UAV-Assisted Content Delivery in Intelligent Transportation Systems-Joint Trajectory Planning and Cache Management

Ahmed Al-Hilo, Moataz Samir, Chadi Assi, *Fellow, IEEE*, Sanaa Sharafeddine, *Senior Member, IEEE*, and Dariush Ebrahimi, *Member, IEEE*

***Abstract*— Unmanned Aerial Vehicles (UAVs) are gaining growing interests due to the paramount roles they play, particularly these days, in enabling new services that help modernize our transportation, supply chain, search and rescue, among others. They are capable of positively influencing wireless systems through enabling and fostering emerging technologies such as autonomous driving, vertical industries, virtual reality and so many others. The Internet of Vehicles is a prime sector benefiting from the services offered by future cellular systems in general and UAVs in particular, and this paper considers the problem of content delivery to vehicles on road segments with either overloaded or no available communication infrastructure. Incoming vehicles demand service from a library of contents that is partially cached at the UAV; the content of the library is also assumed to change as new vehicles carrying more popular contents arrive. Each inbound vehicle makes a request and the UAV decides on its best trajectory to provide service while maximizing a certain operational utility. Given the energy limitation at the UAV, we seek an energy efficient solution. Hence, our problem consists of jointly finding caching decisions, UAV trajectory and radio resource allocation which is formulated mathematically as a Mixed Integer Non-Linear Problem (MINLP). However, owing to uncertainties in the environment (e.g., random arrival of vehicles, their requests for contents and their existing contents), it is often hard and impractical to solve using standard optimization techniques. To this end, we formulate our problem as a Markov Decision Process (MDP) and we resort to tools such as Proximal Policy Optimization (PPO), a very promising Reinforcement Learning method, along with a set of crafted algorithms to solve our problem. Finally, we conduct simulation-based experiments to analyze and demonstrate the superiority of our solution approach compared with four counterparts and baseline schemes.**

***Index Terms*— Unmanned aerial vehicle (UAV), UAVs' trajectories, resource allocation, Deep Reinforcement Learning, Vehicular Ad-hoc Networks (VANETs).**

Manuscript received May 1, 2020; revised June 24, 2020; accepted August 13, 2020. Date of publication September 9, 2020; date of current version August 9, 2021. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and in part by Concordia University. The Associate Editor for this article was S. Garg. *(Corresponding author: Chadi Assi.)*

Ahmed Al-Hilo and Moataz Samir are with the CIISE, Concordia University's Gina Cody School of Engineering and Computer Science, Montreal, QC H3G 2W1, Canada.

Chadi Assi is with the CIISE, Concordia University, Montreal, QC H3G 1M8, Canada (e-mail: assi@ciise.concordia.ca).

Sanaa Sharafeddine is with the Department of Computer Science, Lebanese American University, Beirut 1102 2801, Lebanon.

Dariush Ebrahimi is with the Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada.

Digital Object Identifier 10.1109/TITS.2020.3020220

I. Introduction

A. Preliminaries and Motivation

THE demand for digital contents and advanced streaming services is experiencing unprecedented growth lately [1]–[3]. Hence, futuristic wireless networks, i.e. beyond 5G and 6G, desire high speed, low latency, massive capacity, ultra-reliable, and fit-for-purpose connectivity in order to fulfill the rigorous requirements needed to satisfy the market needs. Additionally, it is strongly envisioned that the next cellular generations will enable emerging technologies such as Internet of Connected Vehicles (IoCVs), Virtual and Augmented Reality (VR/AR), Vertical Industries, Telemedicine services and so many others. Side by side with cellular networks, IoCVs is foreseen as an integral component of wireless communication to deliver and distribute multimedia and other content types [4].

Meanwhile, the automotive industry is witnessing a qualitative shift, with its widespread adoption of technology and digitalization. This shift has enabled the internet of vehicles to provide a myriad of capabilities and services to vehicle users such as safer self-driving experiences and higher network qualities. Generally speaking, Vehicular Ad-hoc Networks (VANETs) admit two kinds of services precisely, safety and infotainment [5]. The latter type imposes massive data traffic volumes as it comprises Video on Demand, web browsing, 4K streams, VR, and so on. Such type of applications require low latency and seamless connectivity to achieve the desired Quality of Service (QoS) levels. Indeed, 5G (and beyond) technologies offer the right platform for the internet of intelligent vehicles to become a reality and therefore enable their anticipated revolutionary services. UAVs, owing to their agility and flexibility, can be deployed to assist a VANET infrastructure by providing vehicles' users with the same services (e.g., infotainment, road safety and assistance, content delivery, etc.) or could help vehicles when the infrastructure is not available [6]. As opposed to Base Stations (BS) and Road Side-Units (RSU), a UAV is mobile and, therefore, can follow moving objects, i.e. vehicles. A UAV can plan its trajectory to get closer to those vehicles that need to establish connections.

Armed with our previous contributions in [7] where buffer-equipped stationary RSU is proposed, in this work we propose to further enhance the experience of content serving in the context of intelligent vehicular network by leveraging content-bearing UAVs. The network operator dispatches a UAV or multiple to help serving end users in vehicles demanding contents from the network/content provider. We can assume each UAV is augmented with a cache that enables it to store, carry and deliver contents to the vehicles. Caching is shown to have good impact on the network by reducing the access on the backhaul [8]–[10] through averting repeated fetching for similar contents and remarkably lowering latency down. Now, unlike User Equipment (UEs) such as smartphones which are limited by size and capacity, vehicles have sufficient energy and they can carry large components, i.e. cache and processing units. It is anticipated that new smart vehicles, especially those which are equipped with OnBoard Units (OBUs), would have caching capabilities. Hence, when contents are downloaded, it is preferred to keep them in their caches for a little while to prevent redundant requests. In fact, the prevailing trend of vehicle users is to visit several zones in their journey. Sometimes, they send requests for diverse contents from different entities, e.g., BS, RSU, Access Points, and others. A UAV can make profit out of this by harvesting these contents from the passing by vehicles in order to serve other vehicles which demand similar contents.

Serving contents to non-stationary entities through cache-augmented UAVs given the uncertainties of the traffic and requests remains unaddressed [11]. The existing works have either overlooked caching techniques to offload traffic or the energy consumption incurred by the UAVs movement. Therefore, in this paper, we aim at addressing this problem and solve it efficiently leveraging recently developed machine learning tools.

B. Challenges

In contrast to stationary entities such as terrestrial base stations, a UAV can control its trajectory. However, this mobility comes with a set of pros and cons. On one hand, the UAV is able to move closer to the desired vehicles to create stronger connection with higher data rates. On the other hand, as the UAV moves towards a vehicle, it may become distant from others. Hence, the pre-scheduled transmission process may falter if the process is still incomplete. Besides, the UAV consumes different amounts of energy based on its velocity. Thereby, the UAV should plan its trajectory taking into account the mobility of vehicles and the total energy consumption in order to avoid fast battery depletion and wasting its resources. Hence, finding a good UAV trajectory to deliver contents to existing and newly arriving vehicles is not a straightforward problem. To further complicate this problem, caching decisions can be difficult to make since the cache capacity is limited while the library of contents is relatively large, and the UAV is completely oblivious to vehicles' arrivals and their demands for contents.

C. Contribution

In this work, we design a platform to assist content delivery in vehicular networks. We can summarize our novelties and contributions as follows:

- We present a UAV caching based system where the UAV helps vehicle users to download their desired contents while it makes decisions to fetch contents for its own cache from the same vehicles. The proposed system also considers the limited amount of energy on the UAV and, thus, it aims at striking a balance between the traffic offloaded and energy consumption. Then, we formulate the problem of UAV trajectory, radio resources, and caching replacement mathematically as an optimization problem to find a suitable trajectory to maximize the energy efficiency of the UAV.
- Next, owing to the complexity of the addressed problem and ambiguity of input parameters, we proposed PPO algorithm towards solving the aforementioned issue of UAV mobility in an efficient manner.
- Finally, to answer the challenges of uplink and downlink resource allocation, we develop two effective yet light designed algorithms to schedule for wireless resources. Additionally, we conduct extensive tests to prove their efficiency and compatibility with PPO model.

D. Paper Structure

The paper is structured as follows. Section II presents the related works, while section III explains our system model. In Section IV, we formulate our problem mathematically. Section V presents our solution approach in details. The numerical results and analysis are shown in section VI. Finally, section VII, the conclusion wraps up the paper with directions for future work.

II. Related Work

UAVs have been widely studied and they attract the telecommunication community. They have been used in the literature as cache units, flying base stations, and relay nodes. Despite the numerous amount of works in this domain, only few studies exist that address the problem of caching UAV to serve highly dynamic requester objects.

