

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC DUY TÂN**



Nguyễn Quỳnh Anh

**LƯU TRỮ VÀ PHÂN PHỐI TÀI NGUYÊN TỐI ƯU
CHO DỊCH VỤ TRUYỀN VIDEO TRONG UAV-VANET**

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

Đà Nẵng, 2026

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC DUY TÂN

LUU TRỮ VÀ PHÂN PHỐI TÀI NGUYÊN TỐI ƯU
CHO DỊCH VỤ TRUYỀN VIDEO TRONG UAV-VANET

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

Chuyên ngành: Khoa học máy tính

Mã ngành: 9.48.01.01

NGƯỜI HƯỚNG DẪN KHOA HỌC:

Hướng dẫn 1: TS. VÕ NGUYỄN SƠN

Đà Nẵng, 2026

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài “LƯU TRỮ VÀ PHÂN PHỐI TÀI NGUYÊN TỐI ƯU CHO DỊCH VỤ TRUYỀN VIDEO TRONG UAV-VANET” mà tôi thực hiện là công trình nghiên cứu của riêng tôi, dưới sự hướng dẫn trực tiếp của TS. Võ Nguyên Sơn.

Toàn bộ thông tin, cũng như các kết quả nghiên cứu và các kết luận trong luận án này là trung thực, không sao chép từ bất kỳ nguồn nào và dưới bất kỳ hình thức nào. Việc tham khảo các nguồn tài liệu đều đã được thực hiện trích dẫn và ghi nguồn tài liệu tham khảo đúng quy định. Nếu phát hiện bất cứ gian dối nào trong luận án, tôi xin chịu hoàn toàn trách nhiệm và nhận mọi sự xử lý, kỷ luật từ nhà trường.

Đà Nẵng, ngày tháng năm 2026

Tác giả

Nguyễn Quỳnh Anh

LỜI CẢM ƠN

Luận án tiến sĩ này được hoàn thành nhờ sự nỗ lực của bản thân tôi, cùng với sự hướng dẫn tận tâm của quý thầy cô và sự động viên, ủng hộ quý báu từ gia đình và đồng nghiệp.

Trước hết, tôi xin bày tỏ lòng biết ơn sâu sắc đến TS. Võ Nguyên Sơn, Viện Nghiên cứu Khoa học Cơ bản và Ứng dụng – Đại học Duy Tân, Thành phố Hồ Chí Minh. Thầy đã luôn tận tình hướng dẫn, định hướng khoa học và dành nhiều thời gian trao đổi chuyên môn trong suốt quá trình tôi học tập, nghiên cứu và hoàn thiện luận án. Sự nghiêm túc trong nghiên cứu và tinh thần làm việc khoa học của Thầy là nguồn động lực lớn giúp tôi kiên trì theo đuổi đề tài đến cùng.

Tôi xin trân trọng cảm ơn PGS. TS. Nguyễn Gia Như và quý Thầy, Trường Khoa học Máy tính và Trí tuệ Nhân tạo – Đại học Duy Tân, Thành phố Đà Nẵng, đã luôn quan tâm, định hướng và hỗ trợ kịp thời trong suốt thời gian tôi tham gia chương trình nghiên cứu sinh. Những góp ý chuyên môn sâu sắc của quý Thầy đã giúp tôi hoàn thiện hơn tư duy nghiên cứu và cách tiếp cận vấn đề.

Tôi xin chân thành cảm ơn Đại học Duy Tân đã tạo điều kiện về môi trường học thuật và nghiên cứu thuận lợi, giúp tôi có cơ hội phát triển chuyên môn và hoàn thành luận án này.

Cuối cùng, tôi xin gửi lời cảm ơn sâu sắc đến gia đình – những người luôn bên cạnh, động viên và sẻ chia, là điểm tựa tinh thần vững chắc để tôi vượt qua những giai đoạn khó khăn trong quá trình thực hiện luận án.

Mặc dù tôi đã nỗ lực hoàn thành luận án trong phạm vi khả năng của mình, song chắc chắn không tránh khỏi những thiếu sót. Tôi rất mong nhận được những ý kiến đóng góp quý báu từ quý thầy cô và các nhà nghiên cứu để công trình được hoàn thiện hơn.

Đà Nẵng, ngày tháng năm 2026

Tác giả

Nguyễn Quỳnh Anh

MỤC LỤC

LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN.....	ii
MỤC LỤC	iii
DANH MỤC CÁC TỪ VIẾT TẮT.....	vii
DANH MỤC HÌNH ẢNH.....	x
DANH MỤC BẢNG	xi
MỞ ĐẦU	1
1. Lý do chọn đề tài	1
2. Mục tiêu, đối tượng, phạm vi và phương pháp nghiên cứu	2
2.1. Mục tiêu nghiên cứu.....	2
2.2. Đối tượng nghiên cứu	3
2.3. Phạm vi nghiên cứu.....	3
2.4. Phương pháp nghiên cứu.....	4
3. Nhiệm vụ nghiên cứu và kết quả đạt được.....	4
3.1. Nhiệm vụ nghiên cứu	4
3.2. Kết quả đạt được	5
4. Bố cục luận án	5
CHƯƠNG 1: TỔNG QUAN VỀ LƯU TRỮ TRONG UAV-VANET.....	7
1.1. GIỚI THIỆU VỀ MẠNG UAV-VANET.....	8
1.1.1. Giới thiệu về UAV-VANET	8
1.1.2. Các ứng dụng của UAV-VANET	10
1.1.3. Các hướng nghiên cứu của UAV-VANET	11
1.1.4. Các thách thức của UAV-VANET	12

1.2. CÁC NGHIÊN CỨU LIÊN QUAN ĐẾN LƯU TRỮ TRONG UAV-VANET	13
1.2.1. Truyền video và QoS/QoE trong VANET	13
1.2.2. Lưu trữ và lưu trữ tại biên trong VANET	15
1.2.3. Lưu trữ hợp tác và lưu trữ xác suất	16
1.2.4. Mạng VANET có UAV hỗ trợ	18
1.2.5. Khoảng trống nghiên cứu	19
1.3. GIỚI THIỆU VỀ CÁC KỸ THUẬT TỐI ƯU	20
1.3.1. Giới thiệu về tối ưu	20
1.4. LỘ TRÌNH NGHIÊN CỨU	23
1.5. CÁC KHÁI NIỆM VÀ MÔ HÌNH NỀN TẢNG	32
1.6. KẾT LUẬN CHƯƠNG 1	37
CHƯƠNG 2: KẾT HỢP LƯU TRỮ XÁC ĐỊNH VÀ LƯU TRỮ XÁC SUẤT TRONG VANET	39
2.1. GIỚI THIỆU	40
2.2. CÁC NGHIÊN CỨU LIÊN QUAN.....	44
2.2.1. Các ứng dụng và dịch vụ được hỗ trợ bởi lưu trữ trong VANET	44
2.2.2. Phương pháp DC.....	46
2.2.3. Phương pháp PC	48
2.3. MÔ HÌNH HỆ THỐNG.....	50
2.3.1. Quá trình Poisson đồng nhất cho các VU	52
2.3.2. DC và PC	53
2.3.3. Kênh truyền V2V	53
2.3.4. Kênh truyền R2V	55
2.3.5. Thời gian phục vụ trung bình.....	56

2.4. THIẾT KẾ TỐI ƯU DPC	58
2.4.1. Bài toán tối ưu DPC	58
2.4.2. Giải bài toán bằng GA	59
2.4.3. Triển khai giải pháp DPC	62
2.5. ĐÁNH GIÁ HIỆU SUẤT	64
2.5.1. Thiết lập thông số hệ thống	64
2.5.2. Đánh giá sự hội tụ của GA	66
2.5.3. Đánh giá hiệu suất DPC	68
2.5.4. Độ phức tạp của mô hình DPC	74
2.6. KẾT LUẬN CHƯƠNG 2	76
CHƯƠNG 3: KẾT HỢP LƯU TRỮ VÀ XÁC ĐỊNH VỊ TRÍ PHỤC VỤ TỐI ƯU TRONG TRUYỀN VIDEO HỢP TÁC QUA UAV-VANET	78
3.1. GIỚI THIỆU	78
3.2. MÔ HÌNH HỆ THỐNG	81
3.3. TÍNH TOÁN THÔNG SỐ HỆ THỐNG	83
3.3.1. Mô hình T1D-PPP cho RV	83
3.3.2. Lưu trữ tại RSU và UAV	83
3.3.3. Dung lượng truyền từ RSU đến phương tiện (R2V)	84
3.3.4. Dung lượng truyền từ UAV đến phương tiện (U2V)	85
3.3.5. Thời gian phục vụ trung bình	87
3.4. THIẾT KẾ TỐI ƯU CAHO	89
3.4.1. Bài toán tối ưu CAHO	89
3.4.2. Giải bài toán bằng GA	91
3.4.3. Triển khai giải pháp CAHO	94
3.5. ĐÁNH GIÁ HIỆU SUẤT	95

3.5.1. Thiết lập thông số	95
3.5.2. Đánh giá sự hội tụ của GA.....	98
3.5.3. Đánh giá hiệu suất CAHO	100
3.5.4. Độ phức tạp của mô hình CAHO.....	106
3.6. KẾT LUẬN CHƯƠNG 3	107
KẾT LUẬN	108
1. Kết quả đạt được.....	108
2. Các hướng nghiên cứu tiếp theo	110
CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ CỦA NGHIÊN CỨU SINH.....	111
Công trình đã công bố của luận án	111
Các công trình khác	111
TÀI LIỆU THAM KHẢO	112

