro-BSs, Macro-BSs, Cloud)	1000 MB, 10000 MB, 100000 MB, unlimited
Energy Model	Σ E(Transmission,Processing,Mobility)
Arrival rate of video service request distribution	Poison
Transmit Power (Edge, Micro-BSs, Macro-BSs)	23 dBm, 23 dBm, 46 dBm
Pathloss Model	Nakagami and LOS
Mobility Model	RandomWalk with pause time
UAV height	30m

Regarding adaptive video evaluation, QoS metrics are not enough to evaluate video services' quality level since they fail to capture subjective aspects of video content related to human experience and subjectivity. In this context, QoE metrics overcome those limitations to reflect subjective aspects of human perception, and thus we rely on a well-known objective QoE metric. Specifically, we investigate the impact of service migration for adaptive video services in an FEC environment from the QoE perspective (*i.e.,* stalls duration, number of stalls, and average bitrate). Moreover, we eval-

[1] http://www.nsnam.org/

uate the impact of service migration based on the service provider perspective method (*i.e.,* service provider distribution and service migration success rate) for a different number of video service requests.

4.2 Results and analysis

In this section, extensive numerical analysis is presented to verify the effectiveness of the proposed migration strategy DVSM. We compared the proposed strategy with an optimal solution, obtained solving the MDP in Section 3.2, and three baseline approaches, named, Storage-aware like Gao et al. (2017), Mobility-aware like Ouyang et al. (2018), and PMVCS. Storage-aware and Mobility-aware are service migration strategies that consider one main feature as a parameter to determine the appropriate location to migrate a service (storage and mobility, respectively). In this work, we aimed to have a proof of concept of our proposal to verify its strength against some classic service migration strategies. Therefore, Storage- and mobility-aware solutions were selected, since they are generally standard procedures in terms of migration. On the other hand, PMVCS is a more modern approach, and it refers to the migration strategy of Zhu et al. (2019) that considers other contextual factors such as network bandwidth and delay in choosing the best location for the migration of the service.

Figure 4 illustrates the success rate of service migration, distribution of services, and the number and duration of stalls for different migration strategies, and average bitrate obtained for the first simulation setup (*i.e.,* Scenarios 1-4). **Figure 4a** indicates DVSM has allocated most of the requested service in edge nodes followed by micro-BS, macro-BS, and cloud. This performance is explained by the fact that DVSM prioritizes video service migration to edge nodes with lower delay, taking into account available storage space, remaining power on the target device, available bandwidth, and user locations, enabling DVSM to choose the good locations (usually in the same category of optimal one). Before performing a service migration, the migration mechanism chooses a location to allocate the service and then executes the migration. If the service migration was not migrated properly, the requested service is migrated directly from the cloud to the user, causing delays and taking up the bandwidth of the backhaul link. The distribution metric in **Figure 4a** considers this scenario, increasing the number of services allocated in the cloud since the mechanism did not correctly allocate the requested video.

The DVSM mechanism provides results close to the optimal solution. Specifically, DVSM allocated 36% of requested service in edge nodes, 26% in micro-BS, 22% in macro-BSs, and 16% in the cloud. Optimal solution allocated 37% of requested services in edge nodes, 27% in micro-BSs, 22% in macro-BSs, and 14% in the cloud, totaling 2% difference compared to DVSM. On the other hand, PMVCS allocated 29% of services in edge nodes, 22% in micro-BSs, 20% in macro-BSs, and 29% in the cloud, totaling 13% difference compared to the optimal solution. Furthermore, mobility-aware and storage-aware mechanisms have raised the number of services allocated in the cloud to as much as 40% and 59%, respectively.

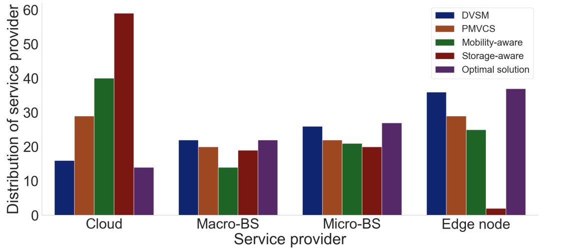
Defining the optimal solution service migration as a successful migration, we analyzed the other migration mechanisms through the service migration success rate metric as illustrated in **Figure 4b**. The proposed mechanism had the highest success rate, followed by PMVCS, Mobility-aware, and Storage -aware because DVSM and PMVCS are multiple-feature-based mechanisms while Mobility-aware and Storage-aware take into account only mobility or storage in order to find the service migration.