The study of [12] investigates into the performance of UAVs which cache, relay, and serve multimedia contents to IoT devices. The authors of this work solve the joint problem of UAVs placement and content caching to optimize wireless throughput. First, as apposed to our work, the entertainment large contents are served to stationary IoT devices. Thus, dealing with fixed or low mobility environment is much easier than dynamic one. Second, the contents are assumed of similar size and the UAV is assumed to cache contents proactively based on the popularity and other factors. While we address a different scenario where the UAV has no access to the backhaul and, hence, it needs to fetch the contents from the passing by vehicles. In [10], a proactive caching scheme is studied where a UAV is dispatched to provide content delivery service for vehicle users in a certain area. The objective of this work is to find UAV trajectory besides content placement and delivery mechanism to optimize the overall UAV throughput taking into account its limited battery capacity. In [9], the authors present a networking framework consisting of caching UAVs which store popular contents in advance to alleviate heavy traffic on the congested backhaul. The UAVs then serve mobile users through their cache units instead of fetching the contents from the internet. The authors of [13] suggest a new architecture for UAV to deliver content in vehicular environments. Their scenario does not consider caching or library of contents, but rather one content which is requested by multiple vehicles. In [14], the authors present a solution for the joint problem of caching and resource allocation to serve fixed users from the cache unit or via server-UAV-user link. They leverage liquid state machine (LSM) to realise the content popularity distribution. The work of [11] tries to improve the QoE of wireless devices through caching UAV. As a solution, they leverage the history and information of user in order to find out his request patterns. In [15], authors use UAV to provide coverage for vehicles in a particular area where the infrastructure becomes out of service due to disaster situations. This work aims at satisfying certain QoS to the end users. The authors of [16] utilize UAVs to help the infrastructure operation taking into account the cost incurred by deploying more UAVs. Thus, it optimizes the number of UAVs dispatched in order to provide coverage for a certain region. Moreover, in [17], the authors study the communication between a swarm of UAVs moving at fixed velocity with passing vehicles and they model the average data packet delivery delay in such scenario. In addition, [18] proposes deployment of UAVs to mitigate security problems in vehicular networks where the vehicles communication with the RSU is jammed. The UAV is leveraged to relay contents or messages in case the correspondent RSU is being attacked or there is high link interference, hence, the link is obscured.

Alternatively, a plenty of works have been devoted for static infrastructure for improving content delivery in vehicular networks. The study of [19] suggests cooperative caching among a set of RSUs by replicating contents. The incentive behind this work is that RSUs located at the same roads are highly correlated during the service of the traffic flow. Furthermore, [20] studies the stochastic delay of content delivery in cache-based system in vehicular networks. Here, a stochastic network calculus is employed to evaluate the stochastic playback delay upper bounds of vehicular video content delivery with cache-enabled RSUs. In [21], authors proposed a caching system to deliver large contents such as videos to fast moving vehicles. They addressed the efficient content delivery problems in VANET by caching popular files in the RSUs with large storage capacity. In [22], the authors solve the problem where multiple content providers (CPs) aim to improve the data dissemination of their own contents by utilizing the storage of RSUs. They used multi-object auction-based solution to get a sub-optimal solution for the competition among the CPs. Also, [23] proposed a Q-learning based caching strategy where the system predicts the the path trajectory of vehicles and based on that it decides which RSUs should cache the contents in order to reduce the latency. Some works such as [24], suggest cooperation between RSUs and vehicles where a model is proposed to determine from where to obtain the content. Moreover, the work of [25] suggests caching using different types of infrastructure entities to minimize the total delay of content delivery.

TABLE I
LIMITATIONS OF EXISTING WORKS

Paper	Caching	Dynamic Environment	Energy consumption
[12]	×	✓	×
[10]	✓	✓	×
[9]	✓	×	✓
[13]	×	✓	×
[14]	✓	×	×
[11]	✓	×	✓
[15]	×	✓	×
[16]	×	✓	×
[17]	×	✓	×

The aforesaid works have addressed the problem of content delivery through leveraging UAVs and fixed entities, however, the existing works either suggest proactive caching mechanisms (which depends on a previous knowledge regarding the environment such as request for contents and arrival rate of requesters which, most of the time, may not be provided beforehand.) or utilize the congested and hard-to-implement backhaul link [10]. Assuming a UAV with sufficient backhaul link capacity to fetch, cache, and serve looks impractical scenario for the case of UAVs as they move and, hence, cannot maintain stable connection to the backhaul especially when they are required to follow and serve moving objects such as vehicles. Most importantly, the energy consumption of the UAV was overlooked in the previous works with caching. We list the key limitations of the existing related work on UAV servicing users in table I.

Finally, none of these studies has looked into how cache-equipped UAVs can offer communication and content delivery services for dark areas, which are commonly assumed to be covered through UAVs, where no internet connection is available.

III. System Model

Consider a one-way highway segment of a certain length (G) where a standalone UAV is utilized to provide content delivery service for the visitor vehicles as illustrated in Fig 1. The highway segment is presumed to entrust the UAV to respond to the content requests with a cache unit mounted on the UAV having limited space (η) to fetch, buffer, and relay contents. Meanwhile, a set of vehicles (I) travels through this particular highway where each vehicle $i \in I$ buffered a content before they approach this highway segment. The incoming vehicles are also in need of contents while they are within the highway. Assume each vehicle has one content to request and this content can be downloaded while the vehicle is within the highway segment. Therefore, we assume that each requested content can tolerate time not less than that taken by a vehicle to cross the highway.

The UAV is dispatched to carry out this operation for a certain amount of time (N). The service time consists of several time slots ($n = 1, \ldots, N$) of length δ. Next, the details of each aspect of the system model is given in a subsection.

Remark: In this work, for tractability, we deal only with one UAV. However, the same approach may apply for several

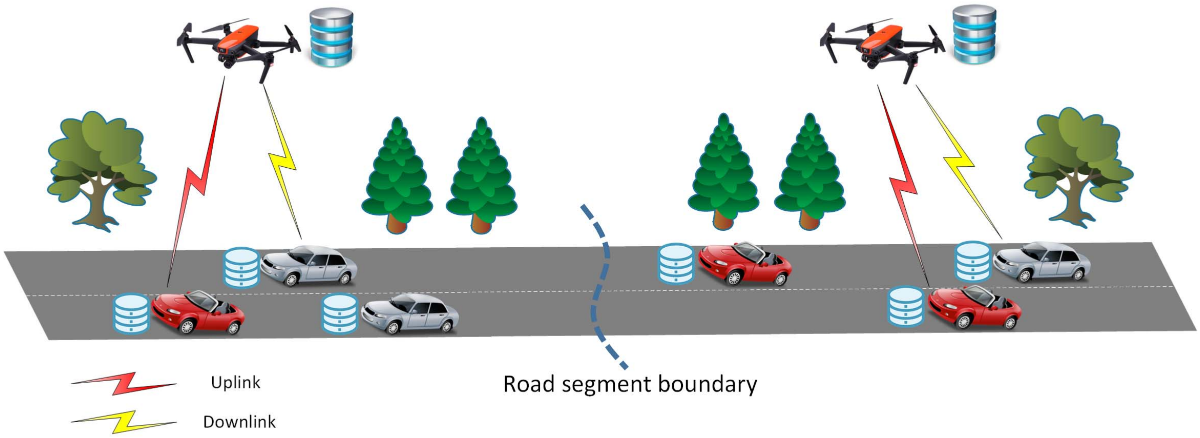

Fig. 1. System model.

UAVs covering a certain highway. In this context, the overall network performance could be notably improved since multiple UAV are much powerful and resource-rich to serve vehicles than a solo UAV. Given that scenario, the highway can be fragmented into smaller segments where there is one UAV dedicated to serve each segment separately as in Fig. 1. The first UAV, which is located at the beginning of the highway, may serve some of the vehicles while the remaining unserved ones will probably receive service through the consequent UAVs. The adjacent UAVs may operate on different wireless spectrum and, hence, the interference could be neglected. Interestingly, as multiple UAVs are deployed, one can leverage cooperation among them to improve cooperation among them. For instance, UAVs covering adjacent segments may get closer to each other in order to establish wireless communication and transfer contents between them. However, this is beyond the scope of this paper paper and is kept for future work.

A. U2V and V2U Communication

We assume the communication between UAV and vehicles is established with orthogonal time division multiplexing access (TDMA). Furthermore, we assume the communication is full-duplex and over two different spectrum. For generality, we further assume the transmission power of vehicles and UAV are constant, thus, the uplink and downlink data rates (U_i^n and D_i^n respectively) vary with the distance only. At each time slot n, the UAV is either fetching a content from a vehicle, and/or forwarding a content to a vehicle, or idle.

B. Content Model

The library of contents encompasses a number of items (J) which have different sizes ($Z_0, \ldots, Z_j, \ldots, Z_J$). The probability of content being requested is driven from Zipf distribution where the skewness parameter α_1 characterizes this distribution as in Eq(1).

$$\frac{\frac{1}{k^{\alpha_1}}}{\sum_{j=0}^{J} \frac{1}{k^{\alpha_1}}}, \tag{1}$$

where k is the item rank.