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng Việt
1D-PPP	One-dimensional homogeneous Poisson point process	Quá trình điểm Poisson 1 chiều
A&Ss	Advanced Applications and Services	Ứng dụng & dịch vụ tiên tiến
AGMEN	Air–Ground Mobile Edge Network	Mạng di động biên tích hợp Không–Đất
AR	Augmented Reality	Thực tế tăng cường
C-V2X	Cellular Vehicle-to-Everything	Kết nối xe dựa trên mạng di động
CAHO	Caching and Hovering Optimization	Tối ưu lưu trữ và vị trí dừng của UAV
CoT	Cooperative Transmission	Truyền hợp tác
CSI	Channel State Information	Thông tin trạng thái kênh
CV	Caching VU	Xe lưu trữ
D2D	Device-to-Device	Truyền thiết bị-thiết bị
DAV	DPC with average individuals	DPC với cá thể trung bình
DC	Deterministic Caching	Lưu trữ xác định
DPC	Deterministic and Probabilistic Caching	Lưu trữ kết hợp xác định & xác suất
DSRC	Dedicated Short-Range Communications	Truyền thông tầm ngắn chuyên dụng
DT	Digital Twin	Bản sao số
DWO	DPC with worst individuals	DPC với cá thể kém nhất
FANET	Flying Ad-hoc Network	Mạng bay tự tổ chức
GA	Genetic Algorithm	Giải thuật di truyền
IoT	Internet of Things	Internet vạn vật
IoV	Internet of Vehicles	Internet vạn vật trong giao thông
ITS	Intelligent Transportation Systems	Hệ thống Giao thông Thông minh

LoS	Line-of-Sight	Đường truyền nhìn thẳng
MBS	Macro Base Station	Trạm gốc
MEC	Mobile Edge Computing	Tính toán biên di động
MU	Mobile User	Người dùng di động
NUAV	No Unmanned Aerial Vehicle	Không có UAV
NLoS	Non-Line-of-Sight	Đường truyền không nhìn thẳng
OCR	Only caching in RSUs	Chỉ lưu trữ tại RSU
OCV	Only caching in VUs	Chỉ lưu trữ tại VU
OTA	Over-the-Air	Cập nhật qua mạng
PC	Probabilistic Caching	Lưu trữ xác suất
PPP	Poisson Point Process	Quá trình điểm Poisson
QoE	Quality of Experience	Chất lượng trải nghiệm
QoS	Quality of Service	Chất lượng dịch vụ
R2V	RSU-to-Vehicle	Truyền từ RSU đến phương tiện
RACR	Random Caching at RSU	Lưu trữ bất kỳ tại RSU
RACU	Random Caching at UAV	Lưu trữ bất kỳ tại UAV
RAHO	Random Hovering	Dừng bất kỳ tại vị trí nào
RSU	Roadside Unit	Trạm phát ven đường
RV	Requesting VU	Xe yêu cầu
SBS	Small Base Station	Trạm gốc tế bào nhỏ
T1D-PPP	Thinned one-dimensional homogeneous Poisson point process	Poisson 1 chiều làm mỏng độc lập
TrA	Transmission Adaptation	Thích nghi tốc độ truyền
TrS	Transmission Scheduling	Lập lịch truyền
U2V	UAV-to-Vehicle	Truyền từ UAV đến phương tiện
UAV	Unmanned Aerial Vehicle	Thiết bị bay không người lái
UAV-VANET	UAV-VANET	VANET có UAV hỗ trợ
V2I	Vehicle-to-Infrastructure	Xe-tới-hạ tầng

V2V	Vehicle-to-Vehicle	Xe-tới-xe
VANET	Vehicular Ad-hoc Network	Mạng xe tự tổ chức
VASs	Video Applications and Services	Dịch vụ/ứng dụng video
VEC	Vehicular Edge Computing	Tính toán biên trên xe
VR	Virtual Reality	Thực tế ảo
VU	Vehicular User	Người dùng phương tiện

DANH MỤC HÌNH ẢNH

Hình 1.1: Mô hình mạng UAV-VANET	9
Hình 1.2: Ứng dụng và dịch vụ của mạng UAV-VANET	10
Hình 1.3: Mô hình nghiên cứu.....	26
Hình 2.1: Mô hình DPC cho truyền video trong VANET.....	51
Hình 2.2: Ví dụ mô hình DPC với $h_i = 5$	57
Hình 2.3: Tốc độ hội tụ của GA trong phương pháp DPC	67
Hình 2.4: Độ ổn định và chính xác của GA trong phương pháp DPC	67
Hình 2.5: Thời gian phục vụ theo α	69
Hình 2.6: Thời gian phục vụ theo λ	70
Hình 2.7: Thời gian phục vụ theo τ	71
Hình 2.8: Dung lượng lưu trữ theo α	72
Hình 2.9: Dung lượng lưu trữ theo λ	73
Hình 2.10: Dung lượng lưu trữ theo τ	74
Hình 3.1: Mô hình CAHO cho truyền video trong UAV-VANET.....	82
Hình 3.2: Ví dụ về lưu trữ và dừng với $h_i = 5$	84
Hình 3.3: Mô hình vùng phủ của UAV trong đoạn đường n.....	86
Hình 3.4: Tốc độ hội tụ của GA trong phương pháp CAHO	98
Hình 3.5: Độ ổn định và chính xác của GA trong phương pháp CAHO	99
Hình 3.6: Thời gian phục vụ theo α	100
Hình 3.7: Dung lượng lưu trữ theo α	101
Hình 3.8: Thời gian phục vụ theo M	103
Hình 3.9: Dung lượng lưu trữ theo M	104
Hình 3.10: Thời gian phục vụ theo λ	105
Hình 3.11: Dung lượng lưu trữ theo λ	105

DANH MỤC BẢNG

Bảng 1.1: Vùng phủ của một RSU	35
Bảng 2.1: Đặc điểm của kỹ thuật lưu trữ DC và PC	43
Bảng 2.2: Bảng thông số hệ thống mô hình DPC.....	52
Bảng 2.3: Thiết lập thông số cho DPC	64
Bảng 3.1: Thông số hệ thống CAHO	96
Bảng 3.2: Thông số môi trường.....	97
Bảng 3.3: Giá trị tối ưu của các phương pháp so sánh	97

MỞ ĐẦU

1. Lý do chọn đề tài

Sự phát triển của các hệ thống giao thông thông minh (ITS - Intelligent Transportation Systems) trong những năm gần đây đã làm gia tăng đáng kể nhu cầu truyền và xử lý dữ liệu trong mạng VANET (Vehicular Ad-hoc Network). Trong đó, dịch vụ truyền video thời gian thực đóng vai trò quan trọng đối với nhiều ứng dụng như giám sát giao thông, hỗ trợ lái xe an toàn và cung cấp dịch vụ giải trí trên xe. Tuy nhiên, việc truyền video trong môi trường VANET vẫn gặp nhiều thách thức do đặc tính di động cao của các phương tiện và sự biến đổi liên tục của cấu trúc mạng, khiến việc đảm bảo chất lượng dịch vụ (QoS – Quality of Service) và trải nghiệm người dùng (QoE – Quality of Experience) cho truyền video trở thành một bài toán đầy thách thức.

Trong thực tế, các giải pháp truyền video hiện nay trong VANET vẫn dựa chủ yếu vào hạ tầng cố định như các trạm phát ven đường (RSU – Roadside Unit) kết nối về máy chủ trung tâm. Cách tiếp cận này tuy khả thi trong một số kịch bản, nhưng bộc lộ nhiều hạn chế khi mật độ RSU thưa, tải mạng lõi tăng cao hoặc môi trường giao thông phức tạp. Điều đó đặt ra nhu cầu phải xây dựng một cơ chế lưu trữ và phân phối nội dung linh hoạt hơn, thích ứng tốt hơn với đặc tính động của mạng.

Kỹ thuật lưu trữ nội dung (caching) đã được chứng minh là một hướng tiếp cận hiệu quả nhằm giảm thời gian phục vụ và cải thiện QoS/QoE cho dịch vụ video trong VANET. Tuy nhiên, qua quá trình khảo sát các công trình liên quan, NCS nhận thấy rằng nhiều nghiên cứu mới chỉ tập trung vào lưu trữ tại RSU hoặc khai thác lưu trữ tại phương tiện theo cách cục bộ, thiếu một khung tối ưu thống nhất cho toàn hệ thống. Sự tách rời giữa các tầng lưu trữ khiến hiệu quả tổng thể chưa được khai thác tối đa.

Sự xuất hiện của thiết bị bay không người lái (UAV – Unmanned Aerial Vehicle) mở ra một hướng tiếp cận mới cho mạng VANET. Với khả năng di chuyển linh hoạt trong không gian ba chiều, UAV có thể bù đắp những khoảng trống hạ tầng, mở rộng vùng phủ sóng và đóng vai trò như một nút lưu trữ – phân phối nội dung di động. Tuy nhiên, việc tích hợp UAV vào VANET không chỉ đơn thuần là bổ sung thêm một tầng

truyền thông, mà còn đặt ra yêu cầu tối ưu đồng thời vị trí UAV, chiến lược lưu trữ và cơ chế phân phối nội dung. Đây là một bài toán phức tạp, chịu ảnh hưởng đồng thời của đặc tính di động, độ phổ biến nội dung và điều kiện kênh truyền.