Results illustrated in **Figures 4c and 4d** indicate DVSM provided the lowest number of stalls with the shortest duration, which is an essential behavior since high values may induce users to leave the video service entirely. They are consequences of a higher average bitrate obtained by DVSM if compared to PMVCS, Mobility-aware, and Storage-aware, illustrated in **Figure 4e**. While the optimal solution had an average bitrate up to 4920 Kbps, DVSM had an average bitrate up to 4898 Kbps, PMVCS had an average bitrate up to 4224 Kbps, and mobility-aware and storage-aware had an average bitrate 3718 Kbps and 2814 Kbps, respectively. This equates to a 32% higher proposal quality compared to the mobility-aware strategy and 74% higher than the storage-aware strategy.

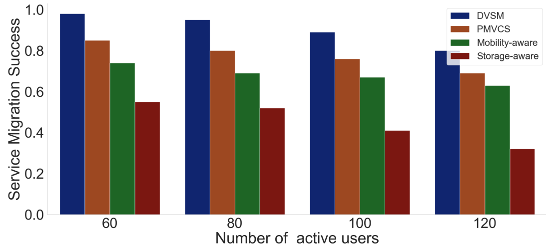
Observing results of **Figures 4a, 4c, and 4d**, we note system quality decreases in user QoE perspective and service provider perspective as long as the number of active users increases. It occurs because the number of requests also increases, overloading the network infrastructure.

In order to evaluate the performance of service migration mechanisms in another way, we performed the second setup of simulations (*i.e.,* involving Scenarios 1, 5-7). Thus, we kept the scenarios of 60 active nodes per macro-BS and varied the number of inactive users (users who do not require service migration) per macro-BS, from 300 to 600. As illustrated in **Figures 5a and 5b**, the number and duration of stalls decreased as the number of idle nodes also increased. This happens because the greater the number of idle nodes, the greater the possibility of it serving as a relay node that allocates services requested by other nearby nodes, and transmits them through a D2D link.

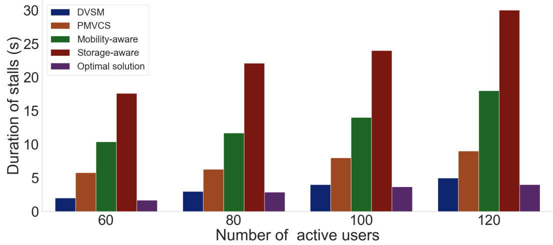
Our performance evaluation analysis identified that DVSM produces the closest performance of Optimal solution for QoE and service provider perspectives compared to other service migration mechanisms. DVSM is a multiple-feature-based mechanism that considers delay, available storage, available bandwidth, and energy, which is desirable for service migration and adaptive video applications. Therefore, DVSM provides greater efficiency in distributing content since its decision-making prioritizes efficiency in favor of balancing the network, or vice-versa, when necessary, delivering more QoE to users with better resource usage. On the other hand, other strategies are associated with the deficiency of metrics for a decision, and relying only on mobility or storage, leads to inadequate allocation decisions, especially for high-demanding scenarios. Finally, PMVCS has a poor performance related to user QoE and service provider perspective compared to DVSM due to metrics used for decision-making methods for predicting contextual features.



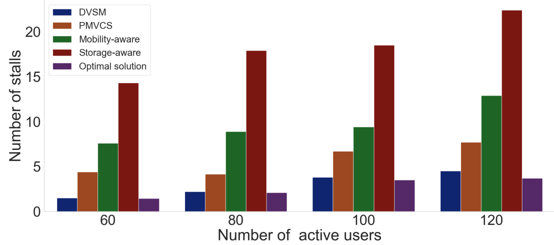
(a) Service provider distribution in a scenario of 60 active nodes and 300 inactive nodes per macro-BS.



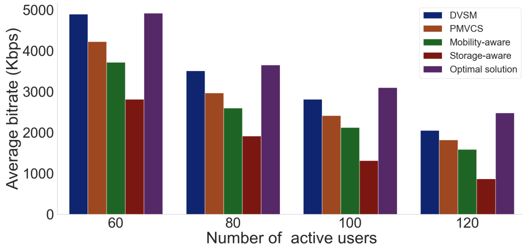
(b) Service migration success rate per number of active per macro-BS.