The same type of distribution models the probability of content being buffered on a vehicle, yet, with different skewness parameter α_2. Moreover, for the convenience of exposition, all contents are assumed to tolerate time larger than the travelling time of vehicles to cross the highway to avoid having varying request deadlines. Also, we assume that contents do not allow partial service. That means, in order to serve a content, it has to be fully downloaded, else, it is considered corrupted, hence, it does not count similar to [26], [27].

C. Traffic Model

To model the arrivals and velocities of passing vehicles, we assume a free-flow model similar to [28] where the relationship between velocity and traffic can be expressed as $v = v_f(1 - \rho/\rho_{jam})$ where v_f is the expected velocity, ρ is the traffic density, i.e., vehicle(s) per Km, and ρ_{jam} is the jam density (jam density is extreme traffic density when traffic flow stops completely). Moreover, the arrival rate of vehicles is calculated by $\lambda = \rho v$. The vehicle arrivals follow a Poisson distribution and velocities are generated using truncated normal distribution.

For a certain period of time (N), there is a subset of $I^n \subset I$ vehicles coming where $I = \{I^0 \cup I^1 \cup I^2 \ldots \cup I^N\}$. Each vehicle $i \in I^n$ has a certain velocity and arrival time (a_i) as mentioned before. Also, d_i represents the departure time of vehicle i.

D. Operation Phase and Objective

Initially, we assume the UAV is positioned at the middle of the highway segment ($w^0 = \frac{G}{2}$). When the UAV moves, it consumes energy and the amount of energy consumed with each move depends on the velocity selected at that time slot n. In practice, UAV has a maximum velocity of V_{max} which should not be violated. Also, the energy consumed for movement is modeled as non-linear function with distance traveled [29]. It must be stressed that since the energy consumed for operating the UAV is much larger than that of serving vehicle (transmission power), we only take the former into account. Besides considering energy consumption, the UAV needs to decide for the following actions:

- UAV trajectory (w^n).[1]
- Serve content to a vehicle (x_i^n)
- Fetch content from a vehicle (y_i^n).
- Caching decision or replacement policy (f_j^n).

In case that the UAV is unable to serve or fetch a content completely before the vehicle leaves, the process must be terminated.[2]

Finally, the ultimate objective of utilizing the UAV is to provide service for a specific region. However, exploiting UAV comes with costs due to energy consumption. UAVs are known to be energy constrained, therefore, the UAV needs to operate smartly such that it makes the most of its energy and cache capability. In this work, we address the trade-off between content delivery and energy consumption by optimizing energy efficiency. Here, the energy efficiency is defined as the power required to transmit one data unit or the number of bits that can be sent over a unit of power consumption which is usually quantified by bits per Joule [30].

E. Problem Definition

Definition 1: Given a set of vehicles travelling through a highway covered by a UAV. Assume the UAV travelling velocity affects its energy consumption, also, the UAV is equipped with a finite cache unit. Assume the coming vehicles have one content in their buffers and raise a request for another content. Assume that the UAV has a particular service time consists of several time slots where at each time slot, the UAV can move or hover, besides, it can serve and/or fetch contents from vehicles at different data rates based on the distance. In the light of the foregoing, what is the optimal movement, serving, fetching, cache replacement actions for the UAV such that the energy efficiency is maximized.

TABLE II

MATHEMATICAL NOTATION

Parameters	
Z_j	content size
G	highway length
J	list of contents
I	set of vehicles
N	time horizon
a_i	arrival time of vehicle i
d_i	departure time of vehicle i
L_i^n	location of vehicle i at time n
$R_{i,j}$	1: if i requests content j and 0 otherwise
$C_{i,j}$	1: if i cached content j and 0 otherwise
V_{max}	UAV maximum velocity
η	UAV cache capacity
δ	length of one time slot
τ	path loss exponent
h_0	median of the mean path gain at reference distance = 1m
σ	thermal noise power
h_i^n	channel gain between the UAV and vehicle i at time slot n
H	Fixed altitude of the UAV.
$P_{V\to U}^{i,n}$	Received power of vehicle i from the UAV at time slot n.
$P_{U\to V}$	Received power of the UAV from vehicle i at time slot n
P_V	Transmission power of vehicles.
P_U	Transmission power of the UAV.
D_i^n	Immediate data rate to vehicle i from the UAV at time slot n
U_i^n	Immediate data rate to the UAV from vehicle i at time slot n
W	Channel bandwidth available.
δ	Time slot size.
K	Constant of the UAV dimension blade.
F	The UAV drag coefficient.
π	Air density.
ν^n	Velocity of the UAV in time slot n.
Variables	
f_j^n	1: if UAV removes content j at n and 0 otherwise
x_i^n	1: if UAV serves vehicle i at n and 0 otherwise
y_i^n	1: if UAV fetches from vehicle i at n and 0 otherwise
w^n	location of UAV at n
k_j^n	1: if UAV cached content j at n and 0 otherwise
Q_i	1: if UAV served vehicle i sufficiently and 0 otherwise

IV. PROBLEM FORMULATION

This section formulates the system model described above mathematically. To make it clearer and more organized, we categorize the different aspects of the system. Additionally, Table II provides a summery of the variables and parameters used in the formulation.

A. Wireless Communication

Typically, the communication channel between UAVs and end users is usually modeled as large- and small-scale fading. However, in the context of highways where there exists a clear link between transmitter and receiver, the communication link can be characterized by strong line-of-sight and, hence, small and large scale fading can be omitted [15], [16]. Thereby, the channel gain between the UAV and vehicle i at time slot n can be written as follows:

$$h_i^n = h_0\left(\sqrt{(L_i^n - w^n)^2 + H^2}\right)^{-\tau}, \quad \forall i \in I^n,\ \forall n \tag{2}$$

[1]Since this work concentrates mainly on improving content delivery via caching UAV and for the sake of simplicity, we assume that the altitude of the UAV is fixed.

[2]We assume once the UAV runs out of energy another one is dispatched caching similar contents of the last UAV.

where h_0 is the mean path gain at reference distance = 1m. H denotes the altitude of the UAV and τ denotes path loss exponent.

Let us define $P_{V\to U}$ and $P_{U\to V}$ to denote the received power of UAV and vehicles respectively. Now, $P_{V\to U} = h_i^n P_V$ and $P_{U\to V} = h_i^n P_U$ where P_V and P_U are constant values representing the transmission power of vehicles and the UAV respectively.

Now, let us introduce variables D_i^n and U_i^n which are the instantaneous data rates for down and uplink scheduled to the passing vehicle i at time slot n respectively and they are calculated as follows:

$$D_i^n(x_i^n, w^n) = \begin{cases} x_i^n W \log_2\left(1 + \dfrac{P_{U\to V}^{i,n}}{\sigma^2}\right) & \text{if } a_i < n < d_i, \forall i \in I^n, \forall n, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

$$U_i^n(y_i, w^n) = \begin{cases} y_i^n W \log_2\left(1 + \dfrac{P_{V\to U}^{i,n}}{\sigma^2}\right) & \text{if } a_i < n < d_i, \forall i \in I^n, \forall n, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where W is the available channel bandwidth (in Hz). σ is the thermal noise power which is linearly proportional to the allocated bandwidth. a_i and d_i represent the arrival and departure time of vehicle i, respectively.[3] $x_i^n \in \{0, 1\}$ is a decision variable to allocate downlink resources to vehicle i. If $x_i^n = 1$ then vehicle i is receiving content from the UAV at time slot n. $y_i^n \in \{0, 1\}$ is also a decision variable to schedule uplink resources to vehicle i at time slot n. Since we assume Time-Division Multiple Access (TDMA) as a channel access method for the UAV and vehicle links, then only one transmission is allowed per time slot:

$$\sum_{i=0}^{I^n} x_i^n \leq 1, \quad \forall n. \tag{5}$$

$$\sum_{i=0}^{I^n} y_i^n \leq 1, \quad \forall n. \tag{6}$$

B. UAV Mobility

The maximum distance the UAV can pass during one time slot should not exceed its maximum velocity.

$$|w^n - w^{n+1}| \leq V_{max}\delta, \quad \forall n. \tag{7}$$

In addition, the movement of the UAV incurs energy consumption which depends on the velocity of the UAV similar to [16], [29].

$$P(v^n)_{total} = \underbrace{K\left(1 + 3\frac{M_{tip}^2}{w_b^2}\right)}_{\text{Blade profile power}} + \underbrace{\frac{1}{2}\pi (v^n)^3 F}_{\text{Parasite power}} + \underbrace{m_u g\sqrt{\left(\frac{-(v^n)^2 + \sqrt{(v^n)^4 + (\frac{m_U g}{\pi A})^2}}{2}\right)}}_{\text{Induced power}}. \tag{8}$$

where v^n denotes UAV velocity at time slot n. M_{tip} represents the blade's rotor speed, K and F are two constants which depend on the dimensions of the blade and the UAV drag coefficient, respectively, π is the air density, m_U and g respectively denote the mass of the UAV and the standard gravity, A is the area of the UAV. The total energy consumption to cover a distance d at a constant velocity UAV w can be computed as $E(v)_{total} = \int_0^{d/v} P(v)dt = P(v)\frac{d}{v}$ as in [16].