Từ góc độ mô hình hóa, bài toán lưu trữ và phân phối tài nguyên trong UAV-VANET có độ phức tạp cao. Các biến tối ưu liên quan đồng thời đến chiến lược lưu trữ nội dung, vị trí hoạt động của UAV và đặc tính truyền thông của hệ thống. Những yếu tố này khiến bài toán trở thành một bài toán tối ưu phi tuyến với không gian tìm kiếm lớn và nhiều ràng buộc.

Các nghiên cứu trước đây cho thấy sự phát triển của các mô hình lưu trữ trong VANET thường diễn ra theo từng bước. Ban đầu, nội dung chủ yếu được lưu trữ tại các RSU. Sau đó, nhiều công trình đã mở rộng mô hình sang lưu trữ kết hợp giữa RSU và phương tiện nhằm tận dụng tài nguyên phân tán trong mạng. Gần đây, sự xuất hiện của UAV đã mở ra khả năng bổ sung thêm một tầng lưu trữ linh hoạt trên không. Có thể thấy rõ một khoảng trống khoa học cần được giải quyết. Đó là xây dựng một khung tối ưu tích hợp, khai thác đồng thời không gian 3D của UAV và cơ chế lưu trữ đa tầng nhằm trực tiếp giảm thời gian phục vụ cho dịch vụ truyền video.

Trên cơ sở đó, NCS lựa chọn đề tài "Lưu trữ và phân phối tài nguyên tối ưu cho dịch vụ truyền video trong UAV-VANET" nhằm xây dựng khung tối ưu tích hợp cho ba tầng RSU–phương tiện–UAV và đề xuất thuật toán giải phù hợp với cấu trúc bài toán hỗn hợp phi tuyến.

2. Mục tiêu, đối tượng, phạm vi và phương pháp nghiên cứu

2.1. Mục tiêu nghiên cứu

Mục tiêu tổng quát của luận án là xây dựng và tối ưu hóa cơ chế lưu trữ và phân phối nội dung trong mạng VANET có UAV hỗ trợ nhằm giảm thiểu thời gian truyền video đến người dùng phương tiện (VU – Vehicular User), đồng thời sử dụng hiệu quả tài nguyên mạng và tài nguyên lưu trữ. Hai mô hình DPC và CAHO cùng hướng tới giảm thời gian phục vụ video. Cụ thể, luận án hướng đến các mục tiêu sau:

- Xây dựng mô hình hệ thống UAV-VANET phục vụ truyền video hướng đến thời gian thực, có xét đến tính di động của phương tiện và UAV.
- Đề xuất mô hình lưu trữ đa tầng (RSU - phương tiện - UAV) dựa trên đặc tính độ phổ biến nội dung và nhu cầu người dùng.
- Xây dựng hàm mục tiêu tối thiểu hóa thời gian phục vụ và phân tích các ràng buộc tài nguyên lưu trữ, truyền dẫn.
- Phát triển giải thuật tối ưu phù hợp nhằm tìm kiếm nghiệm gần tối ưu cho bài toán có tính phi tuyến và không lồi.
- Đánh giá hiệu năng hệ thống thông qua mô phỏng và so sánh với các phương pháp khác.

2.2. Đối tượng nghiên cứu

Đối tượng nghiên cứu của luận án bao gồm:

- Phương tiện tham gia trong VANET;
- RSU;
- UAV;
- Nội dung video được phân phối trong mạng;
- VU với nhu cầu truy cập nội dung trong môi trường giao thông động;
- Các phương pháp/thuật toán tối ưu.

Luận án tập trung vào mối quan hệ tương tác giữa các thành phần trên trong bối cảnh truyền video hướng đến thời gian thực.

2.3. Phạm vi nghiên cứu

Phạm vi nghiên cứu được giới hạn trong các nội dung sau:

- Xem xét dịch vụ truyền video trong mạng VANET có UAV hỗ trợ;
- Tập trung vào bài toán lưu trữ và phân phối nội dung đồng thời sử dụng hiệu quả tài nguyên nhằm giảm thời gian phục vụ;
- Xây dựng mô hình toán học cho hệ thống UAV-VANET trong môi trường ba chiều có tính di động;
- Đề xuất các giải thuật lưu trữ, phân phối và giải thuật tối ưu.

2.4. Phương pháp nghiên cứu

NCS kết hợp các phương pháp nghiên cứu sau:

- *Phương pháp phân tích – tổng hợp tài liệu*: Khảo sát và hệ thống hóa các nghiên cứu liên quan về truyền video, lưu trữ tại biên, lưu trữ hợp tác và VANET có UAV hỗ trợ nhằm xác định khoảng trống nghiên cứu.
- *Phương pháp mô hình hóa toán học*: Xây dựng mô hình hệ thống UAV-VANET, thiết lập hàm mục tiêu tối thiểu hóa thời gian phục vụ và các ràng buộc tài nguyên.
- *Phương pháp tối ưu hóa*: Phân tích đặc tính bài toán (phi tuyến, không lồi, biến hỗn hợp) và đề xuất giải thuật tối ưu phù hợp, trong đó sử dụng giải thuật di truyền (GA – Genetic Algorithm) để tìm nghiệm tối ưu hoặc gần tối ưu.
- *Phương pháp mô phỏng và đánh giá hiệu năng*: Thực hiện mô phỏng bằng công cụ lập trình để đánh giá các chỉ tiêu như thời gian phục vụ, hiệu suất lưu trữ và mức độ cải thiện QoS/QoE; so sánh với các phương pháp tham chiếu.

3. Nhiệm vụ nghiên cứu và kết quả đạt được

3.1. Nhiệm vụ nghiên cứu

- **Nhiệm vụ nghiên cứu 1**: Khảo sát và phân tích các mô hình lưu trữ, phân phối video trong VANET, bao gồm lưu trữ tại RSU, lưu trữ xác suất tại phương tiện và các mô hình có UAV hỗ trợ; từ đó xác định khoảng trống nghiên cứu về tối ưu tích hợp đa tầng.
- **Nhiệm vụ nghiên cứu 2**: Xây dựng mô hình toán học cho bài toán lưu trữ và truyền video trong VANET, tối ưu bán kính lưu trữ tại RSU và xác suất lưu trữ tại phương tiện nhằm tối thiểu hóa thời gian phục vụ dưới ràng buộc tài nguyên; đề xuất và triển khai GA để giải bài toán tối ưu hỗn hợp.
- **Nhiệm vụ nghiên cứu 3**: Mở rộng mô hình sang UAV-VANET, tối ưu đồng thời lưu trữ tại RSU và UAV cùng vị trí dừng của UAV nhằm giảm thời gian phục vụ và sử

dụng hiệu quả tài nguyên lưu trữ; phát triển GA cải tiến để giải bài toán tối ưu đa biến.

- **Nhiệm vụ nghiên cứu 4:** Xây dựng khung tối ưu thống nhất cho bài toán lưu trữ và phân phối video trong UAV-VANET, đánh giá hiệu năng thông qua mô phỏng và so sánh với các phương pháp tham chiếu.

3.2. Kết quả đạt được

Trong quá trình nghiên cứu và hoàn thành luận án, NCS đã có 01 công bố trên tạp chí ISI, cùng 02 công bố trên tạp chí trong nước, với các kết quả cụ thể như sau:

- **Kết quả 1:** Đề xuất mô hình DPC (Deterministic and Probabilistic Caching) nhằm tối ưu kết hợp lưu trữ xác định tại RSU và lưu trữ xác suất tại phương tiện trong VANET. Mô hình DPC tích hợp truyền dẫn R2V (RSU to Vehicle) và V2V (Vehicle to Vehicle), tối thiểu hóa thời gian phục vụ thông qua tối ưu đồng thời bán kính lưu trữ tại RSU và xác suất lưu trữ tại phương tiện, có xét đến phân bố phương tiện, độ phổ biến video và ràng buộc tài nguyên lưu trữ. GA được phát triển để xử lý bài toán tối ưu hỗn hợp (biến nguyên và biến thực), đảm bảo tính hội tụ và ổn định.
- **Kết quả 2:** Đề xuất mô hình CAHO (Caching and Hovering Optimization) cho UAV-VANET, tối ưu đồng thời lưu trữ nội dung tại RSU và UAV cùng vị trí dừng (hovering) của UAV nhằm bù đắp khoảng trống lưu trữ và giảm thời gian phục vụ. Mô hình CAHO tích hợp truyền dẫn R2V và U2V (UAV to Vehicle), tối ưu các biến bán kính lưu trữ tại RSU, quyết định lưu trữ tại UAV, và vị trí dừng của UAV dưới ràng buộc tài nguyên lưu trữ và số lượng UAV hạn chế. Bài toán được giải bằng GA kết hợp hàm phạt và chiến lược xử lý đa biến, hướng đến cải thiện hiệu năng truyền video trong môi trường giao thông động.

