Reinforcement Learning-Based Optimization of Relay Selection and Transmission Scheduling for UAV-Aided mmWave Vehicular Networks

B .Tech. Project

Submitted in partial fulfillment of the requirements

of the degree of

Bachelor of Technology

in

Electrical Engineering

By

Aditya Guhagarkar

(Roll No: 21002005)

Under the guidance of

Prof. Vimal Bhatia

Department of Electrical Engineering

Indian Institute of Technology Indore

Dec. 2024

Abstract

Millimeter-wave (mmWave) communications offer abundant bandwidth for vehicular networks, however it is prone to blockages due to buildings, topology, and other environmental factors. To address these challenges, we propose a novel unmanned aerial vehicle (UAV)-aided two-way relaying system to enhance vehicular connectivity and coverage. We formulate a joint optimization problem for relay selection and transmission scheduling to minimize transmission time while ensuring throughput requirements. Proximal policy optimization, deep Q network and constraint programming models are employed to solve the optimization problem. Extensive evaluations reveal that the proximal policy optimization model achieves 100% accuracy with respect to the constraint programming model. In the dynamic scenario, the PPO model outperforms the JRDS model when the number of flows are greater than 40.

Keywords – **Concurrent scheduling, deep Q-network, proximal policy optimization, relay selection, vehicular networks.**

Contents

List of Figures

List of Tables

List of Abbreviations

DQN	Deep Q-Network
FD	Full-Duplex
HD	Half-Duplex
JRDS	Joint Relay Selection with Dynamic Scheduling
LoS	Line-of-Sight
NLoS	Non-Line-of-Sight
mmWave	Millimeter-Wave
PPO	Proximal Policy Optimization
QoS	Quality of Service
RIS	Reconfigurable Intelligent Surface
RL	Reinforcement Learning
SNR	Signal-to-Noise Ratio
SINR	Signal-to-Interference-plus-Noise Ratio
TDMA	Time-Division Multiple Access
UAV	Unmanned Aerial Vehicle
U2V	UAV-to-Vehicle
V2V	Vehicle-to-Vehicle
V2U	Vehicle-to-UAV
DS	Dynamic Scheduling
LEO	Low Earth Orbit

Nomenclature

Symbols

h_u	Altitude of the UAV (meters)
G_0	Maximum antenna gain
$P_r(s_i, r_i)$	Received signal power at receiver r_i from transmitter s_i
$\mathbf{SINR}_{s_i,r_i}$	Signal-to-Interference-plus-Noise Ratio between transmitter s_i and receiver r_i
P_t	Transmit power of the vehicle (dBm)
d_{s_i,r_i}	Distance between transmitter s_i and receiver r_i (meters)
α_v	Path loss exponent for V2V communication
N_0	Noise power spectral density (W/Hz)
W	System bandwidth (Hz)
a_i^j	Binary variable indicating interference presence from transmitter j for link i
b_i^j	Binary variable indicating self interference for link i
R_{s_i,r_i}	Achievable data rate between source s_i and receiver r_i (bps)
η	Transceiver efficiency factor
k_u	UAV transmission constant
P_u	Transmit power of the UAV (dBm)
d_{u,r_k}	Distance between UAV u and vehicle r_k (meters)
α_u	Path loss exponent for U2V communication
Ω_{u,r_k}	Small-scale fading parameter for U2V links
K	Rician factor representing the ratio of LoS to scattered path power
$I_0(\cdot)$	Zero-order modified Bessel function of the first kind
$\mathbf{SINR}_{u,r_k}$	Signal-to-Interference-plus-Noise Ratio for the link between UAV u and vehicle r_k
k_v	V2U communication constant
$d_{s_k,u}$	Distance between vehicle s_k and UAV u (meters)
$\Omega_{s_k,u}$	Small-scale fading parameter for V2U links
$\mathbf{SINR}_{s_k,u}$	Signal-to-Interference-plus-Noise Ratio for the link from vehicle s_k to UAV u
m	Shape parameter for the Gamma distribution in V2V channel modeling

β	Parameter representing interference factor adjustment
f_i	Flow i in the network.
s_i	Source node (transmitter) of the flow f_i.
r_i	Destination node (receiver) of the flow f_i.
d_i	Distance between the source s_i and destination r_i.
Q_i	Achieved throughput for flow f_i.
$T(f_i)$	Transmission time interval for flow f_i.
$H_{i,j}$	Number of communication hops between transmitter s_i and receiver r_j.
$Q_{\mathbf{min}}$	Minimum required throughput for reliable communication.
ξ_a	Number of time slots required for flow a.
Q_a	Flow throughput in the absence of interference (in Gbps).
M	Multiplier based on additional factors like system load.
T	Slot time.
R_a	Data rate of flow a (in Gbps).

Chapter 1

Introduction

1.1 Introduction

The increasing data demands in vehicular networks have driven the need to utilize mmWave frequency bands, which offer higher bandwidths for next-generation communication systems. Despite the advantages, mmWave V2V communication encounters significant challenges due to high path loss, susceptibility to obstacles, and frequent signal blockages, all of which can severely degrade link quality and affect overall network performance. To mitigate these issues, incorporating relay-aided transmission systems using UAVs and terrestrial vehicles has emerged as a promising solution. UAVs, in particular, offer a flexible and dynamic deployment capability that can enhance coverage, boost throughput, and increase communication reliability, especially in fast-changing environments like urban landscapes [1]. This adaptability allows UAVs to act as aerial relays, swiftly repositioning themselves to establish or maintain LoS connections, thereby overcoming the inherent limitations of traditional terrestrial communication infrastructure.

However, optimizing the deployment of UAV resources and the scheduling of communication tasks presents several challenges. These challenges stem from the dynamic nature of the environment, the high relative mobility of vehicles, constraints on UAV power and payload capacity, and varying QoS requirements. Traditional optimization techniques often fall short in handling the inherent unpredictability of vehicular networks and the rapidly changing conditions they entail [2, 3]. In this context, RL methods have shown considerable promise. Techniques like DQN and PPO are well-suited for these complex environments, as they can learn optimal scheduling and resource allocation strategies through iterative interactions and feedback from the environment [4, 5]. DQN models are adept at estimating long-term rewards for various scheduling actions, making them effective in discrete action spaces, while PPO models excel in optimizing policies within continuous action domains, which is crucial for the nuanced and fine-grained control required in managing mmWave V2V communications [6, 7].

1.2 Motivation for BTP and Problem Statement

Reliability continues to be a significant challenge in mmWave vehicular communication systems, particularly when relying solely on terrestrial relays. These relays can fail in V2V scenarios involving distant transceivers due to obstructions, adverse weather, or network congestion. Additionally, while UAVs present a promising alternative as relays, they are constrained by limited battery capacity, which can be particularly problematic under conditions of high traffic demand. To address these issues, the concept of two-way relaying using UAVs has been proposed, allowing simultaneous bidirectional communication to enhance the efficiency and reliability of the network. This approach has the potential to mitigate some of the inherent limitations of single-directional relay methods, making it more suitable for high-density vehicular environments.

In this work, we focus on implementing two-way relaying with UAVs to enhance the robustness of mmWave vehicular communications. To evaluate the effectiveness of different strategies, we developed and tested three models: DQN, PPO, and Constraint Programming. These models were benchmarked against a traditional TDMA algorithm to provide a comprehensive comparison of their ability to optimize network performance and maintain reliability under challenging conditions. The aim is to determine how well these advanced scheduling methods can adapt to dynamic vehicular environments, optimize resource allocation, and ensure reliable communication links, especially when dealing with the high path loss and frequent blockages typical of mmWave bands.