C. Cache Management

Let us define $\beta_i^n \in \{0, 1\}$ as an indicator holding value 1 if the content available on vehicle i is fully fetched by the UAV before or at time slot n and 0 otherwise.

$$\beta_i^n = \begin{cases} 1 & \sum_{n'=0}^{n} \delta U_i^{n'} \geq \sum_{j=0}^{J} C_{i,j} Z_j, \forall i \in I, n \in N. \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

In addition, Eq (10) prevents wasting radio resources to vehicle that completely uploaded its content to the UAV.

$$y_i^{n+1} \leq 1 - \beta_i^n, \quad \forall i \in I, \ n \in N. \tag{10}$$

Next, we introduce $k_j^n \in \{0, 1\}$ as a binary variable where it is equal to 1 if content j is available on the UAV at time slot n and 0 otherwise. The value of k has three cases; it is either equal to 1 when the content is just fetched or 0 if it is removed in the previous time slot. The third case is that it remains the same if no change occurs on its value.[4]

$$k_j^n(y_i^n, f_j^n) = \begin{cases} 1 & \sum_{i=0}^{I} y_i^n \beta_i^n C_{i,j} = 1, \forall n \\ 0 & f_j^{n-1} = 1, \\ k_j^{n-1} & \text{otherwise,} \end{cases} \tag{11}$$

where $f_j^n \in \{0, 1\}$ is a decision variable where it equals to 1 if content j is removed from the UAV cache at time slot $n + 1$, 0 otherwise.

Eq(12) prevents violating the limited cache capacity of the UAV.

$$\sum_{j=0}^{J} k_j^n Z_j \leq \eta, \quad \forall n. \tag{12}$$

A content cannot be removed from the UAV cache if it does not exist in the cache beforehand.

$$f_j^n \leq k_j^n, \quad \forall j \in J, n. \tag{13}$$

D. Service Management

Before we formulate our objective function mathematically, let us define $Q_i \in \{0, 1\}$ which denotes whether a vehicle has been sufficiently served or not during its sojourn time.

$$Q_i(x_i^n, y_i^n, w^n, f_j^n) = \begin{cases} 1 & \sum_{n=0}^{N} \delta D_i^n \geq \sum_{j=0}^{J} R_{i,j} Z_j, \forall i \in I^n, n, \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where $R_{i,j}$ indicates whether content j is requested by vehicle i ($R_{i,j} = 1$) or not ($R_{i,j} = 0$).

In fact, the UAV cannot serve a vehicle if the requested content is not available in its cache as illustrated in Eq (15).

$$x_i^n \leq \sum_{j=0}^{J} R_{i,j} k_j^n, \quad \forall i \in I^n, n. \tag{15}$$

E. Objective

The ultimate objective is to maximize energy efficiency (bits per Joule) and can be written as:

$$\begin{aligned} \max_{w^n, x_i^n, y_i^n, f_j^n} \quad & \sum_{j=0}^{J} \sum_{i=0}^{I} \frac{Q_i(x_i^n, y_i^n, w^n, f_j^n) Z_j R_{i,j}}{\sum_{n=0}^{N} P(w^n)}, \\ \text{s.t.} \quad & Eq(5), Eq(6), Eq(7), Eq(10), \\ & Eq(12), Eq(13), Eq(15), \end{aligned} \tag{16}$$

where Q_i is function of the four decision variables.

The presented problem is mixed integer non-linear programming (MINLP), which is known to be difficult to solve,

[3]The instantaneous rate depends on the instantaneous location of the vehicles. However, we are interested in the instantaneous rate of vehicles present within the highway segment, therefore, for more tractable analysis, the instantaneous rate is reduced to zero outside the given highway segment.

[4]For more tracebility, the initial value of $k_j^0 = 0, \forall j \in J$ since the cache is assumed empty at the beginning.

owning to the binary variables, x_i^n, f_j^n, y_i^n and the real-value decision variable w^n, as well as, our objective function in Eq (16) is non-convex because of the trajectory variable in Eq (7) [31]. Furthermore, the solution of the problem (if exists) is dependant of the knowledge of instantaneous positions of vehicles at each time slot during their journey along the highway segment. However, in practice, the upcoming set of vehicles alongside their information (their requests and available contents in their buffers) are revealed once they approach the highway segment. Particularly, the values of parameters L_i, $R_{i,j}$, and $C_{i,j}$ remains hidden as long as vehicle i has not yet reached the highway segment. Given the high complexity of the problem and the numbers of uncertainties, definitely there is a need for alternate solution with lower complexity and high efficiency to tackle such scenario [32]. Therefore, we model our problem as MDP and, then, we propose PPO-Clip, which we will introduce next, and effective algorithms to solve for the aforementioned problem. PPO-Clip is efficient to deal with ambiguities in the environments as it is able to learn and estimate values through observations. This MDP model and PPO-Clip will be laid out in the next section.

V. Solution Approach

This section provides a complete solution for the predefined problem of UAV mobility and caching. Due to the intractability of the optimum problem and in order to simplify the arduous challenges addressed by this paper, we divide our problem into three sub-problems. This first one is to find trajectory of the UAV while the other sub-problems deal with wireless transmission and caching replacement.

A. PPO-Clip to Control UAV Trajectory

In this work, we resort to Deep Reinforcement Learning to solve for the trajectory decisions. Many applications such as video games, autonomous vehicles, UAVs, and so forth, encompass uncountable number of states and actions. These high-dimensional spaces create phenomena of curse of dimensionality. Here, neural networks come in handy where Deep Learning techniques are leveraged alongside RL in order to optimize the number of states and actions through approximation [33]. Precisely, we leverage Proximal Policy Optimization (PPO). PPO is a cutting-edge, benchmark, model-free, on-policy, policy gradient Reinforcement Learning algorithm designed at OpenAI [34], [35]. It is famous for its tunability besides its outstanding performance and lower complexity. In fact, the goal of PPO is to balance between implementation, batch sampling efficient, and ease of tuning.

There are two types of PPO: PPO-Penalty and PPO-Clip. The former uses Kullback–Leibler (KL) divergence to update policies while the latter, which we use, relies on a particular clipping technique in the objective function [34]. Genuinely, PPO-Clip approximates the hard constraint applied to PPO-Clip by using much more effortless equations. Therefore, PPO-Clip is much simpler version and demonstrates remarkable efficiency.

To go for the implementation part, we first formulate our problem as MDP which is denoted by a 5-tuple $\langle \mathcal{S}, \mathcal{A}, \gamma, \mathcal{R}, \mathcal{P} \rangle$ where:

- $\mathcal{S}$ is the state space where the size of the state space is very large. Each state $s_n \in \mathcal{S}$ is a vector containing several parameters which represent the current conditions of the highway and the UAV. Namely, the state includes current position of the UAV w^n, information of each vehicle existing at that time slot, and cached contents. The second part, which indicates that vehicles, has 7 values, namely, position of the vehicle (L_i), available content ($j \mid C_{i,j} = 1$), amount fetched D_i^n, requested content ($j \mid R_{i,j} = 1$), amount served U_i^n, sizes of contents buffered and requested (Z_j), and the cached contents on the UAV (k_j^n). Overall, state s contains 8 parameters as follows: $s_n = (w^n, D^n, U^n, L^n, C^n, R^n, Z, k^n)$ where D^n, U^n, L^n, C^n, R^n represent amounts served, amounts fetched, location of vehicle, requests, and availabilities, respectively, for every vehicle $i \in I^n$. Z is the content sizes and k^n denotes which contents are cached at time slot n.
- $\mathcal{A}$ is the action space where it includes all the feasible actions ($a \in \mathcal{A}$) that can be taken by the agent. In this work, in order to void the high complexity of continuous action space, we approximate the action space by descritizing the UAV velocities. Thus, our action space includes 0 velocity for hovering and a certain number of velocities to step the UAV forward and backward and those velocities are predetermined.
- $\mathcal{P}$ is the state transition probabilities. Which denotes the probability of being in s' after executing action a in state s, $Pr(s_{n+1} = s' \mid s_n = s, a_n = a)$. There is a certain probability to move from one state to another. This transition depends on the current state only and can be written as follows:

$$\begin{aligned} Pr(s_{n+1} \mid s, a_n) = \; & Pr(w^{n+1} \mid w^n, a_n) \\ & \times Pr(D^{n+1} \mid D^n, x_i^n, w^n) \\ & \times Pr(U^{n+1} \mid U^n, y_i^n, w^n) \\ & \times Pr(k^{n+1} \mid k^n, f_j^n, R^n, U^n, y_i^n) \\ & \times Pr(L^{n+1} \mid L^n) \times Pr(C^{n+1} \mid C^n) \\ & \times Pr(R^{n+1} \mid R^n). \end{aligned} \tag{17}$$