4. Bố cục luận án

Luận án được bố cục gồm phần Mở đầu, Chương 1, Chương 2, Chương 3 và phần Kết luận, cụ thể như sau:

Chương 1: Tổng quan về lưu trữ trong UAV-VANET

Chương 1 trình bày các khái niệm nền tảng về UAV-VANET, phân tích các ứng dụng và những thách thức kỹ thuật khi triển khai (giới hạn tài nguyên lưu trữ, biến động topo mạng, yêu cầu QoS/QoE). Trên cơ sở khảo sát các công trình liên quan về lưu trữ nội dung và truyền video trong VANET, chương xác định khoảng trống nghiên cứu về tối ưu tích hợp đa tầng. Đồng thời, các kỹ thuật tối ưu được tổng quan và GA được lựa chọn do phù hợp với đặc tính phi tuyến và ràng buộc phức tạp của bài toán. Chương 1 cũng xác lập các mô hình nền tảng: độ phổ biến video theo Zipf, mô hình di chuyển Poisson, cơ chế lưu trữ tại RSU và định nghĩa thời gian phục vụ, làm cơ sở cho các chương tiếp theo.

Chương 2: Kết hợp lưu trữ xác định và lưu trữ xác suất trong VANET

Chương 2 đề xuất mô hình DPC kết hợp lưu trữ xác định tại RSU (theo bán kính lưu trữ) và lưu trữ xác suất tại phương tiện (theo xác suất) cho truyền video hợp tác trong VANET. Phân bố phương tiện được mô hình hóa bằng quá trình Poisson một chiều, kênh truyền R2V và V2V được mô hình hóa lần lượt theo vùng phủ và phân bố Rician. Bài toán tối ưu DPC có cấu trúc hỗn hợp nguyên–thực, được giải bằng GA với hai chuỗi con mã hóa riêng. Kết quả mô phỏng cho thấy DPC cải thiện đáng kể thời gian phục vụ và sử dụng hiệu quả tài nguyên lưu trữ, đồng thời tạo nền tảng cho việc mở rộng sang mô hình CAHO với sự hỗ trợ của UAV trong chương tiếp theo.

Chương 3: Kết hợp lưu trữ và xác định vị trí phục vụ tối ưu trong truyền video hợp tác qua UAV-VANET

Chương 3 mở rộng bài toán sang môi trường UAV-VANET với mô hình CAHO. Khác với DPC chỉ phân bố lưu trữ giữa hai tầng đã tồn tại, CAHO xem xét chiều không gian trên không bằng cách khai thác UAV như các nút lưu trữ và kênh truyền linh hoạt. Kênh truyền U2V được mô hình hóa theo xác suất LoS/NLoS phụ thuộc góc ngẩng và môi trường đô thị. Bài toán có cấu trúc hỗn hợp nguyên–nhị phân với hai ràng buộc tổ hợp được giải bằng GA với hàm phạt hai thành phần. Kết quả cho thấy CAHO cải thiện đáng kể thời gian phục vụ và sử dụng tài nguyên hiệu quả hơn so với các phương pháp khác, đồng thời mở rộng mô hình lưu trữ từ mặt đất lên không gian ba chiều.

Kết luận: Trình bày về các kết quả đạt được của luận án và hướng nghiên cứu tiếp theo.

CHƯƠNG 1: TỔNG QUAN VỀ LƯU TRỮ TRONG UAV-VANET

Tóm tắt: Chương này trình bày các khái niệm và nền tảng liên quan đến mạng UAV-VANET (Unmanned Aerial Vehicle-assisted Vehicular Ad-hoc Network). Trước hết, chương 1 giới thiệu kiến trúc tổng thể của hệ thống, trong đó UAV được tích hợp với mạng VANET nhằm hỗ trợ các dịch vụ truyền dữ liệu trong môi trường giao thông thông minh. Tiếp theo, các ứng dụng và hướng nghiên cứu của UAV-VANET được phân tích dựa trên các công trình đã công bố. Cuối cùng, chương này trình bày các mô hình nền tảng và các khái niệm cần thiết cho việc xây dựng các mô hình tối ưu trong các chương tiếp theo của luận án.

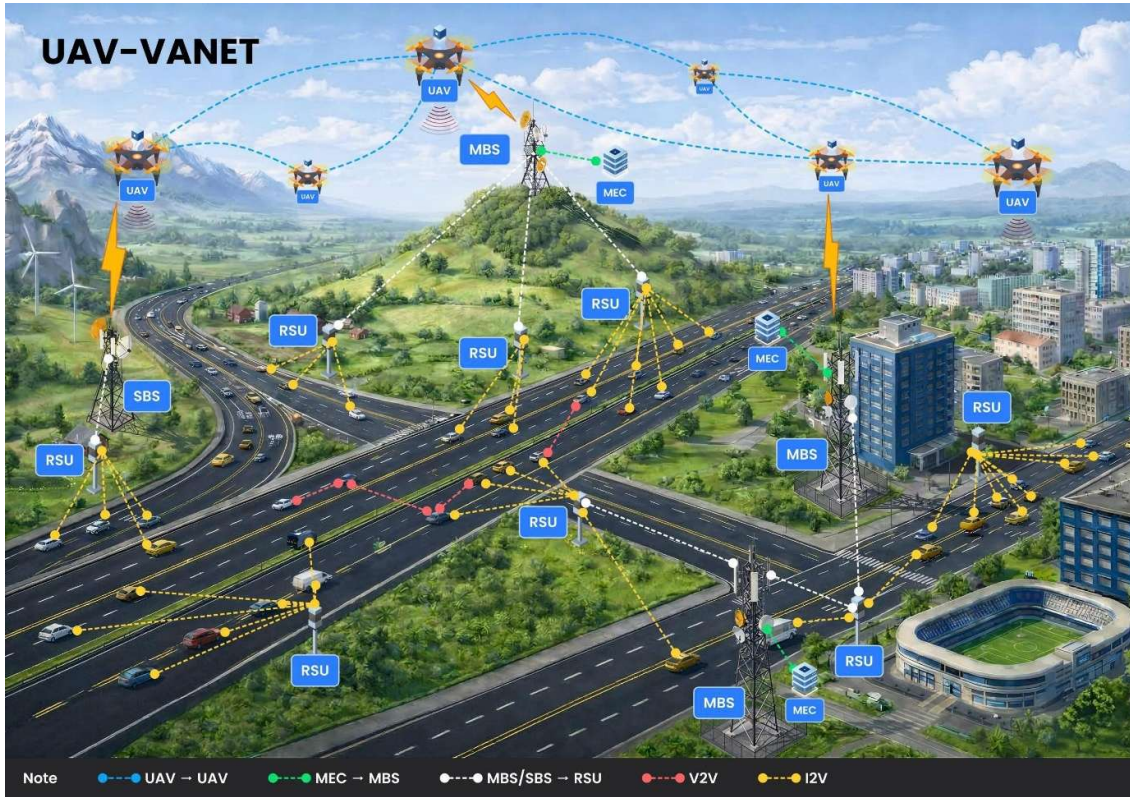
Cụ thể, chương 1 trình bày các ứng dụng tiêu biểu của UAV-VANET, đồng thời phân tích những hướng nghiên cứu chính và các thách thức kỹ thuật đặt ra, như giới hạn tài nguyên lưu trữ, biến động topo mạng, độ trễ truyền dẫn và yêu cầu đảm bảo QoS/QoE (Quality of Service/Quality of Experience). Trên cơ sở khảo sát các công trình liên quan, đặc biệt trong lĩnh vực truyền video, lưu trữ tại biên, lưu trữ hợp tác và mô hình có UAV hỗ trợ, chương này chỉ ra rằng các nghiên cứu vẫn xem xét các thành phần riêng lẻ, trong khi bài toán tối ưu tích hợp đa tầng nhằm giảm trực tiếp thời gian phục vụ còn chưa được khai thác đầy đủ. Trong chương 1, NCS cũng giới thiệu các kỹ thuật tối ưu thường được áp dụng trong UAV-VANET và thảo luận lý do lựa chọn GA (Genetic Algorithm) cho luận án này. Với đặc trưng bài toán có tính phi tuyến, ràng buộc phức tạp và không gian tìm kiếm lớn, GA được xem là một hướng tiếp cận phù hợp để tìm nghiệm gần tối ưu. Cuối cùng, chương 1 xác lập lộ trình nghiên cứu và trình bày các khái niệm, mô hình nền tảng như độ phổ biến video, mô hình di chuyển phương tiện, cơ chế lưu trữ tại trạm phát ven đường và định nghĩa thời gian phục vụ – tiêu chí trung tâm đánh giá hiệu năng hệ thống xuyên suốt các chương tiếp theo. Những nội dung này đặt nền móng cho việc xây dựng và phân tích các mô hình tối ưu được phát triển trong luận án.

1.1. GIỚI THIỆU VỀ MẠNG UAV-VANET

1.1.1. Giới thiệu về UAV-VANET

Quá trình đô thị hóa nhanh chóng trong những năm gần đây đã làm gia tăng đáng kể mật độ phương tiện giao thông tại các thành phố lớn. Điều này đặt ra nhiều thách thức đối với việc quản lý giao thông và bảo đảm an toàn cho người tham gia giao thông. Trong bối cảnh đó, mạng xe tự tổ chức VANET được xem là một thành phần quan trọng của hệ thống giao thông thông minh (ITS - Intelligent Transport System), cho phép các phương tiện trao đổi thông tin trực tiếp với nhau hoặc với hạ tầng ven đường nhằm nâng cao hiệu quả vận hành của hệ thống giao thông. Để giải quyết các vấn đề về an toàn và hiệu quả, mạng lưới xe cộ tự tổ chức VANET đã ra đời như một thành phần cốt lõi của ITS [1]. Về bản chất, VANET cho phép các phương tiện liên lạc trực tiếp với nhau (V2V - Vehicle to Vehicle) hoặc với hạ tầng ven đường (V2I - Vehicle to Infrastructure) thông qua các công nghệ không dây để trao đổi thông tin về tình trạng giao thông và cảnh báo nguy hiểm với độ trễ cực thấp [1], [2]. Mục tiêu hàng đầu của công nghệ này là giảm thiểu tai nạn giao thông, tối ưu hóa luồng xe để tránh ùn tắc và cung cấp các dịch vụ giải trí đa phương tiện cho hành khách [1], [2], [3].