Through these evaluations, our goal is to identify the strengths and weaknesses of each model, offering insights into the most effective techniques for enhancing communication in next-generation vehicular networks. This research has the potential to contribute to the development of robust mmWave V2V communication systems that can support the growing data demands of intelligent transportation systems, smart cities, and autonomous vehicles.

1.3 Organization of the Report

The paper is organized as follows: Chapter 2 contains the literature review. Chapter 3 introduces the system model. Chapter 4 formulates the joint scheduling problem. Chapter 5 details the data processing and proposed schemes. Chapter 6 presents simulation results. Finally, Chapter 7 concludes the paper and discusses future research directions.

Chapter 2

Literature Review

2.1 Introduction

Relay-aided mmWave communications have demonstrated significant potential to enhance system performance in various aspects, such as coverage, spectral efficiency, and energy savings. Wu et al. [8] conducted a study revealing that two-hop device-to-device relaying can substantially expand coverage and improve spectral efficiency in mmWave networks, making it a promising solution for densely populated urban areas where direct LoS paths are often blocked. Similarly, Ruiz et al. [9] explored optimal relay positioning for 5G networks, emphasizing that strategically placed relays can bridge coverage gaps and provide more consistent data rates, particularly in environments with severe shadowing effects. Eltayeb [10] further contributed to the field by proposing relay-aided techniques to mitigate channel estimation errors, which are a common issue in mmWave communication due to the high frequency and narrow beams involved. These techniques enhance the accuracy of channel estimation, leading to more reliable communication links.

In addition, Xiao et al. [11] focused on energy efficiency, presenting a study on effective relay selection strategies in complex scenarios involving multiple sources and relays. Their results demonstrated that appropriate relay selection could lead to considerable energy savings, reducing the operational costs of mmWave networks without sacrificing performance. Furthermore, the use of UAVs as flexible and mobile relays has been a game-changer in the field, particularly for vehicular communication. Jing et al. [12] optimized UAV-assisted communication systems through a combination of joint resource allocation and scheduling, successfully reducing latency and maximizing throughput in dynamic vehicular environments. These UAV-based solutions are particularly advantageous in scenarios where infrastructure-based relays are not feasible or when rapid deployment is necessary.

Efficient scheduling remains a critical factor in maximizing the potential of mmWave communication systems. Hadded et al. [13] demonstrated the effectiveness of TDMA techniques for vehicular applications, showing that TDMA can significantly reduce interference and improve communication reliability in high-mobility scenarios. Additionally, Qiao et al. [14] developed a multi-hop concurrent transmission scheme that leverages the spatial capacity of mmWave relay systems. This approach enables multiple data streams to be transmitted simultaneously over different relays, significantly outperforming traditional single-hop methods in terms of throughput and spectral efficiency. However, while these strategies represent significant progress, challenges remain in adapting them to real-world environments where factors such as mobility, channel variability, and interference are less predictable and harder to manage.

2.2 Conclusions

In conclusion, while relay-aided mmWave communications and advanced scheduling techniques have pushed the boundaries of what is possible in next-generation networks, ongoing research is essential to overcome the practical challenges that come with their real-world deployment. Solutions that address environmental dynamics, user mobility, and real-time resource allocation will be crucial to fully unlocking the potential of mmWave technology in future wireless communication systems.

Chapter 3

System Model

This chapter provides a detailed description of the system model utilized for the UAV-aided mmWave vehicular network. The model includes specifications on the mobility behavior of vehicles and UAVs, the communication patterns adopted, and the channel characteristics that influence the network's performance. The primary goal is to create a realistic environment for evaluating the performance of UAV-assisted vehicular communication.

3.1 Mobility Models

Mobility models play a crucial role in accurately representing the movement and positioning of vehicles and UAVs in a dynamic environment. This section describes the models adopted for simulating both vehicle and UAV mobility within the network.

3.1.1 Vehicle Mobility Model

Vehicles are positioned along a three-laned road, with each lane allowing traffic in a single direction. A minimum separation of two meters is maintained between consecutive vehicles to avoid potential collisions, simulating realistic traffic behavior in urban and highway scenarios. Each vehicle is equipped with a dedicated V2V communication module, enabling them to share data, such as safety warnings and traffic updates. This communication capability allows for dynamic routing and load balancing in the network, improving overall reliability and coverage.

Vehicle mobility follows a predefined speed profile, accounting for acceleration and deceleration due to traffic conditions, obstacles, and the behavior of surrounding vehicles.

3.1.2 UAV Mobility Model

To support V2V communication, multiple UAVs are deployed as relay nodes in the network. Each UAV is programmed to maintain a fixed altitude, denoted by h_u, providing a stable LoS communication channel to ground vehicles. The UAVs operate within a coverage radius of 500 meters, ensuring seamless connectivity between distant vehicles. The deployment of UAVs enhances communication reliability, particularly in areas with dense vehicular traffic or obstacles that obstruct direct V2V communication.

3.2 Antenna Pattern

The communication efficiency between vehicles and UAVs is significantly influenced by the antenna characteristics of the devices. This section discusses the antenna configurations employed in the system, focusing on the modes of operation and associated parameters.

Vehicles are equipped with FD antennas, enabling them to transmit and receive signals simultaneously. This capability allows for uninterrupted data exchange, improving latency and throughput in V2V communication. The FD antennas are designed to handle high data rates and minimize interference using advanced signal processing techniques.

In contrast, UAVs utilize HD communication, where they alternate between transmitting and receiving modes. This operational strategy reduces complexity and power consumption in UAVs, which are constrained by weight and battery capacity. The UAVs employ directional antennas with a maximum antenna gain G_0, optimizing signal strength and focusing the beam towards target vehicles. The directional nature of UAV antennas helps in reducing interference with other nodes and enhancing the overall SNR in the communication links.

The choice of antenna modes ensures a balanced trade-off between communication efficiency and power consumption, aligning with the system's goals of achieving reliable and high-capacity V2V communication facilitated by UAV relays.

3.3 Channel Models

This section discusses the various channel models used in the UAV-aided mmWave vehicular network, focusing on V2V, U2V, and V2U communication links. These models account for the unique propagation characteristics of mmWave communication, including path loss, fading, and interference.

3.3.1 V2V Links

In the V2V communication scenario, the channel is influenced by the urban or highway environment, where both LoS and NLoS conditions may exist. The channel power gain G for V2V links is modeled as a Gamma distribution with a shape parameter m, which reflects the multipath fading conditions commonly experienced in vehicular environments [15].

The received signal power $P_r(s_i, r_i)$ at a receiver r_i from a transmitter s_i is affected by path loss, fading, and interference from other concurrent transmissions. SINR for the V2V link between transmitter s_i and receiver r_i is expressed as:

$$\mathrm{SINR}_{s_i,r_i} = \frac{P_r(s_i, r_i)}{N_0 W + \sum_{j\in\{1,\ldots,N\}\setminus\{i\}} \left(a_i^j P_r(s_j, r_i) + b_i^j \beta P_t\right)},$$

The achievable data rate R_{s_i,r_i} for the V2V link is determined using the Shannon capacity formula [16], which provides an upper bound on the data rate given the SINR:

$$R_{s_i,r_i} = \eta W \log_2\left(1 + \mathrm{SNR}_{s_i,r_i}\right), \tag{3.1}$$

where η is the efficiency factor that accounts for imperfections in the transceiver hardware, such as modulation, coding, and hardware non-linearities.