Particularly, the transition from one state to another is driven from the following. The probability of UAV being in w^{n+1} depends on the UAV previous location and the action a_n. The probability of vehicles being served by amount D^{n+1} depends on the current served amounts, plus, the downlink resource allocation decision and UAV position. Likewise, the probability of UAV being received amounts U^{n+1} of vehicles' contents depends on the current amount fetched with uplink resource allocation decision and UAV position. Next, the probability of content being cached (k^{n+1}) on UAV relies on the present cache status, content replacement decision, amount fetched, and uplink decision variables. Moreover, the probabilities of vehicles being in location L^{n+1}, requesting contents

(C^{n+1}), and having contents (R^{n+1}) rely only on the current status. The content sizes have no impact on this transition as they are constant.
- γ is the discount factor where $0 \leq \gamma \leq 1$.
- $\mathcal{R}$ denotes the discounted cumulative reward which is the product of summing all the discounted immediate reward (step-reward) r_n as illustrated in Eq(18).

$$\mathcal{R} = \sum_{n=0}^{N} \gamma^{n-1} r_n. \tag{18}$$

Worth mentioning that the objective function in Eq (16) emphasizes the total reward is a product of the division of total amount served over energy consumption. However this cannot be realised with RL. As RL calculates the immediate reward in each step, the total energy consumption value remains unknown and only reveals once the episode ends. Thereby, we modify our immediate reward to be as follows:

$$r_n = \sum_{j=0}^{J} \sum_{i=0}^{I} Q_i^n Z_j R_{i,j} - \psi P(w^n), \quad \forall n, \tag{19}$$

where ψ is a weighting factor to balance the impact of the two terms. In Eq (19), to solve the aforesaid issue, instead of dividing the amount served over energy consumption, we subtract them. However, in order to not neglect the impact of any terms, we add ψ to tackle this issue. Now, the step-reward contains two parts; positive reward and negative reward (penalty). The positive reward is awarded to the agent when the UAV serves a vehicle sufficiently while the agent is penalized due to the energy consumed with each UAV move.

As illustrated in algorithm 1, the PPO interacts with the environment to collect the samples through several iterations (typically thousands of epochs) and realises the actual rewards.

First the PPO initializes random sampling policy and value function for the neural networks as in lines 3 and 4. Then, in each epoch, the agent observes the environment which consists of the set of vehicles and their availabilities, requests, cached contents, UAV position and so on as in lines 7. Then, an action is selected based on the policy as line 8 and the action is used to move the UAV to its new position, however, the UAV should remain inside its service region (lines 9 to 12). In line 13 to 25, Algorithm (3) and (2) are used to realise the resource allocation among the set of vehicles present at that time slot. For the downlink, the UAV computes a positive reward if a content is served to a vehicle. In the uplink, the UAV stores a content in the cache once it is fully fetched. A replacement may be required if there is no space available to fit the new content.

After gathering the set of samples and computing the rewards (lines 27-28), next, PPO finds out the advantage function (line 29), $\hat{A}_n$, which is defined as the resultant of subtracting the expected value function from the actual reward. Here, $\hat{A}_n$ is the estimated advantage function or relative value of the selected action. $\hat{A}_t$ helps the system to understand how good it is preforming based on its normal estimate function value [35].

In PPO, similar to Trust Region Policy Optimization (TRPO), the policy update is done only if the difference between the old policy and the new policy does not exceed certain threshold. In line 9, there are two terms in the minimization. First term is the difference between the two policies, new and old where $r(\theta) = \pi_{new}(a_n, s_n) - \pi_{old}(a_n, s_n)$. The second term is the clip function output. Here, ϵ denotes the threshold and clip cancels out the incentive if r_t becomes out side $[\epsilon + 1, \epsilon - 1]$ where ϵ equals to 0.2 as recommended. As stated, the policy is updated if the advantage function is within reasonable value. If A is very large, it means the action becomes much more likely after the last step of gradient ascent and, hence, it might become out of its trusted region, so we should not update to avoid having worse results in the future. In addition, the same applies if the value of A becomes much less probable in the current step (lines 30-32).

Lastly, the PPO optimizes the mean-squared error via gradient descent as in line 32. PPO usually is implemented in Actor-Critic framework, where more objective functions are added to the surrogate objective. Based on [36], the complexity of connected network with P layers is $O(\sum_{p=0}^{P} n_p n_{p-1})$ where n_p denotes the total number of neurons in layer p.

B. Heuristic Algorithms to Wireless Resource Allocation

The second sub-problem of UAV and vehicle radio resources scheduling is solved by two algorithms for the up and downlink.

Concerning the uplink, as laid out in Algorithm 2, the UAV sorts the contents cached based on their popularities. Next, it checks whether the content available on each present vehicle is already cached or not and whether it is downloaded earlier. If the conditions above are satisfied, then the UAV makes sure the content size will only replace content(s) which are less in size. In this step, the UAV needs to ensure that it will not require to remove contents which are more popular for those which receive less hits. The complexity of Algorithm 2 is $O(\partial J)$ where ∂ represents the maximum number of vehicles present simultaneously in the highway segment. The value of ∂ depends on the highway density and, in reality, $\partial \ll I$.

For the downlink, the UAV is committed to serve a vehicle as long it resides within the highway and not fully served. The UAV also need to make sure that the distance between the UAV and the vehicle does not exceed a certain threshold (Ω) as in Algorithm 3. This condition can improve the download experiences since it prevents the UAV from sticking to only one far vehicle which will not be served after all due to far distance. In case no vehicle is being served at that time, the UAV will compare among the available vehicles based on data rates. The vehicle with higher data rate or shortest distance will be selected to download. The complexity of Algorithm 3 is $O(\partial)$.

C. Cache Replacement

The cache replacement policy used in this work is as follows. Whenever the cache is full and a new content is fetched, the UAV starts freeing up space by removing contents

Algorithm 1 PPO-Clip to Find UAV Trajectory

Inputs: L_i^n, $C_{i,j}$, R_{i_j}, N, Learning Rate, γ, ϵ.
Outputs: The UAV velocity control policy π.
Initial policy π with random parameter θ
initial value function V with random parameters ϕ
for each episode $k \in \{0, 1, 2, \ldots\}$ **do**
  for $n : \{0, 1, 2, \ldots, N\}$ **do**
    Observe state $L_i, C_{i,j}, R_{i,j}, U_i^n, D_i^n, \forall i \in I^n, Z_j, w^n$.
    Perform action $a_n \in A$ based on policy π_k.
    Set $W^{n+1} = W^n + a_n$
    if UAV is outside the highway segment **then**
      Keep the UAV at the previous position.
    end if
    Set $x_i^n \in \{0, 1\}$ using Algorithm (3)
    if vehicle i content is completely served **then**
      Calculate step reward as in Eq (19).
    else
      $r_n = 0$
    end if
    Set $y_i^n \in \{0, 1\}$ using Algorithm (2)
    if vehicle i content is fully received by the UAV **then**
      if vehicle i content needs space **then**
        Remove other content(s) as in V-C.
      end if
      Store vehicle i content in the UAV cache.
    end if
  end for
  Collect set of samples $\mathcal{D}_k$ by running policy π_k in the environment.
  Compute reward $\hat{R}_n$.
  Compute advantage estimates, $\hat{A}_n$, which is $\hat{R}_n - V_{\phi k}$.
  Update the policy by maximizing the PPO-Clip objective:

$$\theta_{k+1} = \arg\max_{\theta} \theta \frac{1}{|\mathcal{D}_k| N} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{N} \min \Big(r_n(\theta) \hat{A}^{\pi \theta k}(s_n, a_n), \text{clip}\big(r_n(\theta) \hat{A}^{\pi \theta k}(s_n, a_n), \epsilon + 1, \epsilon - 1\big)\Big) \tag{20}$$

  Fit value function by minimizing mean-squared error, as:

$$\phi_{k+1} = \underset{\theta}{\text{argmin}}\ \theta \frac{1}{|\mathcal{D}_k| N} \sum_{\tau \in \mathcal{D}_k} \sum_{n=0}^{N} \big(V_\phi(s_n) - \hat{R}_t\big)^2 \tag{21}$$

end for

Algorithm 2 Uplink Radio Resource Allocation

Inputs: n, Z_j, $C_{i,j}$.
Outputs: Π as the vehicle to fetch from at n.
Set $\Pi = None$.
Sort cache items by popularity
for $\forall i \in I$ where $a_i < n < d_i$ **do**
  if i's content does not exist in the RSU cache and the content is not fully fetched yet **then**
    Compute the total size of cached contents on the UAV.
    if the vehicle i content cannot fit in the cache **then**
      Define $R = 0$ as the amount to remove in order to cache the new content.
      for each content $j : k_j^n = 1$ **do**
        if vehicle i content is more popular that of Π and j **then**
          Set $R = R + Z_j$
          if vehicle i content size $\leq R$ **then**
            Set $\Pi = i$
            break
          end if
        else
          break
        end if
      end for
    else
      if Π' content is less popular than i's **then**
        Set $\Pi = i$
      end if
    end if
  end if
end for

which are less popular until enough room is made to store the recently fetched content.