Tuy nhiên, việc triển khai VANET trong thực tế gặp nhiều thách thức do những đặc thù riêng biệt như tốc độ di chuyển của các nút mạng (phương tiện giao thông) rất cao và bị giới hạn bởi lộ trình đường sá, dẫn đến cấu trúc mạng thay đổi liên tục [1], [2]. Việc phụ thuộc vào các trạm phát ven đường (RSU - Roadside Unit) cố định cũng tạo ra hạn chế về vùng phủ sóng tại những khu vực thưa thớt hoặc địa hình phức tạp, nơi kết nối thường xuyên bị gián đoạn [1], [4]. Sự phát triển của các chuẩn kết nối DSRC (Dedicated Short-Range Communications) và C-V2X (Cellular Vehicle-to-Everything) tạo điều kiện để VANET tích hợp với mạng 5G, 6G và IoT, hỗ trợ các ứng dụng như xe tự lái và điều khiển giao thông thông minh [1], [2], [5]. Cùng với đó, các mô hình mạng lai mặt đất–trên không đang mở rộng phạm vi triển khai của VANET lên không gian ba chiều [1], [6].



Hình 1.1: Mô hình mạng UAV-VANET

Để khắc phục các vùng phủ hạn chế của hạ tầng mặt đất trong môi trường đô thị phức tạp, UAV được tích hợp vào VANET tạo thành kiến trúc UAV-VANET (Hình 1.1) [7]. UAV hoạt động như một lớp mạng trên không, bổ sung cho phương tiện, RSU và các trạm gốc di động ở tầng mặt đất. Trong kiến trúc này, UAV đảm nhiệm vai trò trạm cơ sở trên không hoặc nút trung chuyển dữ liệu, kết nối các phương tiện với hạ tầng mặt đất và mạng lõi qua các liên kết không-đất [2], [7].

UAV-VANET có ưu điểm thiết lập được liên kết nhìn thẳng (LoS — Line-of-Sight) giữa UAV và phương tiện hoặc trạm mặt đất, giảm ảnh hưởng của che khuất và suy hao tín hiệu trong môi trường đô thị nhiều vật cản [2], [7]. Tầng phủ sóng trên không mở rộng vùng phục vụ vượt giới hạn của các RSU cố định và duy trì kết nối tại các khu vực hạ tầng thưa thớt. Trong các tình huống khẩn cấp khi hạ tầng mặt đất bị hư hại, UAV còn đóng vai trò hạ tầng truyền thông tạm thời phục vụ điều phối và ứng phó [2].

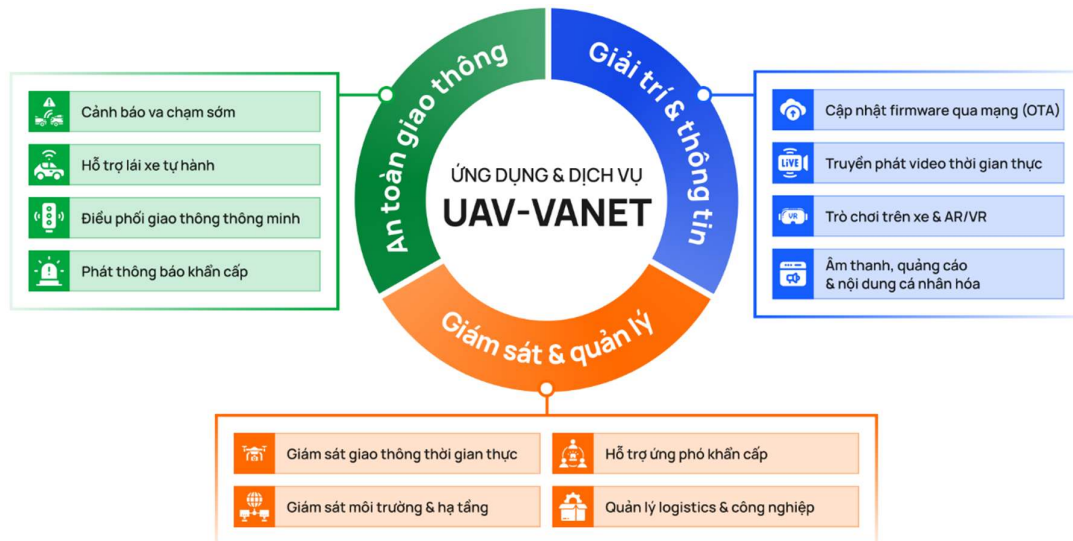
Một bước phát triển tiếp theo của UAV-VANET là mô hình AGMEN (Air-Ground Mobile Edge Network) [8], trong đó UAV tích hợp năng lực MEC và lưu trữ dữ liệu cùng với chức năng truyền thông. Kiến trúc này đưa nội dung và tác vụ tính toán

nặng gần người dùng hơn, giảm tải mạng lõi và rút ngắn độ trễ dịch vụ [9]. AGMEN phù hợp với hướng phát triển của IoV trong kỷ nguyên 6G, nơi UAV, phương tiện và hạ tầng mặt đất phối hợp trong một hệ thống mạng ba chiều.

Vận hành UAV-VANET trong khuôn khổ mạng bay tự tổ chức (FANET – Flying Ad-hoc Network) đặt ra ba thách thức kỹ thuật chính: tối ưu quỹ đạo bay, quản lý năng lượng pin UAV và điều phối nhiều UAV đồng thời trong không gian ba chiều. Các phương pháp tối ưu hóa và học máy là hướng tiếp cận được nhiều nghiên cứu khai thác để giải quyết các bài toán này [2].

1.1.2. Các ứng dụng của UAV-VANET

Các ứng dụng của UAV-VANET trải rộng trên nhiều lĩnh vực, từ an toàn giao thông đến dịch vụ số và quản lý đô thị thông minh (Hình 1.2). Trong nhóm ứng dụng an toàn giao thông, UAV-VANET hỗ trợ cảnh báo va chạm sớm, chia sẻ tình trạng giao thông và phát thông báo khẩn cấp đến phương tiện trong khu vực ảnh hưởng [6], [10]. Đối với xe tự hành và hệ thống hỗ trợ lái nâng cao, UAV cung cấp góc nhìn toàn cục từ trên cao, bổ sung dữ liệu cho các cảm biến gắn trên xe trong các tình huống cảm biến mặt đất bị che khuất [11].



Hình 1.2: Ứng dụng và dịch vụ của mạng UAV-VANET

Trong lĩnh vực giải trí và dịch vụ thông tin trên xe, UAV-VANET hỗ trợ truyền phát video thời gian thực qua liên kết LoS có chất lượng cao hơn so với liên kết R2V

truyền thông [12], [13]. Các dịch vụ điển hình bao gồm trò chơi trực tuyến trên xe, thực tế tăng cường/ảo (AR/VR - Augmented Reality/Virtual Reality), và cập nhật phần mềm qua mạng (OTA - Over the Air) [13]. UAV-VANET cũng phục vụ các dịch vụ quảng cáo và nội dung cá nhân hóa dựa trên vị trí phương tiện [11], [14], [15].

Ngoài an toàn và giải trí, UAV-VANET còn cho thấy giá trị nổi bật trong giám sát và quản lý. Các UAV có thể thực hiện giám sát giao thông theo thời gian thực, phát hiện sớm các điểm ùn tắc, tai nạn hoặc sự cố hạ tầng, từ đó hỗ trợ cơ quan quản lý đưa ra các biện pháp điều phối kịp thời [16]. Trong các tình huống khẩn cấp như thiên tai, cháy nổ hay tai nạn quy mô lớn, UAV-VANET đóng vai trò là hạ tầng truyền thông tạm thời, hỗ trợ công tác cứu hộ, cứu nạn và điều phối lực lượng [17]. Bên cạnh đó, mô hình này còn được ứng dụng trong giám sát môi trường và hạ tầng đô thị, cũng như trong quản lý logistics [18], [19] và các hoạt động công nghiệp, nơi yêu cầu cao về khả năng kết nối linh hoạt và thu thập dữ liệu trên diện rộng.

Từ các ví dụ trên có thể thấy UAV-VANET không chỉ hỗ trợ cải thiện khả năng truyền thông trong mạng giao thông mà còn mở rộng phạm vi ứng dụng sang nhiều lĩnh vực khác nhau như giám sát giao thông, dịch vụ thông tin trên xe và quản lý đô thị thông minh. Sự kết hợp giữa UAV và mạng VANET tạo ra một hạ tầng linh hoạt hơn, có khả năng thích ứng tốt với các kịch bản giao thông động. Chính sự đa dạng về các kịch bản ứng dụng này đã và đang thúc đẩy UAV-VANET trở thành một trong những hướng nghiên cứu và triển khai đầy tiềm năng trong hệ thống giao thông và mạng không dây thế hệ tiếp theo.

1.1.3. Các hướng nghiên cứu của UAV-VANET

Trong những năm gần đây, UAV-VANET đã thu hút nhiều hướng nghiên cứu nhằm nâng cao hiệu năng truyền thông trong môi trường giao thông động. Các hướng nghiên cứu này có thể được khái quát thành ba nhóm chính: mô hình kiến trúc mạng, tối ưu tài nguyên truyền thông và lưu trữ/phân phối nội dung.