3.3.2 U2V Links

In the U2V communication scenario, UAVs provide a relay link from an aerial position at a fixed altitude h_u. Due to the elevated position, U2V links often benefit from LoS conditions, but they are still subject to path loss and fading. The path loss in U2V communication includes a combination of large-scale attenuation and small-scale fading, with the latter being modeled using a Rician distribution, which accounts for the dominance of the LoS component over multipath components.

The SINR for the communication link from UAV u to vehicle r_k is given by:

$$\mathrm{SINR}_{u,r_k} = \frac{k_u P_u G_0 d_{u,r_k}^{-\alpha_u} \Omega_{u,r_k}}{N_0 W + \sum_{w\in\{1,\ldots,N\}\setminus\{k\}} \left(a_k^w P_r(s_w, r_k) + b_k^w \beta P_t\right)},$$

The probability density function (PDF) of the small-scale fading Ω_{u,r_k} is expressed as [17]:

$$
\begin{aligned}
f_{\Omega_{u,r_k}}(\omega) = \frac{(K+1)}{\Omega_{u,r_k}} \exp\left(-\frac{K+1}{\Omega_{u,r_k}}\omega\right) \\
\times I_0\left(\sqrt{\frac{2K(K+1)\omega}{\Omega_{u,r_k}}}\right),
\end{aligned}
\tag{3.2}
$$

3.3.3 V2U Links

V2U communication is characterized by minimal interference due to the aerial vantage point of the UAVs. This results in a cleaner channel compared to V2V links, with LoS conditions being more dominant.

The SINR for V2U communication, where the vehicle s_k communicates with UAV u, is described as :

$$
\mathrm{SINR}_{s_k,u} = \frac{k_v P_t G_0 \ d_{s_k,u}^{-\alpha_v} \Omega_{s_k,u}}{N_0 \ W},
$$

Due to the reduced interference, V2U communication often achieves higher SINR, contributing to more reliable and faster data transmission.

3.4 Dynamic Scheduling

DS optimizes communication by adapting to network conditions through two phases: scheduling and transmission (see Figure 3.1). In the scheduling phase, time slots and channels are allocated based on current conditions and traffic demands, minimizing conflicts. During transmission, data is sent over the allocated slots, enabling simultaneous non-conflicting transmissions, thus improving throughput. DS factors in quality of service and channel conditions to adapt to traffic changes, optimizing resource allocation and reducing idle times, which enhances system performance and reliability. Figure 3.1 illustrates this with five flows: non-adjacent flows (e.g., T1, T2) share time slots, while adjacent flows (e.g., T1, T4) require separate slots.

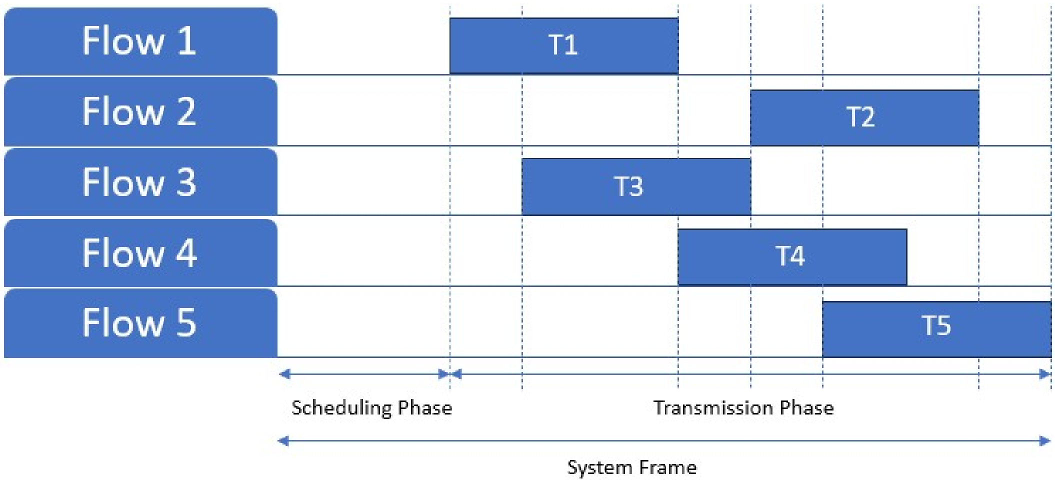

Figure 3.1: Illustration of dynamic scheduling.

Chapter 4

Problem Formulation

This chapter presents the formulation of the joint scheduling problem in a UAV-aided mmWave vehicular network environment. The goal is to efficiently manage data transmission for vehicles on a two-lane highway using UAVs as relay nodes. Effective scheduling is critical to ensure minimal transmission time while meeting quality requirements, such as data throughput and latency constraints, in a highly dynamic and interference-prone communication environment.

4.1 Problem Definition

The problem involves scheduling n data flows, each representing a data transmission request between a source and destination within a three-lane highway setting. In this scenario, UAVs provide relay services to support vehicular communication by acting as intermediary nodes. The primary objective is to minimize the total transmission time for all data flows, which is crucial for real-time vehicular applications and minimizing network congestion. Simultaneously, the solution must adhere to several constraints to maintain the quality and efficiency of the communication, including interference mitigation and throughput guarantees.

Let each flow f_i be characterized by the parameters (s_i, r_i, d_i, Q_i).

The scheduling challenge is to determine the optimal transmission slots for each flow while considering potential interferences, limited communication resources, and constraints related to data delivery quality.

4.2 Constraints

The problem is subject to the following constraints to ensure the reliability and efficiency of communication in the UAV-aided mmWave vehicular network:

4.2.1 Adjacent Flows Constraint

Adjacent flows that share either a transmitter or receiver must not transmit simultaneously. This constraint is necessary to prevent interference caused by overlapping communication resources. Interference can lead to degraded performance, higher latency, and packet loss, which are critical factors in maintaining the quality of service in vehicular networks.

This constraint is formally represented by:

$$T(f_i) \cap T(f_j) = \emptyset, \quad \text{for adjacent flows } f_i, f_j. \tag{4.1}$$

The non-overlapping condition ensures that when two adjacent flows are active, they are assigned to separate time slots to avoid simultaneous transmissions that could cause interference.

4.2.2 Hops Constraint

To reduce latency and maintain efficient data delivery, the number of hops in the communication path for each flow is limited to a maximum of 2. This constraint is particularly important in scenarios involving UAVs, as longer multi-hop paths can lead to increased delays and reduced network performance.

The constraint is mathematically expressed as:

$$H_{i,j} \leq 2. \tag{4.2}$$

A two-hop limit ensures that data is relayed through at most one intermediate UAV, balancing the trade-off between extended coverage and communication latency.

4.2.3 Throughput Threshold

To ensure that each data flow maintains a minimum quality of service, every flow f_i must meet a specified throughput threshold $Q_{\min}$. This constraint guarantees that each flow achieves a baseline data rate, which is essential for applications with stringent communication requirements, such as safety-critical alerts or high-resolution video streaming.