Fig. 2 summarizes the solution framework and the way the different algorithms interact with each other. As PPO is built on top of actor critic, it inherits its architecture. Thus, there are two networks: actor which is responsible for generating the action, and critic which computes the advantage function. There is also a memory to store the samples of the environment to help the networks reducing the loss function.

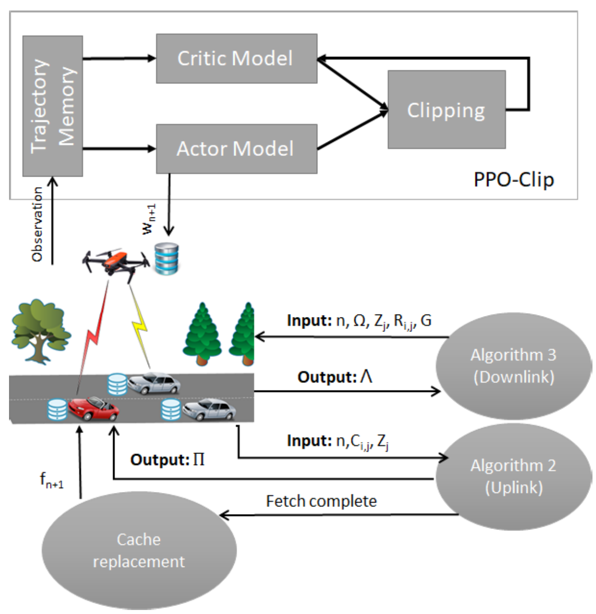

Fig. 2. Solution approach scheme.

VI. Performance Evaluation

A. Simulation Setup

We carry out the simulation studies using Python and PyTorch. To mimic the reality, we take a highway of 2 Km where a UAV is dispatched to cover it with a 2 Gb-cache unit. For the sake of simplicity, we will do our simulation only for one UAV, however, the same can apply for several UAVs covering a large highway. The library of contents is generated having random content sizes while the frequency of content being requested and buffered on vehicles follows Zipf distribution with skewness parameters, α_1 and α_2, equal

Algorithm 3 Downlink Radio Resource Allocation

Inputs: n, Ω, Z_j, $R_{i,j}$, G (denotes vehicle currently being served).
Outputs: Λ as the best vehicle to serve.
if $G \neq None$ **then**
  if $d_G \leq n$ **or** $(Z_j \mid R_{i,G} = 1) \leq D_G^n$ **or** $|L_G - w^n| > \Omega$ **then**
    Set $G = None$
  end if
end if
if $G \neq None$ **then**
  Set $\Lambda = G$
else
  for $i \in I$ where $a_i < n < d_i$ **do**
    if i requested a content that is cached on the UAV, not completely served, has more time slots to remain on the highway segment, and the remaining time slots are enough to receive the whole content **then**
      Set $\Lambda = i$
    end if
  end for
end if
Set $G = \Lambda$

TABLE III
SIMULATION PARAMETERS

Parameter	Value
Highway length	2 Km
δ	1 sec [31]
Number of UAVs	1
Number of Contents	50
Size of Contents	[700-1300] Mb
Bandwidth	10 MHz [41]
σ^2	10^{-14}
τ	3
h	10^{-15}
$P_{U\to V}^{i,n}$, $P_{V\to U}^{i,n}$	0.1 w
UAV discrete velocities	[0, 10, 20, 30, 40] m/s
Density	[2-14] Veh/Km
Learning Rate	0.002
Discount Factor	0.99
Clip	0.02

to 1.3 and 1.6, respectively, and similar to [37]. The set of vehicles is generated at random where the arrivals follow Poisson distribution. Here, as mentioned earlier, we use free flow traffic model to general vehicle arrivals and velocities which has been used widely in the literature to simulate vehicular environment [28], [38], [39]. Meanwhile, the agent is exposed to 1,440,000 samples before results are collected where each sample represents one time slot. Then, different sets are generated for testing. For consistency, the results are averaged over 2000 iterations.

The PPO consists of 2 neural networks, actor and critic, each network has 3 layers, input, hidden, and output. The hidden layer contains 64 nodes and Adam optimizer is designed to train the DNNs. In addition, Hyperbolic Tangent is used mainly as activation function while Softmax is only used for the actor output layer.

The key simulation parameters are listed in Table III.

B. Baseline Methods

We study the performance of our solution approach by comparing its outputs with different methods. Since, to the best our knowledge, there is no work in the literature that addresses similar challenges of UAV mobility, resource allocation, plus cache replacement, we design 4 UAV-steering techniques to use as baselines.

- *Stationary UAV (S-UAV)*: keeps the UAV fixed at the middle of the highway segment.
- *Random UAV Mobility (RAUM)*: selects random velocity to move the UAV at each time slot.
- *Maximum Speed Selection (MASS)*: always chooses the maximum velocity for the UAV. When the UAV reaches one end of the highway segment, it goes back in the opposite direction.
- *Minimum Energy Selection (MISS)*: always chooses the velocity that infers the lowest energy consumption, for the UAV. When the UAV reaches one end of the highway segment, it goes back in the opposite direction.

It is crucial to mention that for fair comparisons, all the UAV trajectory methods suggested above are developed to work based on our proposed algorithms for the up and downlink which are explained in section V.

Fig. 3. PPO-Clip convergence over iterations.

C. Result Analysis

First of all, we examine the convergence of our model. As shown in Fig. 3, the PPO-Clip model converges after around 2000 iterations.

Next, to examine the efficiency of our PPO-based solution, we show in Fig. 4 how the performance of the PPO evolves within time until it reaches convergence state. Here, one can observe that the solution is quite fast at the beginning where PPO reaches above 50% of the maximum performance attained in less than 5 hours and around three-quarters of the maximum performance within a day of training. In addition, we can also notice that exposing the PPO agent to more training samples may enhance the performance further until it converges after less than 6 days of training.

Before we start evaluating the PPO-based solution with other methods, it is very important to see and examine the efficiency of the up and downlink designed algorithms. In order to do that, we suggest two other algorithms to compare with, namely, Greedy and Random. Where Greedy prefers to fetch and serve vehicles that has higher immediate data rate and Random takes random actions. As illustrated

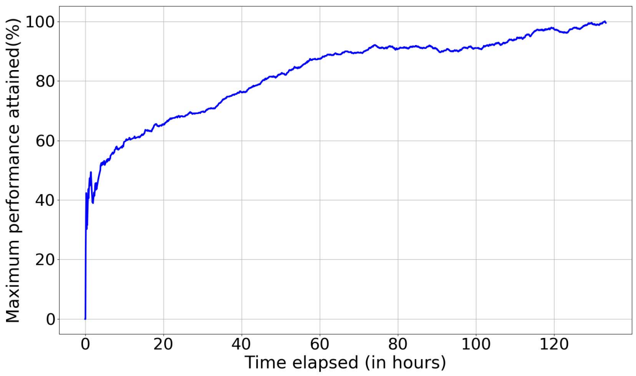

Fig. 4. PPO-Clip performance versus time.

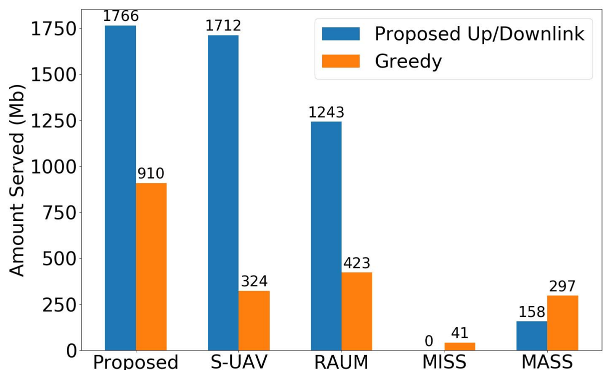

Fig. 5. Proposed wireless allocation algorithms vs Greedy wireless allocation.