Nhóm thứ nhất tập trung vào kiến trúc và mô hình hệ thống UAV-VANET. Trong các nghiên cứu này, UAV được tích hợp vào VANET như một tầng mạng trên không nhằm mở rộng vùng phủ, cải thiện khả năng kết nối và hỗ trợ truyền thông trong những

khu vực hạ tầng mặt đất bị hạn chế. Các mô hình thường xem xét đồng thời đặc tính di chuyển của phương tiện, vùng phủ động của UAV và liên kết truyền thông không–đất.

Nhóm thứ hai nghiên cứu tối ưu tài nguyên trong UAV-VANET, bao gồm vị trí hoạt động của UAV, quỹ đạo bay, công suất truyền, băng thông và tài nguyên tính toán. Mục tiêu của nhóm nghiên cứu này là cải thiện thông lượng, giảm độ trễ, tiết kiệm năng lượng và nâng cao độ tin cậy của hệ thống. Tuy nhiên, các nghiên cứu này thường tập trung nhiều vào khía cạnh truyền thông hoặc tài nguyên mạng, trong khi chưa xem lưu trữ nội dung video là một thành phần tối ưu trung tâm.

Nhóm thứ ba liên quan trực tiếp hơn đến luận án là lưu trữ và phân phối nội dung trong UAV-VANET. Hướng nghiên cứu này khai thác khả năng lưu trữ tại RSU, phương tiện hoặc UAV để đưa nội dung đến gần người dùng hơn, qua đó giảm tải mạng lõi và rút ngắn thời gian phục vụ. Tuy nhiên, các công trình hiện có thường chỉ xem xét từng tầng lưu trữ riêng lẻ hoặc chưa tối ưu đồng thời giữa lưu trữ nội dung, vị trí UAV và thời gian phục vụ video. Đây chính là cơ sở để luận án tiếp tục khảo sát sâu hơn các nghiên cứu liên quan đến lưu trữ trong UAV-VANET ở mục tiếp theo.

1.1.4. Các thách thức của UAV-VANET

Mặc dù UAV-VANET mang lại nhiều lợi ích cho hệ thống giao thông thông minh, việc triển khai mô hình này trong thực tế vẫn gặp nhiều thách thức. Trước hết, UAV-VANET là một mạng dị thể, trong đó phương tiện, RSU, UAV và hạ tầng mạng di động cùng tồn tại và tương tác ở nhiều tầng khác nhau. Sự khác biệt về công nghệ truy nhập, giao thức truyền thông và đặc tính vận hành khiến việc quản lý kết nối, đồng bộ giao thức và duy trì tính liên tục dịch vụ trở nên phức tạp, đặc biệt khi phương tiện di chuyển giữa các vùng phủ hoặc khi UAV tham gia/rời khỏi mạng [11], [6].

UAV-VANET có đặc tính di động cao và hoạt động trong không gian ba chiều. Trong khi phương tiện mặt đất di chuyển nhanh nhưng bị ràng buộc bởi cấu trúc đường sá, UAV có quỹ đạo bay linh hoạt hơn. Sự không đồng nhất này làm topo mạng thay đổi liên tục, gây khó khăn cho việc duy trì liên kết ổn định. Do đó, tối ưu vị trí hoặc quỹ đạo UAV cần cân bằng đồng thời vùng phủ, chất lượng liên kết, độ trễ và mức tiêu thụ năng lượng [14], [16].

Kênh truyền trong UAV-VANET chịu ảnh hưởng của nhiều yếu tố như độ cao UAV, môi trường đô thị, hiện tượng che khuất, nhiễu và sự chia sẻ tài nguyên vô tuyến giữa nhiều liên kết. Vì vậy, việc mô hình hóa kênh truyền không-đất và phân bổ tài nguyên truyền thông hiệu quả là một vấn đề khó, nhất là đối với các ứng dụng thời gian thực có yêu cầu QoS/QoE cao [19], [20].

Bên cạnh truyền thông, lưu trữ và phân phối nội dung cũng là một thách thức quan trọng. Các dịch vụ video và đa phương tiện tạo ra nhu cầu dữ liệu lớn, trong khi tài nguyên lưu trữ tại RSU, phương tiện và UAV đều bị giới hạn. Đồng thời, nhu cầu nội dung của người dùng thay đổi theo không gian và thời gian, đòi hỏi các chiến lược lưu trữ và phân phối nội dung có khả năng thích ứng với trạng thái mạng và độ phổ biến nội dung [21], [12], [22], [23].

Ngoài ra, UAV-VANET còn phải bảo đảm độ trễ thấp, kết nối ổn định và trải nghiệm người dùng nhất quán trong môi trường mạng biến động. Đây là yêu cầu đặc biệt quan trọng đối với truyền video thời gian thực, nơi chất lượng dịch vụ có thể bị ảnh hưởng trực tiếp bởi sự thay đổi topo, điều kiện kênh truyền và vị trí UAV [15], [24]. Bên cạnh đó, dung lượng pin hạn chế của UAV cũng ảnh hưởng đến thời gian hoạt động, vùng phủ và độ tin cậy của hệ thống [11], [6].

Từ các thách thức trên, có thể thấy bài toán lưu trữ và phân phối video trong UAV-VANET cần được xem xét trong một khung tối ưu tích hợp, có xét đến tài nguyên lưu trữ, điều kiện truyền thông, vị trí UAV và yêu cầu thời gian phục vụ. Đây là cơ sở để luận án tập trung vào các mô hình tối ưu nhằm giảm thời gian phục vụ video trong VANET và UAV-VANET.

1.2. CÁC NGHIÊN CỨU LIÊN QUAN ĐẾN LƯU TRỮ TRONG UAV-VANET

1.2.1. Truyền video và QoS/QoE trong VANET

Sự phát triển nhanh chóng của các dịch vụ đa phương tiện trong hệ thống giao thông thông minh đã thúc đẩy nhiều nghiên cứu tập trung vào bài toán truyền video và đảm bảo QoS/QoE trong mạng VANET. Theo các báo cáo được công bố rộng rãi, video hiện chiếm tỷ trọng lớn trong tổng lưu lượng dữ liệu di động, đồng thời là loại dịch vụ

nhạy cảm nhất đối với độ trễ, gián đoạn và biến thiên băng thông, khiến QoS/QoE trở thành tiêu chí đánh giá quan trọng đối với mạng VANET hiện đại [25], [26].

Các nghiên cứu ban đầu về truyền video trong VANET được xây dựng dựa trên kiến trúc tập trung hoặc bán tập trung, trong đó, nội dung video được cung cấp từ máy chủ trung tâm, trạm gốc hoặc các RSU. Các nghiên cứu này cho thấy việc sử dụng RSU cho phép kiểm soát tốt hơn tài nguyên truyền thông và hỗ trợ truyền video theo thời gian thực trong các kịch bản hạ tầng dày đặc [27], [28], [29].

Tuy nhiên, các công trình này cũng đồng thời chỉ ra rằng hiệu quả của kiến trúc tập trung vào RSU phụ thuộc mạnh vào mật độ và sự phân bố của RSU. Trong thực tế, do chi phí triển khai cao, RSU thường được bố trí thưa và không đồng đều trên những tuyến đường khác nhau, đặc biệt trên các tuyến đường dài hoặc khu vực ngoại ô, dẫn đến việc phương tiện không thể duy trì kết nối ổn định với RSU gần nhất [28], [30], [31].

Một thách thức khác được nhấn mạnh trong nhiều nghiên cứu là tính không ổn định của kết nối trong VANET. Do phương tiện di chuyển với tốc độ cao và cấu trúc mạng thay đổi liên tục, các liên kết V2I thường có thời gian tồn tại ngắn, dễ bị gián đoạn, gây ảnh hưởng nghiêm trọng đến tính liên tục của các phiên truyền video [32], [33]. Để đối phó với vấn đề này, một số công trình đã đề xuất các cơ chế truyền video thích nghi, bao gồm điều chỉnh tốc độ truyền, giới hạn số lần truyền lại và thay đổi chất lượng video dựa trên trạng thái kênh và hành vi di chuyển của phương tiện [32], [33]. Các giải pháp này cho thấy khả năng cải thiện QoS/QoE trong một số kịch bản động, tuy nhiên chúng chủ yếu hoạt động ở tầng truyền dẫn và chưa giải quyết được các hạn chế mang tính kiến trúc của mô hình phân phối nội dung.

Sự kết hợp giữa mật độ RSU thưa và kết nối không ổn định dẫn đến hai hệ quả quan trọng được ghi nhận rộng rãi trong các nghiên cứu trước đây. Thứ nhất, độ trễ đầu-cuối của dịch vụ video tăng cao, do nội dung thường phải được truy xuất từ các RSU xa hơn hoặc trực tiếp từ mạng lõi khi không có sẵn tại RSU gần người dùng [27], [29], [34]. Thứ hai, tải truyền trên các liên kết mạng lõi trở nên lớn, khi nhiều phương tiện đồng thời yêu cầu video từ máy chủ trung tâm hoặc RSU, dẫn đến tắc nghẽn mạng

và suy giảm QoE, đặc biệt trong các kịch bản giờ cao điểm hoặc khu vực giao thông đông đúc [27], [28], [34].