(c) Duration of stalls.

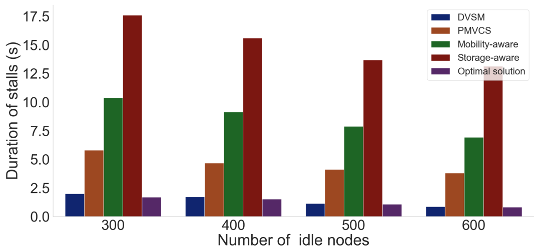


(d) Number of stalls.

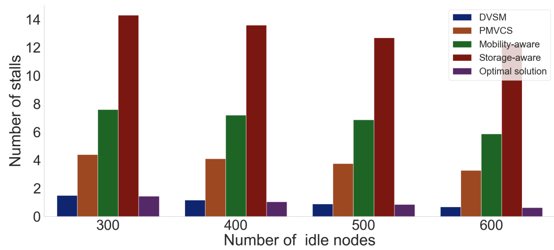


(e) Average bitrate of the simulated FEC scenario.

Figure 4. Results based on number of active nodes per macro-BS for adaptive video service migration.



(a) Duration of stalls.



(b) Number of stalls.

Figure 5. Results based on number of idle nodes per macro-BS for adaptive video service migration.

5 Final Considerations

This work introduced DVSM, a dynamic service migration mechanism for adaptive video stream services. DVSM distributes adaptive video content with QoE support using the flying edge computing environment concepts. For decision-making, the DVSM mechanism uses a cutoff-based algorithm to find a satisfactory solution. Moreover, this work introduced an optimal solution found through an MDP model. Simulations were performed with the DVSM, PMVCS, optimal solution, Mobility-aware, and Storage-aware mechanisms to distribute video contents based on service migration and FEC environment. Results showed that the DVSM mechanism decreases the stalls and stall duration and increases the average bitrate compared to other mechanisms while staying close to the MDP optimal solution. We plan to extend our mechanism for future work by adding new contextual features such as computing capacity and new methods of predicting mobility. Furthermore, we intend to evaluate in other 5G-dense scenarios, migrating Ultra-Reliable Low-Latency applications, and performing against other modern approaches that use different contextual factors.

Declarations

Funding

This study was funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, and also by the Brazilian National Council for Research and Development (CNPq).

Authors' Contributions

All authors contributed to the writing of this article, read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data can be made available upon request.

References

Aggarwal, S. and Kumar, N. (2019). Fog Computing for 5G-Enabled Tactile Internet: Research Issues, Challenges, and Future Research Directions. *Mobile Networks and Applications*, pages 1–28.

Alencar, D., Both, C., Antunes, R., Oliveira, H., Cerqueira, E., and Rosário, D. (2021). Dynamic microservice allocation for virtual reality distribution with qoe support. *IEEE Transactions on Network and Service Management*.

Araújo, F., Rosário, D., Cerqueira, E., and Villas, L. A. (2019). A hybrid energy-aware video bitrate adaptation algorithm for mobile networks. In *2019 15th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pages 146–153. IEEE.

Araújo, F., Sobral, D., Pinheiro, S., Oliveira, H., and Rosário, D. (2021). Estratégia de migração de serviço baseada em informações contextuais em uma arquitetura mec. In *Anais do XLVIII Seminário Integrado de Software e Hardware*, pages 50–57. SBC.

Chen, M., Hao, Y., Hu, L., Huang, K., and Lau, V. K. (2017). Green and mobility-aware caching in 5g networks. *IEEE Transactions on Wireless Communications*, 16:8347–8361.

Cisco (2019). Cisco visual networking index: global mobile data traffic forecast update, 2017–2022. *Update*, 2017:2022.

Costa, A., Pacheco, L., Rosário, D., Villas, L., Loureiro, A. A., Sargento, S., and Cerqueira, E. (2020). Skipping-based handover algorithm for video distribution over ultra-dense vanet. *Computer Networks*, 176:107252.