This requirement can be expressed as:

$$Q_i \geq Q_{\min}, \quad \forall\, i. \tag{4.3}$$

The throughput for each flow is influenced by factors such as the SINR, the allocated bandwidth, and the modulation scheme. The scheduling algorithm must take into account the channel conditions and dynamically adjust to maintain the desired throughput levels.

Chapter 5

Data Processing and Proposed Schemes

This chapter details the data processing approach and the proposed scheduling schemes. It discusses the data flow handling process, including the computation of required transmission times and throughput calculations, and outlines the reinforcement learning-based scheduling models.

5.1 Data Processing

Data processing in this system involves acquiring the vehicle positions and computing the required time for each data flow. This is done by taking into account various factors like data rate, throughput, and time slot requirements:

$$\xi_a = \frac{Q_a \cdot M \cdot T}{R_a \cdot T},$$

After calculating the required time and data rate for each flow, the system filters the flows based on throughput and LoS availability. Flows with LoS are retained for direct communication, while others are routed through cars or UAVs to provide coverage. In case both routing options are available, the one with the shortest completion time is chosen for efficiency.

5.2 Proposed Schemes

We propose several schemes to optimize data flow assignment to address the joint scheduling problem in the UAV-aided mmWave vehicular network. These schemes minimize the total transmission time while meeting network constraints, including interference avoidance and throughput requirements. The proposed solutions utilize both traditional optimization methods and advanced reinforcement learning techniques to handle static and dynamic network scenarios.

5.2.1 Constraint Programming Model

In a static scenario where vehicles are stationary and interference is ignored, we employ a Constraint Programming model to optimize data flow scheduling. This model leverages the OR-Tools library to define and solve the problem, aiming to minimize the total transmission time by assigning time slots and resources efficiently.

Algorithm 1 Flow Scheduling using Constraint Programming

Input: Set of flows $F = \{f_1, f_2, \ldots, f_n\}$
Output: Optimized schedule
Step 1: Adjacency Matrix Creation
for each pair of flows (f_i, f_j) **do**
    if f_i and f_j share the same transmitter or receiver **then**
        Mark flows f_i and f_j as adjacent in the matrix
    end if
end for
Step 2: Define Variables
Define $start\_times[i]$ for each flow $f_i \in F$
Define $total\_time$ as the total number of time slots used
Step 3: Apply Constraints
for each pair of adjacent flows (f_i, f_j) **do**
    Enforce non-overlapping schedule:
        $start\_times[i] + duration(f_i) \leq start\_times[j]$ **OR**
        $start\_times[j] + duration(f_j) \leq start\_times[i]$
end for
$total\_time \geq \max_{f_i \in F} (start\_times[i] + duration(f_i))$
Step 4: Objective Function
Minimize $total\_time$
Step 5: Solve the Model
Use OR-Tools CP-SAT solver to find the optimal schedule

5.2.2 Reinforcement Learning Models

Both DQN and PPO are evaluated for their ability to adapt to dynamic environments and their scalability for large problem sizes. These models are chosen for their proven efficiency in decision-making problems in communication networks. In both DQN and PPO, the data rate for each flow is calculated at each timestep, taking into account the interference from other active flows. Scheduling decisions are dynamically made at each timestep based on the current network state and expected performance.

DQN

DQN is a model-free reinforcement learning technique that uses experience replay and target networks to stabilize learning. It approximates Q-values using a neural network, which is trained to predict the expected cumulative reward for different actions. The algorithm is well-suited for environments with discrete action spaces, such as scheduling tasks in a communication network.

At each timestep, the DQN model evaluates the current state of the network, including the positions of vehicles, active data flows, and UAV availability. The data rate for each flow is computed while accounting for interference from simultaneously active flows. This real-time data rate calculation allows the DQN to adaptively decide which flows should be scheduled or rescheduled to optimize overall network performance. The core features of the DQN model include:

- **State Space**: The state space captures the dynamic configuration of the network, including the status of each flow (0 for idle, 1 for active, and 2 for completed), and current data rates.
- **Action Space**: The actions consist of selecting the next data flow to schedule, assigning UAVs, and considering the potential impact on interference.
- **Reward Function**: The reward function is based on minimizing the total transmission time while maintaining throughput requirements. Negative rewards are applied for high interference levels or throughput violations, encouraging the model to make interference-aware scheduling decisions.
- **Data Rate Calculation**: At each timestep, the data rate for each flow is dynamically computed based on the current interference levels caused by other active flows. The model leverages this calculation to make informed scheduling decisions, ensuring that throughput constraints are met even in the presence of network dynamics.
- **Training Strategy**: DQN employs experience replay to store and learn from past experiences, improving stability. A target network is used to periodically update Q-value predictions, enhancing training reliability in complex environments.

DQN's effectiveness in complex scheduling tasks comes from its ability to make data-driven decisions based on past experiences, dynamically adjusting to the network's changing conditions.

PPO

PPO is an on-policy reinforcement learning algorithm that optimizes a policy by maximizing expected cumulative rewards. Unlike traditional methods, PPO performs stable updates by restricting the changes to the policy through a clipped objective function. This method helps to avoid overfitting and ensures stable learning during training, which is especially useful in highly dynamic network environments.

In the PPO model, the data rate for each flow is recalculated at every timestep, considering interference from all active transmissions in the network. This interference-aware approach allows the PPO to make scheduling decisions that are more responsive to real-time changes. Key components of the PPO model include:

- **State Space**: The state space captures the dynamic configuration of the network, including the status of each flow (0 for idle, 1 for active, and 2 for completed), and current data rates.
- **Action Space**: The actions consist of selecting the next data flow to schedule, assigning UAVs, considering the potential impact on interference.
- **Reward Function**: The reward function is designed to maximize throughput while minimizing latency and interference. Penalties are assigned for scheduling decisions that lead to high levels of interference or fall below the throughput threshold.
- **Data Rate Calculation**: At each time step, the model computes the achievable data rate for each flow based on current network conditions and interference from concurrent transmissions. This calculation is essential for making decisions that adapt to dynamic changes and ensure that minimum throughput requirements are consistently met.
- **Training Process**: The PPO model is trained through interactions with a simulated vehicular environment, continuously adjusting its scheduling policy to improve performance. The clipped objective function prevents abrupt changes, leading to smoother and more stable training.

PPO is particularly well-suited for environments with continuous or large action spaces, making it an excellent choice for complex scheduling scenarios with diverse communication constraints.

Fig. 5.1 shows the flow graphs for both models, illustrating their architectures and training processes.

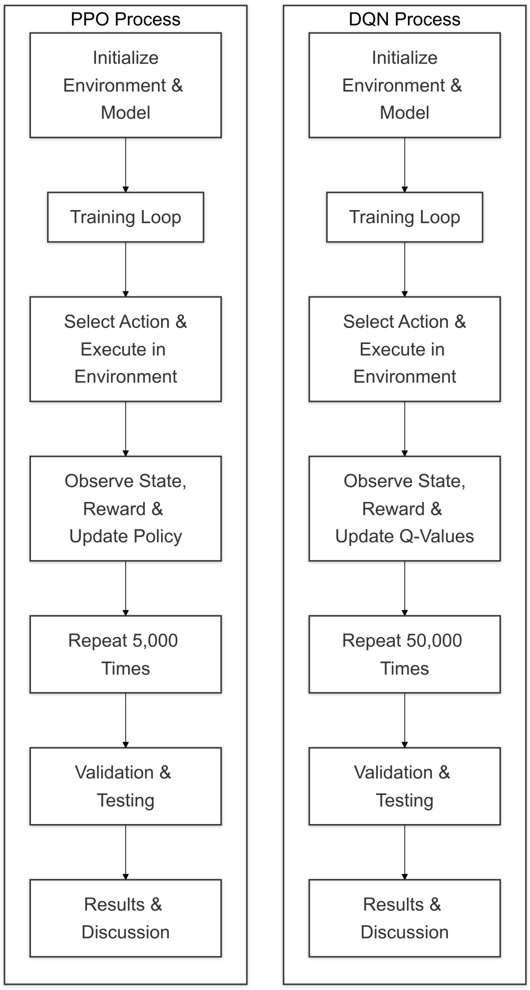

Figure 5.1: Flow graphs for DQN and PPO models.