Mặc dù nhiều nghiên cứu đã đề xuất các giải pháp nhằm cải thiện QoS/QoE cho dịch vụ video trong VANET, các mô hình truyền thống dựa vào máy chủ trung tâm hoặc RSU vẫn gặp phải những hạn chế mang tính cấu trúc. Mật độ RSU thưa và tính không ổn định của kết nối dẫn đến độ trễ cao và tải mạng lõi lớn, trong khi các giải pháp thích nghi ở tầng truyền dẫn chỉ mang tính cải thiện cục bộ. Điều này cho thấy nhu cầu cấp thiết về các kiến trúc phân phối nội dung hiệu quả hơn, có khả năng giảm phụ thuộc vào mạng lõi và thích ứng tốt với môi trường mạng động.

1.2.2. Lưu trữ và lưu trữ tại biên trong VANET

Nhằm khắc phục các hạn chế về độ trễ cao và tải mạng lõi lớn của các mô hình truyền video truyền thống trong VANET, nhiều nghiên cứu đã đề xuất khai thác lưu trữ nội dung và lưu trữ tại biên như một giải pháp hiệu quả. Ý tưởng chính của hướng tiếp cận này là đưa các nội dung phổ biến xuống gần người dùng hơn, từ đó giảm khoảng cách truyền dẫn, rút ngắn thời gian phục vụ và giảm lưu lượng truyền tải trên mạng lõi.

Một nhóm lớn các công trình tập trung vào lưu trữ tại RSU, xem RSU như các nút lưu trữ biên có khả năng phục vụ trực tiếp các phương tiện trong vùng phủ sóng. Các nghiên cứu này chỉ ra rằng việc lưu trữ các nội dung video phổ biến tại RSU giúp giảm đáng kể thời gian truy xuất nội dung và giảm tải cho các liên kết mạng lõi kết nối tới máy chủ trung tâm [28], [34]. Trong bối cảnh dịch vụ video nhạy cảm với độ trễ, cách tiếp cận này mang lại cải thiện rõ rệt về QoS/QoE so với các mô hình truyền thống. Tuy nhiên, hiệu quả của lưu trữ tại RSU phụ thuộc rất nhiều vào mật độ và sự phân bố của RSU. Trong các kịch bản hạ tầng thưa hoặc phân bố không đồng đều, phương tiện có thể nhanh chóng rời khỏi vùng phủ của RSU trước khi hoàn tất quá trình truyền video, làm giảm lợi ích của lưu trữ [30], [31]. Ngoài ra, nhiều nghiên cứu giả định dung lượng lưu trữ của RSU là cố định và tương đối lớn, trong khi trên thực tế tài nguyên lưu trữ và băng thông tại RSU vẫn bị giới hạn.

Mặc dù lưu trữ và lưu trữ tại biên đã được chứng minh là giải pháp hiệu quả để giảm độ trễ và tải mạng lõi trong VANET, các nghiên cứu hiện tại chủ yếu tập trung vào một tầng lưu trữ riêng lẻ hoặc tối ưu lưu trữ một cách tách biệt. Việc tối ưu đồng

thời vị trí lưu trữ và cơ chế phân phối nội dung, đặc biệt trong bối cảnh mạng động và hạ tầng không đồng đều, vẫn chưa được giải quyết triệt để.

Song song với hướng tiếp cận dựa trên RSU, một số nghiên cứu khác đề xuất lưu trữ tại phương tiện, khai thác số lượng lớn các xe như những nút lưu trữ phân tán trong mạng VANET [32], [35]. Cách tiếp cận này tận dụng mật độ phương tiện cao trong đô thị để tăng khả năng tìm thấy nội dung được lưu trữ gần người dùng, đồng thời giảm phụ thuộc vào hạ tầng cố định. Các công trình trong hướng này cho thấy lưu trữ tại phương tiện có thể cải thiện hiệu năng hệ thống trong các kịch bản giao thông đông đúc, đặc biệt khi nội dung có mức độ phổ biến cao và hành vi di chuyển của phương tiện có tính lặp lại [35]. Tuy nhiên, do tính di động cao và kết nối không ổn định giữa các phương tiện, việc đảm bảo độ tin cậy và khả năng sẵn sàng của nội dung vẫn là một thách thức lớn. Ngoài ra, dung lượng lưu trữ trên mỗi phương tiện thường hạn chế và chịu ràng buộc bởi các yếu tố như chi phí và tiêu thụ năng lượng.

Để tận dụng ưu điểm của cả hai hướng tiếp cận trên, một số nghiên cứu đã đề xuất các mô hình lưu trữ hợp tác giữa RSU và phương tiện, trong đó nội dung được phân bố linh hoạt giữa các tầng lưu trữ [34], [35]. Các mô hình này cho phép RSU lưu trữ các nội dung có mức độ phổ biến cao trên phạm vi rộng, trong khi phương tiện lưu trữ các nội dung mang tính cục bộ hoặc theo ngữ cảnh. Kết quả từ các nghiên cứu cho thấy lưu trữ hợp tác giúp cải thiện tỷ lệ truy xuất nội dung lưu trữ thành công và giảm độ trễ so với các mô hình lưu trữ đơn lẻ. Tuy nhiên, nhiều công trình mới chỉ xem xét bài toán lưu trữ ở mức độ tĩnh hoặc bán tĩnh, trong khi chưa xét đầy đủ sự biến thiên mạnh của nhu cầu nội dung theo không gian và thời gian trong môi trường VANET thực tế.

1.2.3. Lưu trữ hợp tác và lưu trữ xác suất

Trong bối cảnh mạng VANET có tính động cao và cấu trúc liên kết thay đổi nhanh, các mô hình lưu trữ xác định (DC – Deterministic Caching) bộc lộ nhiều hạn chế về khả năng thích ứng. Để khắc phục điều này, một số nghiên cứu đã đề xuất các cơ chế lưu trữ hợp tác (cooperative caching) và lưu trữ xác suất (PC – Probabilistic Caching), trong đó quyết định lưu trữ nội dung không còn mang tính cứng nhắc mà được điều chỉnh linh hoạt dựa trên đặc tính nội dung và hành vi mạng.

Các mô hình lưu trữ hợp tác được đề xuất nhằm khai thác sự phối hợp giữa nhiều nút mạng, bao gồm V2V và V2R (Vehicle to RSU), để nâng cao hiệu quả sử dụng tài nguyên lưu trữ phân tán. Trong các nghiên cứu này, nội dung được phân bố giữa các nút mạng theo cách bổ sung lẫn nhau, giảm trùng lặp và tăng khả năng phục vụ cục bộ [34], [35]. Một số công trình tập trung vào lưu trữ hợp tác giữa các phương tiện, trong đó các xe chia sẻ nội dung video với nhau thông qua các liên kết V2V khi gặp gỡ trên đường. Cách tiếp cận này đặc biệt hiệu quả trong các kịch bản mật độ xe cao, nơi xác suất tiếp xúc giữa các phương tiện lớn, giúp tăng tỷ lệ truy xuất nội dung lưu trữ thành công và giảm phụ thuộc vào hạ tầng cố định [35]. Tuy nhiên, hiệu quả của các mô hình này giảm mạnh trong các kịch bản mật độ xe thấp hoặc hành vi di chuyển không đồng nhất.

Bên cạnh đó, một số nghiên cứu cải tiến mô hình hợp tác lưu trữ giữa RSU và phương tiện, trong đó, RSU lưu trữ các nội dung phổ biến trên phạm vi rộng, còn phương tiện lưu trữ các nội dung mang tính cục bộ hoặc theo ngữ cảnh [34]. Các kết quả mô phỏng cho thấy mô hình này cải thiện đáng kể độ trễ và QoS/QoE so với lưu trữ đơn lẻ. Để tăng tính linh hoạt hơn nữa, các mô hình lưu trữ xác suất được đề xuất, trong đó mỗi nút mạng lưu trữ nội dung với một xác suất nhất định, thường dựa trên độ phổ biến của nội dung hoặc lịch sử truy cập [32]. Cách tiếp cận này giúp phân tán nội dung trong mạng một cách tự nhiên, tránh tình trạng lưu trữ trùng lặp quá mức và giảm chi phí điều phối giữa các nút.

Các nghiên cứu cho thấy lưu trữ xác suất đặc biệt phù hợp với môi trường VANET, nơi cấu trúc mạng thay đổi nhanh và khó nắm bắt được thông tin toàn bộ trạng thái mạng. Hơn nữa, do bản chất ngẫu nhiên, các mô hình này khó đảm bảo QoS/QoE trong các kịch bản yêu cầu độ trễ thấp và ổn định cao, chẳng hạn như truyền video thời gian thực. Ngoài ra, việc lựa chọn xác suất lưu trữ tối ưu thường phụ thuộc vào các giả định đơn giản về phân bố nhu cầu nội dung và hành vi di chuyển, làm hạn chế khả năng áp dụng trong các kịch bản thực tế phức tạp.

Mặc dù các mô hình lưu trữ hợp tác và theo xác suất giúp cải thiện tính linh hoạt và khả năng thích ứng so với lưu trữ xác định, chúng vẫn đối mặt với nhiều thách thức. Thứ nhất, hiệu quả của các mô hình này phụ thuộc mạnh vào mật độ RSU và phương

tiện, cũng như mức độ ổn định của các liên kết truyền thông. Thứ hai, chi phí điều phối và trao đổi thông tin giữa các nút mạng thường bị bỏ qua hoặc đánh giá thấp trong nhiều nghiên cứu. Thứ ba, các công trình chưa tối ưu đồng thời lưu trữ và phân phối nội dung, mà chỉ tập trung vào quyết định lưu trữ. Cuối cùng, các nghiên cứu hiện tại vẫn phụ thuộc nhiều vào hạ tầng mặt đất mà chưa khai thác kênh truyền thông không-đất hỗ trợ bởi UAV.