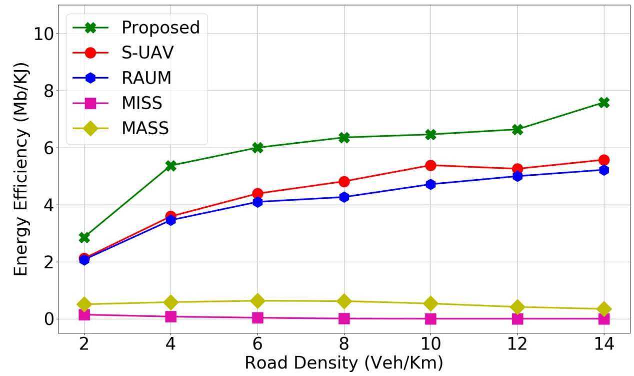

Fig. 6. Impact of road density on energy efficiency.

in Fig. 5, the difference is very significant. With our proposed solution, there is around 50% difference between the proposed algorithms and Greedy. In addition, one can notice that Random does not exist in this figure. Actually, Random produces 0 gain in every attempt. This is because that it cannot complete any transmission due to the randomness in resource allocation. It can also be observed that MISS and MASS have higher amounts of contents served with Greedy method. This can be justified as follows: MASS and MISS are unable to serve or fetch large contents because they do not catch up with vehicles and their connection is not stable as in RAUM and S-UAV. However, with Greedy, it is not mandatory to fetch popular contents which might be large. Thus, MASS and MISS are able to make some gain by fetching and serving smaller and probably unpopular contents.

In terms of evaluation, we compare these four methods with our PPO-based solution approach in terms of the energy efficiency level, amount served to vehicles, and energy consumption.

First, Fig. 6 demonstrates the energy efficiency levels achieved by the five methods. As it can be seen, the values of our proposed solution in addition to S-UAV and RAUM are much higher than MISS and MASS. The reason behind this is that MISS and MASS highly deviate from their regions and, thus, most of their transmission processes are interrupted before they mange to complete. In general, the energy efficiency grows as the number of vehicles increases. Such behavior is normal given that as the density of the highway increases, there will be more chances for the UAV to serve or fetch content using the same amount of energy. We can also observe that the proposed solution scores the highest levels in all cases with a remarkable differences from others. Indeed, it outperforms S-UAV and RAUM by around 20% to 30%. Meanwhile, S-UAV comes in the second place with a slight difference from RAUM. As S-UAV keeps the UAV stable in the middle of the highway, it can attain some stability to the wireless links. However, this method fails in some cases when the content is large and the UAV needs to move toward the vehicle in order to save the link. Moreover, as Eq (8) infers, hovering is not the best choice for power consumption. Indeed, as Eq (8) is non-linear, velocity of 25 m/s consumes the lowest amount of energy. Likewise, RAUM almost obtains identical performance values to S-UAV. Here, we use uniform random distribution, therefore, the UAV goes back and forth over the middle of the highway segment and does not move away from the center.

As shown in the same figure, MISS was the worst technique with very close to 0 energy efficiency regardless of the highway density. The reason behind that is MISS moves the UAV very slow while vehicles move much faster, thus, it will never catch up with vehicles to serve or fetch from them.

Although the amount served to vehicle is not the sole goal of this work, one can note that the PPO-based solution manages to serve more amount of contents than others. In Fig. 7, we can see the total amount served resulted from using the proposed method and S-UAV is 15-35% higher than RAUM and 95% higher than MISS and MASS. The reason behind this good performance of S-UAV is, as mentioned above, owing to its fixed location. We can also notice that the proposed solution starts by 750 Mb at low density while ends up with more than 2000 Mb of content served. That is, the density of the road can notably increase the the service amount and the reason behind this is aforesaid. Moreover, along the y-axis, the proposed solution comes at first place.

Next, Fig. 8 shows how much energy each method can incur. Based on energy consumption function in Eq (8), the amount of energy consumed is correlated with the traveled distance. However, this relationship is non-linear. Hovering consumes higher energy than low velocity movement. Thus, we can see that S-UAV incurs high energy consumption at around

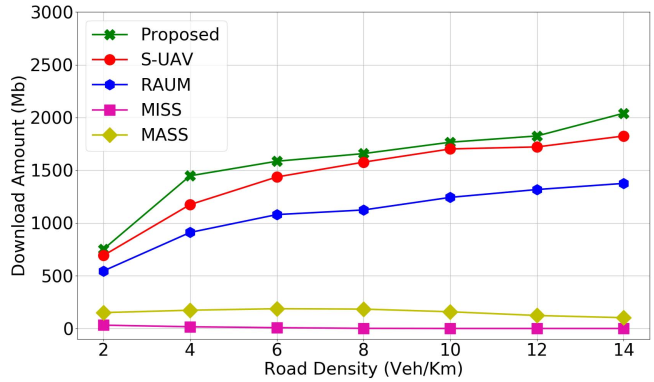

Fig. 7. Impact of road density on amount of content served.

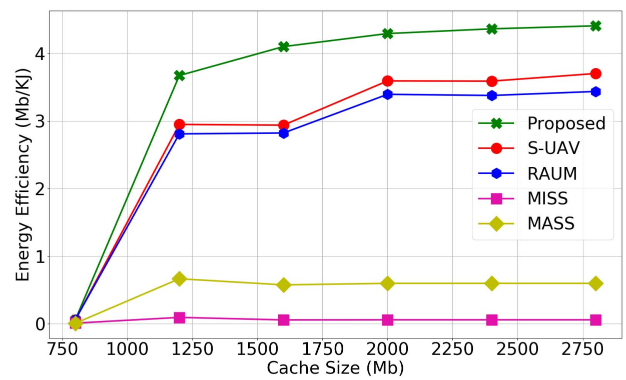

Fig. 9. Impact of cache unit capacity on energy efficiency.

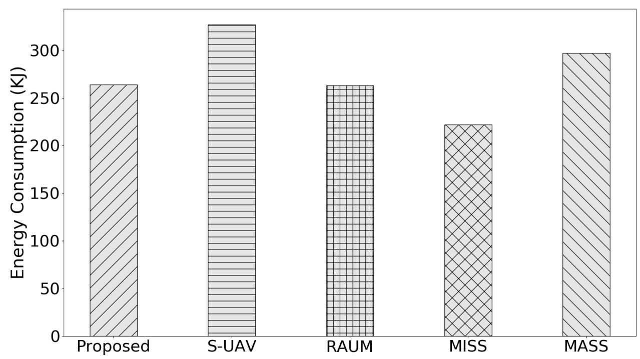

Fig. 8. Amounts of energy consumption incurred using the five methods (per 4 minutes).

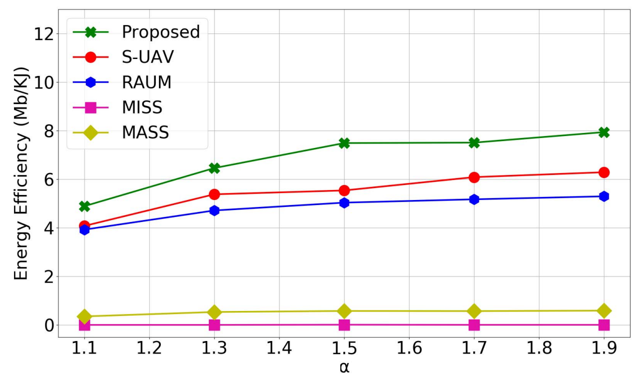

Fig. 10. Impact of Zipf skewness parameter on energy efficiency.

350 KJ. In contrast, MISS wastes less amount of energy at less than 250 KJ since it uses velocity that incurs the lowest amount of energy. Meanwhile, one can also notice that the proposed solution consumes moderate amount of energy at a bit above than 250 KJ. Taking the amount served by the proposed solution shown in Fig. 7 and the amount of energy consumption in this figure, it becomes very clear that the proposed solution is actually the best one.

Next, we investigate into the role that cache capacity can play in improving the efficiency of the system. As it can be seen in Fig. 9, the energy efficiency starts at almost 0 level when the cache unit is very small to fit only one tiny content. The level of energy efficiency jumps surprisingly to 3.7 at cache capacity 1200 Mb. This increase continues as the cache unit grows. However, at some point it saturates. Actually, that depends on the shape of popularity of contents. In this work, as we use Zipf distribution to model the frequency of requests, the top popular contents receive much higher hits than other. Therefore, once these contents can fit in the cache, the extra space will have very little effect. One can also observe that the proposed solution comes in the first place at much higher performance.

Now, let us see the impact of Zipf parameter (α) on the performance. α characterizes the distribution where large values of α means the top popular contents will receive much higher hits than others and vise versa. It can be observed in Fig. 10 that as α increases, the performance of S-UAV, RAUM and besides the proposed solution demonstrate better results. Indeed, as we mentioned earlier, when the α becomes larger, the UAV can serve more vehicles as the overlap over requests raises. We can also note that the proposed solution's performance becomes much higher than others as α increases. That means the PPO-Clip can better adapt and take advantage of the identical requests than other methods which are blind in this regard.