5.3 Throughput Calculations

Throughput is defined as the ratio of total data in the given time to total time:

$$\text{Throughput} = \frac{\text{Total Data}}{\text{Total Time}}. \tag{5.1}$$

Total data is the summation of $R_{s_i,r_i} \times T_i$ for $i = 1$ to N:

$$\text{Total Data} = \sum_{i=1}^{N} R_{s_i,r_i} \times T_i. \tag{5.2}$$

The total time is $T_{\text{total}} = n \times 0.1$, where n is the total number of time slots, T_i is the time taken for flow i, and 0.1 is the duration of each time slot in seconds. Substituting these into the throughput equation gives:

$$\text{Throughput}\,(\mathcal{U}) = \frac{Q_a \times N \times M}{n}. \tag{5.3}$$

Chapter 6

Simulation Results

This chapter presents the simulation results, showing the effectiveness of the proposed scheduling algorithms in terms of performance metrics such as transmission time, throughput, and fairness.

6.1 Simulation Setup

We simulate a 3 km long, 3-lane highway with 30 vehicles per lane, maintaining safety distances. Vehicles are randomly selected as transceivers, each 4.5 m long and 2 m wide. The highway width is 11.25 m, and three UAVs are positioned along the road. Each vehicle moves with a speed of 100 km/hr.

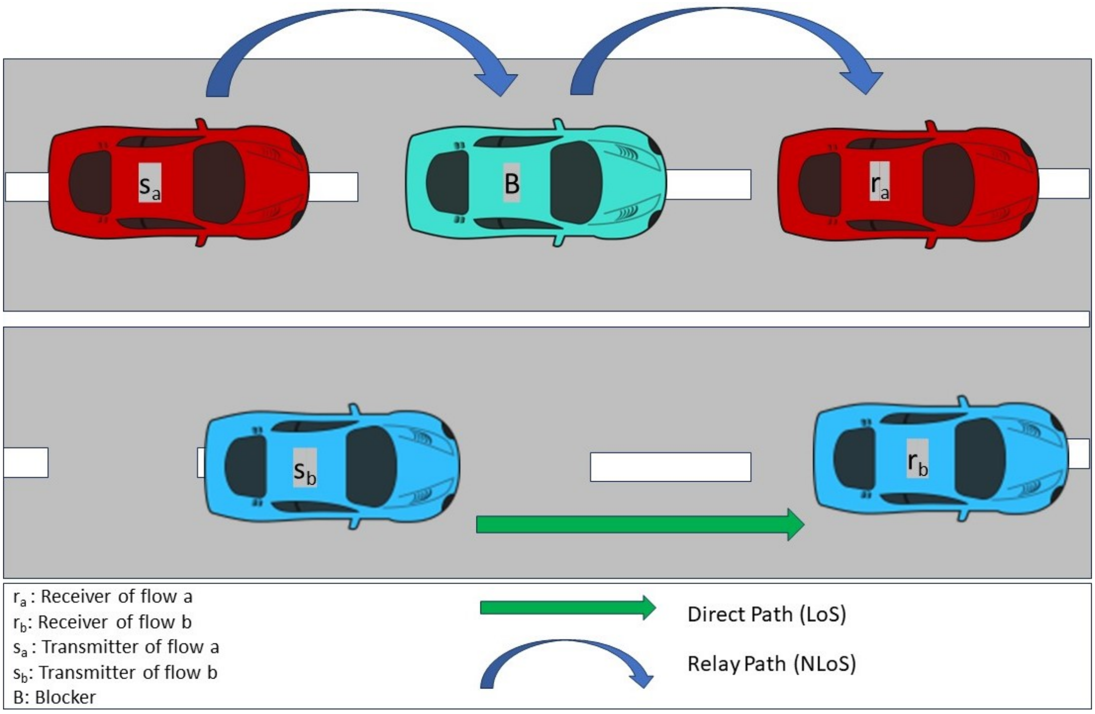

Figure 6.1: Principle of Relay Selection.

Table 6.1: Simulation Parameters

Parameters	Symbol	Value
Carrier Frequency	f	30 GHz
Number of flows	N	60
Number of time slots	M	2000
Slot duration	T	0.1 s
Fading depth	m	2
Background noise	N_0	-134 dBm/MHz
System bandwidth	W	2000 MHz
Transmission power	P_t	40 dBm
Average power of UAV	$\bar{P}_u$	30 dBm
Peak power of UAV	P_u	2 $\bar{P}_u$
Transceiver efficiency	η	0.8
Height of UAV	h_u	100 m
PL factor for V2V	α_v	2.5
PL factor of U2V	α_u	2
Rician K factor	K	9 dB
Throughput threshold	Q_a	0.5 Gbps
maximum antenna gain	G_0	21 dBi

For performance evaluation, two metrics are considered.

- Transmission Time: The total number of time slots consumed to complete flow transmission.
- Network Throughput: The achieved throughput of completed flows in the network [Gbps].

6.2 Numerical Results

The following figures and tables provide an overview of the performance in UAV-aided environments across static and dynamic scenarios, highlighting the efficiency of various scheduling algorithms.

6.2.1 Static Scenario

In the static scenario, environmental conditions remain constant, allowing for a controlled comparison of scheduling algorithms. This entails no vehicular mobility of interference considerations. Performance metrics include transmission time, throughput, and flow completion rates.

Figure 6.2 shows the time performance of different algorithms, with PPO performing notably well, minimizing scheduling and transmission times more effectively than other methods. While the DQN model also reduces transmission time, it suffers from longer scheduling times, which can impact overall efficiency. The JRDS scheme deviates more from the ideal solution as system complexity increases, whereas constraint programming and PPO achieve better consistency across different flow counts.