Costanzo, F., Lorenzo, P. D., and Barbarossa, S. (2020). Dynamic resource optimization and altitude selection in uav-based multi-access edge computing. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4985–4989.

Faraci, G., Grasso, C., and Schembra, G. (2020). Fog in the clouds: Uavs to provide edge computing to iot devices. *ACM Transactions on Internet Technology (TOIT)*, 20(3):1–26.

Gao, Y., Zhang, H., Zhu, Y., Tang, B., and Ma, H. (2017). A load-aware data migration scheme for distributed surveillance video processing with hybrid storage architecture. In *2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pages 563–570.

Hao, Y., Chen, M., Cao, D., Zhao, W., Petrov, I., Antonenko, V., and Smeliansky, R. (2020). Cognitive-caching: Cognitive wireless mobile caching by learning fine-grained caching-aware indicators. *IEEE Wireless Communications*, 27:100–106.

Li, C., Zhu, L., Li, W., and Luo, Y. (2021). Joint edge caching and dynamic service migration in sdn based mobile edge computing. *Journal of Network and Computer Applications*, 177.

Li, J. et al. (2018). Modeling QoE of Virtual Reality Video Transmission over Wireless Networks. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–7.

Montero, E., Rocha, C., Oliveira, H., Cerqueira, E., Mendes, P., Santos, A., and Rosário, D. (2021). Proactive radio- and qos-aware uav as bs deployment to improve cellular operations. *Computer Networks*, 200:108486.

Ouyang, T., Zhou, Z., and Chen, X. (2018). Follow me at the edge: Mobility-aware dynamic service placement for mobile edge computing. *IEEE Journal on Selected Areas in Communications*, 36(10):2333–2345.

Ouyang, W., Chen, Z., Wu, J., Yu, G., and Zhang, H. (2021). Dynamic task migration combining energy efficiency and load balancing optimization in three-tier uav-enabled mobile edge computing system. *Electronics (Switzerland)*, 10:1–30.

Pacheco, L., Oliveira, H., Rosário, D., Zhao, Z., Cerqueira, E., Braun, T., , and Mendes, P. (2021). Towards the Future of Edge Computing in the Sky:Outlook and Future Directions. In *17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 220–227.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Quer, G., Pappalardo, I., Rao, B. D., and Zorzi, M. (2018). Proactive caching strategies in heterogeneous networks with device-to-device communications. *IEEE Transactions on Wireless Communications*, 17:5270–5281.

Rosário, D., Schimuneck, M., Camargo, J., Nobre, J., Both, C., Rochol, J., and Gerla, M. (2018). Service migration from cloud to multi-tier fog nodes for multimedia dissemination with qoe support. *Sensors*, 18(2):329.

Santos, H., Alencar, D., Meneguette, R., Rosário, D., Nobre, J., Both, C., Cerqueira, E., and Braun, T. (2020). A multi-tier fog content orchestrator mechanism with quality of experience support. *Computer Networks*, 177:107288.

Vo, N.-S., Bui, M.-P., Truong, P. Q., Yin, C., and Masaracchia, A. (2020). Multi-tier caching and resource sharing for video streaming in 5g ultra-dense networks. *IEEE Communications Letters*, 7798:1–1.

Wang, S., Urgaonkar, R., Zafer, M., He, T., Chan, K., and Leung, K. K. (2019). Dynamic service migration in mobile edge computing based on markov decision process. *IEEE/ACM Transactions on Networking*, 27:1272–1288.

Xu, D., Li, T., Li, Y., Su, X., Tarkoma, S., Jiang, T., Crowcroft, J., and Hui, P. (2020). Edge intelligence: Architectures, challenges, and applications.

Yuan, Z. and Muntean, G.-M. (2020). Airslice: A network slicing framework for uav communications. *IEEE Communications Magazine*, 58(11):62–68.

Zhao, Z., Cumino, P., Esposito, C., Xiao, M., Rosário, D., Braun, T., Cerqueira, E., and Sargento, S. (2022). Smart unmanned aerial vehicles as base stations placement to improve the mobile network operations. *Computer communications*, 181:45–57.

Zhu, H., Cao, Y., Hu, Q., Wang, W., Jiang, T., and Zhang, Q. (2019). Multi-Bitrate Video Caching for D2D-Enabled Cellular Networks. *IEEE Multimedia*, 26(1):10–20.