VII. Conclusion

In this work, we presented a novel UAV-assisted content delivery in VANETs without having to be connected to the internet. The system model, which is mathematically formulated, is solved efficiently using Deep Reinforcement Learning technique namely, PPO-Clip with two designed algorithms. The solution method is put to test against other baseline methods to examine i ts adequacy. The results illustrate that even UAV without internet connectivity is able to contribute in serving contents to vehicles through taking the chance of collecting contents from the upcoming vehicles.

In terms of future work, the proposed idea can be further extended through considering V2V communication in order to enhance the performance of the system and improve the QoS to the vehicle users. Such scenario is of interest to network operator as the UAVs may belong to another operator and we cannot guarantee their collaboration. Here, the vehicles may need incentives to motivate them in delivering service for each others [41]. The mobility of UAV can also be improved through considering continuous action space which gives the UAVs more flexibility to planning their trajectory.

REFERENCES

[1] Y. Hui, Z. Su, T. H. Luan, and J. Cai, "Content in motion: An edge computing based relay scheme for content dissemination in urban vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 3115–3128, Aug. 2019.

[2] H. T. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and W.-J. Hwang, "Collaborative multicast beamforming for content delivery by cache-enabled ultra dense networks," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3396–3406, May 2019.

[3] L. Liu, Y. Zhou, J. Yuan, W. Zhuang, and Y. Wang, "Economically optimal MS association for multimedia content delivery in cache-enabled heterogeneous cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 7, pp. 1584–1593, Jul. 2019.

[4] A. Moubayed, A. Shami, P. Heidari, A. Larabi, and R. Brunner, "Edge-enabled V2X service placement for intelligent transportation systems," *IEEE Trans. Mobile Comput.*, early access, Jan. 10, 2020, doi: 10.1109/TMC.2020.2965929.

[5] H. Gao, C. Liu, Y. Li, and X. Yang, "V2 VR: Reliable hybrid-network-oriented V2 V data transmission and routing considering RSUs and connectivity probability," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 13, 2020, doi: 10.1109/TITS.2020.2983835.

[6] H. Huang, A. V. Savkin, and C. Huang, "Reliable path planning for drone delivery using a stochastic time-dependent public transportation network," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 7, 2020, doi: 10.1109/TITS.2020.2983491.

[7] A. Al-Hilo, D. Ebrahimi, S. Sharafeddine, and C. Assi, "Revenue-driven video delivery in vehicular networks with optimal resource scheduling," *Veh. Commun.*, vol. 23, Jun. 2020, Art. no. 100215.

[8] H. Zhou, N. Cheng, J. Wang, J. Chen, Q. Yu, and X. Shen, "Toward dynamic link utilization for efficient vehicular edge content distribution," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8301–8313, Sep. 2019.

[9] N. Zhao *et al.*, "Caching unmanned aerial vehicle-enabled small-cell networks: Employing energy-efficient methods that store and retrieve popular content," *IEEE Veh. Technol. Mag.*, vol. 14, no. 1, pp. 71–79, Mar. 2019.

[10] H. Wu, J. Chen, F. Lyu, L. Wang, and X. Shen, "Joint caching and trajectory design for cache-enabled UAV in vehicular networks," in *Proc. 11th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2019.

[11] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1046–1061, May 2017.

[12] B. Jiang, J. Yang, H. Xu, H. Song, and G. Zheng, "Multimedia data throughput maximization in Internet-of-Things system based on optimization of cache-enabled UAV," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3525–3532, Apr. 2019.

[13] S. Ortiz, C. T. Calafate, J.-C. Cano, P. Manzoni, and C. K. Toh, "A UAV-based content delivery architecture for rural areas and future smart cities," *IEEE Internet Comput.*, vol. 23, no. 1, pp. 29–36, Jan. 2019.

[14] M. Chen, W. Saad, and C. Yin, "Liquid state machine learning for resource and cache management in LTE-U unmanned aerial vehicle (UAV) networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1504–1517, Mar. 2019.

[15] M. Samir, M. Chraiti, C. Assi, and A. Ghrayeb, "Joint optimization of UAV trajectory and radio resource allocation for drive-thru vehicular networks," in *Proc. IEEE WCNC*, Apr. 2019, pp. 1–6.

[16] M. Samir, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Trajectory planning of multiple dronecells in vehicular networks: A reinforcement learning approach," *IEEE Netw. Lett.*, vol. 2, no. 1, pp. 14–18, Mar. 2020.

[17] M. Khabbaz, J. Antoun, S. Sharafeddine, and C. Assi, "Modeling and delay analysis of intermittent V2U communication in secluded areas," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3228–3240, May 2020.

[18] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang, and W. Zhuang, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087–4097, May 2018.

[19] F. Chen *et al.*, "Distribution-aware cache replication for cooperative road side units in VANETs," *Peer-to-Peer Netw. Appl.*, vol. 11, no. 5, pp. 1075–1084, Jul. 2017.

[20] S. Fang, P. Fan, and Z. Khan, "Stochastic playback delay upper bounds of vehicular video content delivery networks with cache-enabled RSUs," in *Proc. IEEE VTC-Spring*, Jun. 2018, pp. 1–5.

[21] R. Ding, T. Wang, L. Song, Z. Han, and J. Wu, "Roadside-unit caching in vehicular ad hoc networks for efficient popular content delivery," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 1207–1212.

[22] Z. Hu, Z. Zheng, T. Wang, L. Song, and X. Li, "Roadside unit caching: Auction-based storage allocation for multiple content providers," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6321–6334, Oct. 2017.

[23] L. Hou, L. Lei, K. Zheng, and X. Wang, "A Q-learning-based proactive caching strategy for non-safety related services in vehicular networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4512–4520, Jun. 2019.

[24] Z. Su, Y. Hui, Q. Xu, T. Yang, J. Liu, and Y. Jia, "An edge caching scheme to distribute content in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5346–5356, Jun. 2018.

[25] H. Wu, W. Xu, J. Chen, L. Wang, and X. Shen, "Matching-based content caching in heterogeneous vehicular networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2018, pp. 1–6.

[26] M. H. Cheung, F. Hou, V. W. S. Wong, and J. Huang, "DORA: Dynamic optimal random access for vehicle-to-roadside communications," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 792–803, May 2012.

[27] R. F. Atallah, C. M. Assi, and J. Y. Yu, "A reinforcement learning technique for optimizing downlink scheduling in an energy-limited vehicular network," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4592–4601, Jun. 2017.

[28] L. Sun, H. Shan, A. Huang, L. Cai, and H. He, "Channel allocation for adaptive video streaming in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 734–747, Jan. 2017.

[29] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329–2345, Apr. 2019.

[30] Y. Cai, Z. Wei, R. Li, D. W. Kwan Ng, and J. Yuan, "Energy-efficient resource allocation for secure UAV communication systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–8.

[31] M. Samir, S. Sharafeddine, C. M. Assi, T. M. Nguyen, and A. Ghrayeb, "UAV trajectory planning for data collection from time-constrained IoT devices," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 34–46, Jan. 2020.

[32] M. Zhu, X.-Y. Liu, and X. Wang, "Deep reinforcement learning for unmanned aerial vehicle-assisted vehicular networks," 2019, *arXiv:1906.05015*. [Online]. Available: http://arxiv.org/abs/1906.05015

[33] A. Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep reinforcement learning framework for autonomous driving," *Electron. Imag.*, vol. 2017, no. 19, pp. 70–76, Jan. 2017.

[34] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*. [Online]. Available: http://arxiv.org/abs/1707.06347

[35] E. Bøhn, E. M. Coates, S. Moe, and T. A. Johansen, "Deep reinforcement learning attitude control of fixed-wing UAVs using proximal policy optimization," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, 2019, pp. 523–533.

[36] M. S. Shokry, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Leveraging UAVs for coverage in cell-free vehicular networks: A deep reinforcement learning approach," *IEEE Trans. Mobile Comput.*, early access, Apr. 29, 2020, doi: 10.1109/TMC.2020.2991326.

[37] A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal and scalable caching for 5G using reinforcement learning of space-time popularities," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 180–190, Feb. 2018.

[38] M. J. Khabbaz, W. F. Fawaz, and C. M. Assi, "A simple free-flow traffic model for vehicular intermittently connected networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1312–1326, Sep. 2012.

[39] Q. Zheng, K. Zheng, H. Zhang, and V. C. M. Leung, "Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 7857–7867, Oct. 2016.

[40] Y. K. Tun, A. Ndikumana, S. R. Pandey, Z. Han, and C. S. Hong, "Joint radio resource allocation and content caching in heterogeneous virtualized wireless networks," *IEEE Access*, vol. 8, pp. 36764–36775, 2020.

[41] O. A. Wahab, H. Otrok, and A. Mourad, "A Dempster–Shafer based tit-for-tat strategy to regulate the cooperation in VANET using QoS-OLSR protocol," *Wireless Pers. Commun.*, vol. 75, no. 3, pp. 1635–1667, Apr. 2014.