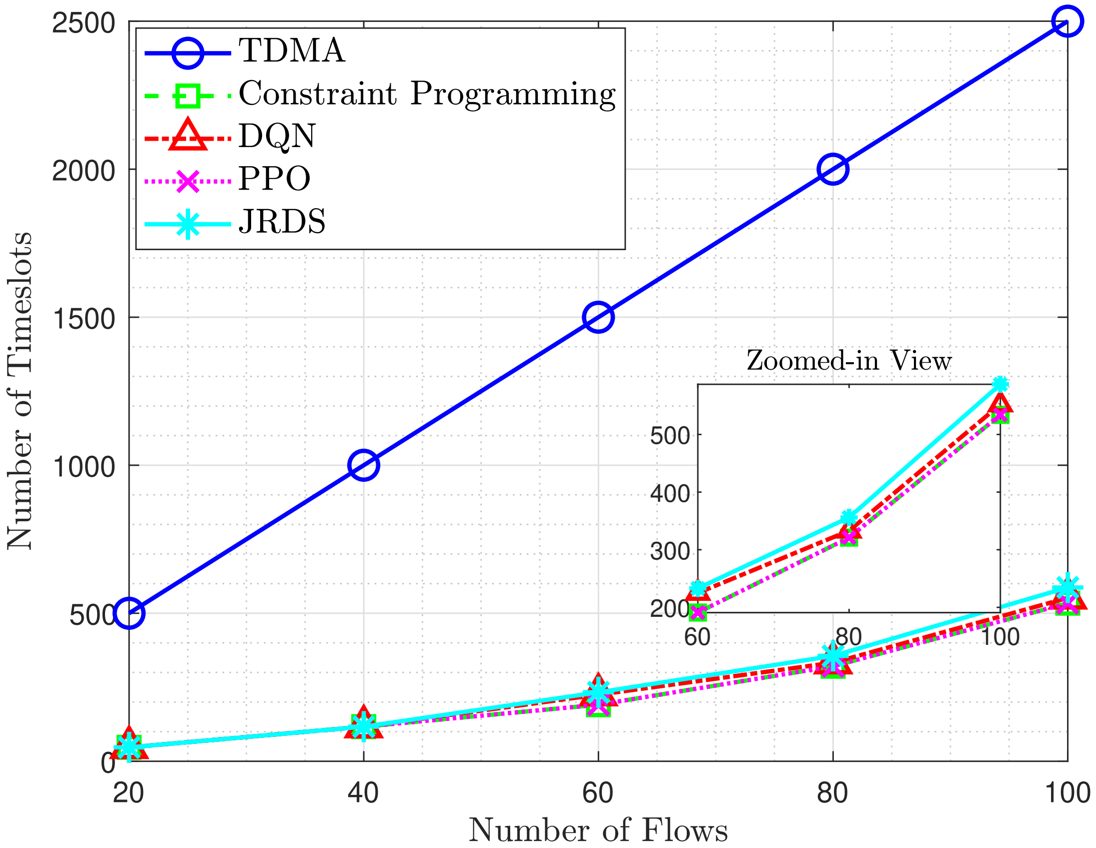

Figure 6.2: Time Performance in UAV-Aided Scenario (Static).

Table 6.2 compares transmission time slots across various flows, demonstrating that PPO achieves a lower time requirement than DQN and other baseline methods as the number of flows increases. This trend highlights PPO's capability to efficiently manage static transmission schedules.

Table 6.2: Transmission Time Slots Comparison in Static Scenario

Flows	TDMA	JRDS	Constraint Programming	DQN	PPO
20	500	47	47	47	47
40	1000	117	117	117	117
60	1500	233	191	225	191
80	2000	356	321	333	321
100	2500	587	534	552	534

PPO, DQN, and constraint programming approaches also show improvements in throughput over TDMA and JRDS, as shown in Fig. 6.3 and Table 6.3. While DQN slightly under-performs at higher flow counts, it still surpasses TDMA. JRDS exhibits a notable decline in throughput beyond 40 flows, reflecting its limited scalability in complex scenarios.

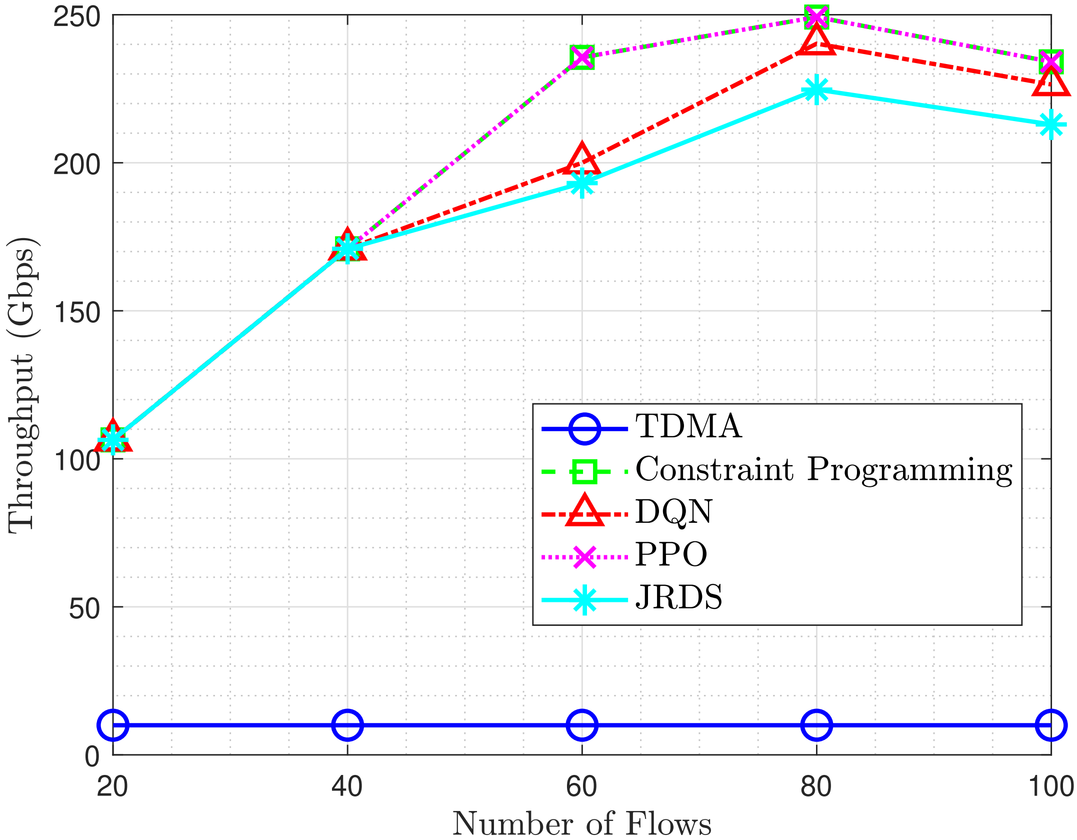

Figure 6.3: Throughput in UAV-aided scenario (Static).

Table 6.3: Network Throughput (Gbps) Comparison in Static Scenario

Flows	TDMA	JRDS	Constraint Programming	DQN	PPO
20	10	106.4	106.4	106.4	106.4
40	10	170.94	170.94	170.94	170.94
60	10	193.13	235.6	200	235.6
80	10	224.72	249.22	240.24	249.24
100	10	212.95	234.1	226.45	234.1

Finally, Fig. 6.4 illustrates flow completion, showing that UAV-aided scenarios enable more completed flows through two-way relaying, enhancing flexibility and resource availability, which PPO and constraint programming efficiently leverage.

6.2.2 Dynamic Scenario

In the dynamic scenario, environmental and flow conditions vary, introducing challenges like interference and vehicle mobility. This scenario simulates real-world conditions, with vehicles moving at an average speed of 100 km/hr.

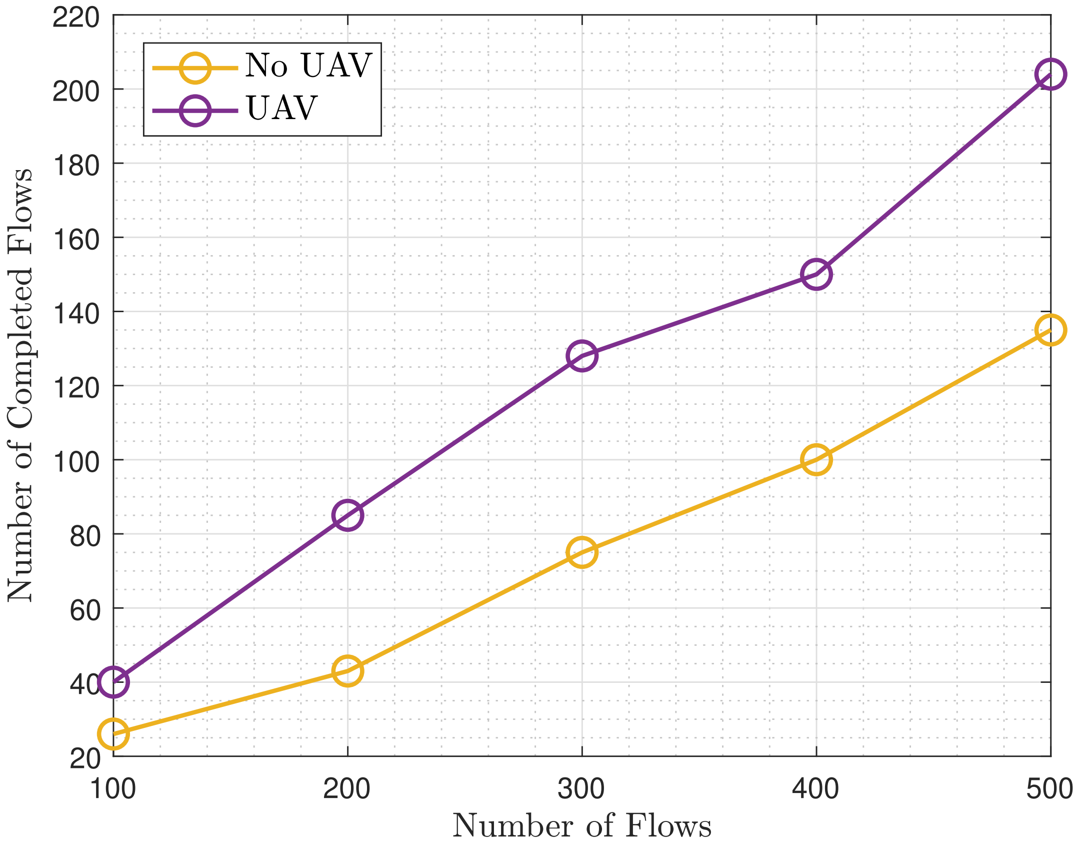

Figure 6.4: Comparison of completed flows: UAV-aided vs UAV-Un-aided.

Fig. 6.5 and Table 6.4 illustrate the time performance in a dynamic environment, where PPO emerges as the best performer, minimizing scheduling and transmission times despite fluctuating conditions. The JRDS scheme performs adequately under lighter loads but becomes less efficient as flows increase.

Table 6.4: Transmission Time Slots Comparison in Dynamic Scenario

Flows	TDMA	JRDS	PPO
20	500	85	85
40	1000	154	154
60	1500	229	200
80	2000	392	354
100	2500	495	458

Throughput results in Fig. 6.6 and Table 6.5 show that PPO outperforms JRDS in managing network throughput under dynamic conditions. PPO maintains consistent performance as flow count increases, demonstrating resilience to changing network states.

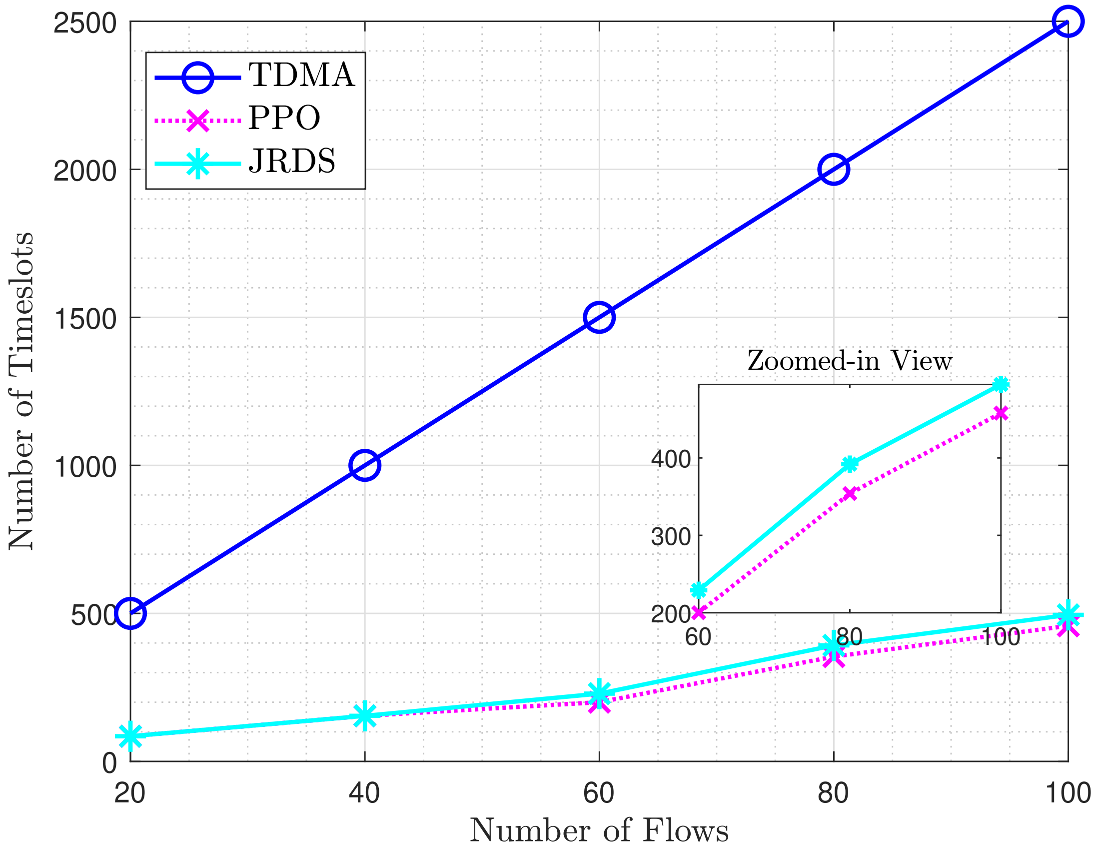

Figure 6.5: Time Performance in UAV-Aided Scenario (Dynamic).

Table 6.5: Network Throughput (Gbps) Comparison in Dynamic Scenario

Flows	TDMA	JRDS	PPO
20	10	58.82	58.82
40	10	129.87	129.87
60	10	196.5	225
80	10	204	227.92
100	10	252.53	272.93

Key Observations in Dynamic Scenario

- **Adaptability:** UAV-based scheduling exhibits strong adaptability to changing conditions, optimizing data flow management.
- **Dynamic Efficiency:** PPO is particularly effective in dynamic environments, minimizing both scheduling and transmission times under fluctuating conditions.
- **Comparative Performance:** PPO consistently outperforms JRDS in both throughput and time efficiency, establishing it as a reliable choice in dynamic conditions. DQN, though not compared here, would likely struggle due to its higher scheduling times observed in the static scenario.

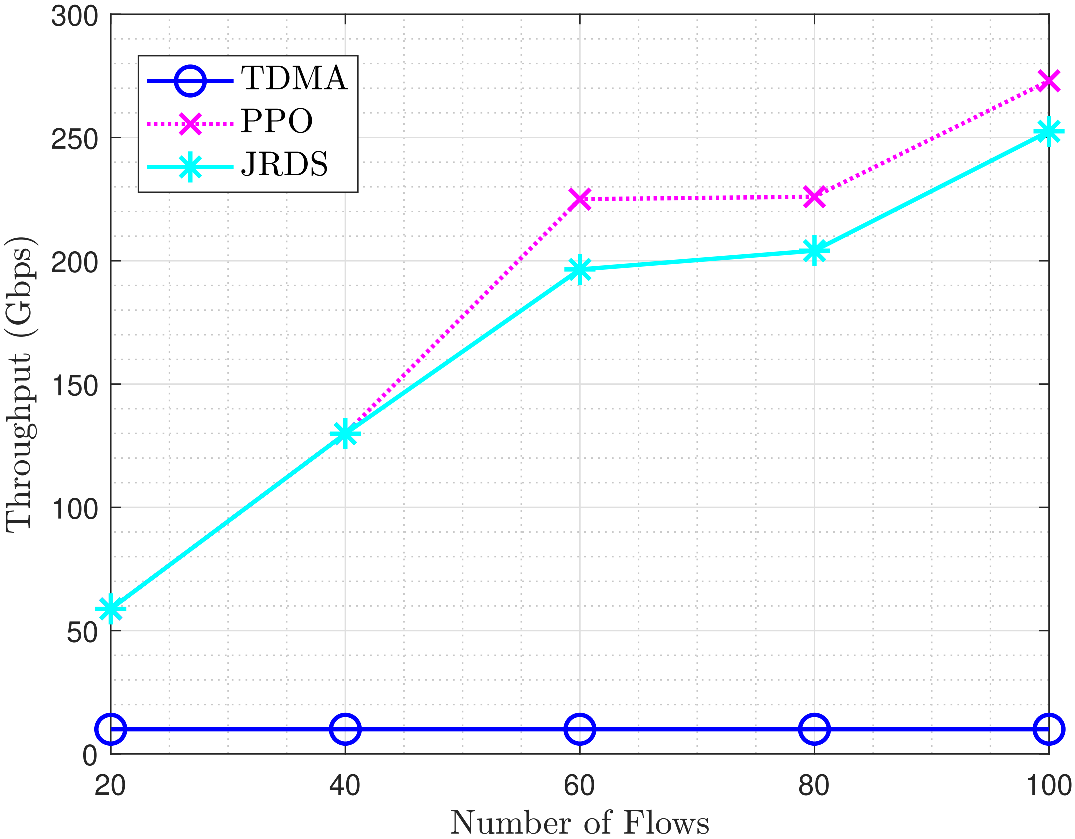

Figure 6.6: Throughput in UAV-aided Scenario (Dynamic).

Chapter 7

Conclusion and Future Work

This chapter concludes the paper and outlines potential future research directions.

7.1 Conclusion

The paper presents a solution for scheduling data flows in a UAV-aided mmWave vehicular network. The proposed reinforcement learning-based scheduling models (DQN and PPO) demonstrate significant improvements in transmission time and throughput efficiency compared to conventional scheduling schemes like TDMA and JRDS. These models are capable of adapting to dynamic network conditions and optimizing the scheduling process, leading to more efficient communication in UAV-assisted vehicular networks. In static environments, we observed that the PPO model gave 100% accuracy concerning the CP model while comparing transmission times. It however had the lowest scheduling time. In dynamic environments where interference and vehicular mobility were in play, we compared the JRDS and PPO models and noticed that the PPO model gives lower transmission times and higher throughput as the number of flows increases.

7.2 Future Research

We are currently working on adding Low Earth Orbit (LEO) satellites, such as Starlink, to the simulation scenario to assist in situations where UAVs are not reachable, enhancing the system's coverage and reliability. The imminent arrival of Starlink in India makes this research particularly relevant and prominent for the future, as it promises to significantly expand connectivity options. Beyond this, future work could explore hybrid reinforcement learning models that combine the strengths of DQN and PPO, optimizing the balance between exploration and exploitation. Additionally, routing algorithms could be further refined to accommodate varying network conditions and meet real-time scheduling demands.

7.3 Publications

Part of the work presented in this report was showcased at the 27th International Symposium on Wireless Personal Multimedia Communications (WPMC 2024) held at Sharda University on 18th December 2024. The publication, titled “Reinforcement Learning-Based Optimization of Relay Selection and Transmission Scheduling for UAV-Aided mmWave Vehicular Networks,” authored by Aditya Guhagarkar, Thushan Sivalingam, Vimal Bhatia, Nandana Rajatheva, and Matti Latva-aho, has been accepted by IEEE.

References

[1] M. Khabbaz, J. Antoun, and C. Assi, "Modeling and performance analysis of uav-assisted vehicular networks," *IEEE Tran. Veh. Technol.*, vol. 68, pp. 8384–8396, Sep. 2019.

[2] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys & Tut.*, vol. 15, no. 3, pp. 996–1019, 2013.

[3] F. Baccelli and B. Błaszczyszyn, "Stochastic geometry and wireless networks: Volume i theory," *Found. Trends Netw.*, vol. 3, p. 249–449, mar 2009.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[6] X. Wang, S. Wang, X. Liang, D. Zhao, J. Huang, X. Xu, B. Dai, and Q. Miao, "Deep reinforcement learning: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 5064–5078, 2024.

[7] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[8] S. Wu, R. Atat, N. Mastronarde, and L. Liu, "Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 2251–2265, 2018.

[9] C. G. Ruiz, A. Pascual-Iserte, and O. Muñoz, "Analysis of blocking in mmwave cellular systems: Characterization of the los and nlos intervals in urban scenarios," *IEEE Transactions on Vehicular Technology*, vol. 69, pp. 16247–16252, 2020.

[10] M. E. Eltayeb, "Relay-aided channel estimation for mmwave systems with imperfect antenna arrays," 2019.

[11] H. Xiao, Y. Hu, K. Yan, and S. Ouyang, "Power allocation and relay selection for multisource multirelay cooperative vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11, pp. 3297–3305, 2016.

[12] J. Li, Y. Niu, H. Wu, B. Ai, R. He, N. Wang, and S. Chen, "Joint optimization of relay selection and transmission scheduling for uav-aided mmwave vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6322–6334, 2023.

[13] M. Hadded, P. Muhlethaler, A. Laouiti, R. Zagrouba, and L. A. Saidane, "TDMA-Based MAC Protocols for Vehicular Ad Hoc Networks: A Survey, Qualitative Analysis, and Open Research Issues," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2461–2492, 2015.

[14] J. Qiao, L. X. Cai, and X. Shen, "Multi-hop concurrent transmission in millimeter wave wpans with directional antenna," in *2010 IEEE International Conference on Communications*, pp. 1–5, 2010.

[15] Y. Wang, H. Wu, Y. Niu, Z. Han, B. Ai, and Z. Zhong, "Coalition game based full-duplex popular content distribution in mmwave vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13836–13848, 2020.

[16] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[17] M. K. Simon and M.-S. Alouini, "Digital communications over fading channels [book review]," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3369–3370, 2008.