1.2.4. Mạng VANET có UAV hỗ trợ

Trước những hạn chế của hạ tầng mặt đất trong VANET, đặc biệt khi mật độ RSU thưa và có thể phân bố không đồng đều ở nhiều tuyến đường khác nhau, việc tích hợp UAV đã được đề xuất như một hướng tiếp cận đầy tiềm năng. Nhờ khả năng cơ động cao và triển khai linh hoạt, UAV có thể nhanh chóng mở rộng vùng phủ sóng và hỗ trợ truyền thông trong các kịch bản mà hạ tầng cố định không đáp ứng được.

Trong nhiều nghiên cứu, UAV được xem như trạm chuyển tiếp hoặc trạm gốc di động, đóng vai trò trung gian giữa các phương tiện và mạng lõi. Nhờ xác suất thiết lập liên kết nhìn thẳng (LoS – Line of Sight) cao, UAV giúp cải thiện chất lượng kênh truyền và giảm suy hao tín hiệu so với các liên kết mặt đất, đặc biệt trong môi trường đô thị nhiều vật cản [36], [37]. Một số công trình tập trung vào tối ưu vị trí hoặc quỹ đạo bay của UAV nhằm cực đại vùng phủ hoặc cải thiện QoS/QoE cho các phương tiện dưới mặt đất [38], [39]. Các kết quả cho thấy việc triển khai UAV có thể giảm đáng kể độ trễ truyền thông và cải thiện QoS/QoE trong các khu vực hạ tầng hạn chế hoặc trong các tình huống khẩn cấp.

Gần đây hơn, các nghiên cứu bắt đầu mở rộng vai trò của UAV sang lưu trữ nội dung và MEC, trong đó UAV không chỉ truyền dữ liệu mà còn lưu trữ và xử lý nội dung video ngay trên không [40], [41]. Cách tiếp cận này cho phép nội dung phổ biến được phục vụ trực tiếp từ UAV, giảm phụ thuộc vào mạng lõi và rút ngắn thời gian phục vụ. Tuy nhiên, việc tích hợp lưu trữ và tính toán trên UAV đặt ra nhiều thách thức mới, bao gồm giới hạn năng lượng của pin, dung lượng lưu trữ hạn chế và độ phức tạp trong điều phối nhiều UAV. Ngoài ra, các công trình hiện tại chỉ xem xét tối ưu một trong các yếu

tổ như vị trí UAV, truyền thông hoặc lưu trữ, trong khi chưa có một khuôn khổ tối ưu thống nhất.

Mặc dù UAV-VANET cho thấy tiềm năng lớn, các nghiên cứu hiện tại vẫn còn nhiều hạn chế. Thứ nhất, nhiều công trình giả định số lượng UAV nhỏ và môi trường lý tưởng, chưa phản ánh đầy đủ các kịch bản thực tế phức tạp. Thứ hai, sự tương tác giữa UAV và các tầng lưu trữ khác (RSU, phương tiện) thường chưa được mô hình hóa đầy đủ. Thứ ba, các vấn đề về an toàn, độ tin cậy và quản lý năng lượng dài hạn vẫn chưa được giải quyết triệt để. Cuối cùng, các nghiên cứu hiện tại chưa tối ưu đồng thời lưu trữ nội dung, phân phối video và vị trí/quỹ đạo UAV trong một khuôn khổ thống nhất. Đặc biệt, việc kết hợp lưu trữ đa tầng với động học không gian 3D của UAV để đảm bảo QoS/QoE trong môi trường mạng động cao vẫn là một hướng nghiên cứu mở.

1.2.5. Khoảng trống nghiên cứu

Nhiều nghiên cứu hiện có về UAV-VANET xem xét từng tầng lưu trữ riêng lẻ, tại RSU, tại phương tiện, hoặc tại UAV, mà chưa có khung tối ưu kết hợp đồng thời ba tầng này dưới một hàm mục tiêu thống nhất. Mặc dù UAV cung cấp khả năng mở rộng vùng phủ trong không gian ba chiều, các nghiên cứu hiện có chưa kết hợp tối ưu vị trí UAV với chiến lược lưu trữ và phân phối nội dung trong một khung tối ưu thống nhất. Đây là khoảng trống nghiên cứu mà luận án nhằm giải quyết qua mô hình CAHO. Trong UAV-VANET, truyền video là loại dịch vụ có yêu cầu cao về độ trễ và tính liên tục của quá trình phân phối nội dung. Tuy vậy, các nghiên cứu hiện có chủ yếu tiếp cận bài toán tối ưu hiệu năng thông qua các chỉ tiêu gián tiếp như tỷ lệ cache hit, thông lượng truyền dẫn hoặc vị trí triển khai UAV, trong khi thời gian phục vụ video chưa được đặt làm hàm mục tiêu tối ưu trực tiếp.

Từ những phân tích trên có thể thấy rằng, việc lấp đầy các khoảng trống nghiên cứu là cần thiết, nhằm xây dựng một giải pháp tối ưu kết hợp cho hệ thống UAV-VANET, vừa khai thác hiệu quả tài nguyên lưu trữ và truyền thông đa tầng, vừa đáp ứng các yêu cầu ngày càng khắt khe về độ trễ và QoS/QoE của các dịch vụ truyền video hướng đến thời gian thực trong các hệ thống giao thông thông minh thế hệ mới.

1.3. GIỚI THIỆU VỀ CÁC KỸ THUẬT TỐI ƯU

1.3.1. Giới thiệu về tối ưu

Trong các hệ thống mạng hiện đại, đặc biệt là các mạng có tính động cao và cấu trúc phức tạp như UAV-VANET, tối ưu hóa đóng vai trò trung tâm trong việc khai thác hiệu quả các tài nguyên hạn chế của hệ thống. Bản chất của bài toán không chỉ nằm ở việc cải thiện một chỉ số riêng lẻ, mà là tìm kiếm sự cân bằng tối ưu giữa nhiều yếu tố cạnh tranh, bao gồm tài nguyên lưu trữ, băng thông truyền thông, năng lượng, cũng như yêu cầu ngày càng khắt khe về độ trễ và QoE.

Trong bối cảnh dịch vụ truyền video, bài toán tối ưu trở nên đặc biệt thách thức do đặc tính nhạy cảm với độ trễ và sự biến thiên mạnh theo không gian và thời gian của cả nhu cầu nội dung lẫn trạng thái mạng. Các quyết định tối ưu, chẳng hạn như vị trí lưu trữ nội dung, cơ chế phân phối hay vị trí triển khai các nút hỗ trợ, đều có mối quan hệ chặt chẽ và ảnh hưởng lẫn nhau. Do đó, việc tối ưu từng thành phần một cách độc lập thường không mang lại hiệu quả toàn cục, mà đòi hỏi một khung tối ưu hóa tổng thể, trong đó các quyết định được xem xét đồng thời và phối hợp một cách nhất quán.

Hơn nữa, bài toán tối ưu trong UAV-VANET thường mang tính phi tuyến, không lồi và có không gian tìm kiếm lớn, do sự kết hợp giữa các biến rời rạc (quyết định lưu trữ) và các biến liên tục (vị trí dừng của UAV). Những đặc điểm này khiến các phương pháp tối ưu truyền thống dựa trên giải tích gặp nhiều hạn chế, đồng thời mở ra nhu cầu sử dụng các thuật toán tối ưu xấp xỉ và heuristic, có khả năng tìm kiếm nghiệm tốt trong thời gian tính toán chấp nhận được.

Từ góc nhìn đó, việc xây dựng bài toán tối ưu không chỉ nhằm tìm ra nghiệm tối ưu về mặt toán học, mà còn hướng tới tính khả thi và khả năng triển khai trong thực tế. Một mô hình tối ưu hiệu quả cần phản ánh đúng đặc tính của hệ thống, cân nhắc các ràng buộc thực tế, và đồng thời cho phép đánh giá rõ ràng mức độ cải thiện hiệu năng thông qua các chỉ số như thời gian phục vụ, độ trễ và QoE. Đây chính là nền tảng để phát triển các giải pháp tối ưu có ý nghĩa thực tiễn cho dịch vụ truyền video trong UAV-VANET.

1.3.1. Các phương pháp tối ưu trong UAV-VANET

1) Phương pháp tối ưu lời

Tối ưu lời là một phương pháp được sử dụng rộng rãi trong thiết kế hệ thống truyền thông không dây nhờ khả năng tìm nghiệm tối ưu toàn cục trong thời gian tính toán đa thức. Trong UAV-VANET, một số bài toán con như phân bổ công suất, băng thông hoặc tài nguyên truyền thông có thể được mô hình hóa dưới dạng lời và giải bằng tối ưu lời với độ ổn định và khả năng hội tụ được đảm bảo về mặt lý thuyết [42].

Trong nhiều thập kỷ qua, lý thuyết tối ưu lời đã chứng minh vai trò nền tảng trong xử lý tín hiệu và truyền thông không dây, và được áp dụng rộng rãi cho các bài toán thiết kế mạng, quản lý tài nguyên và phân tích hiệu năng. Sự phong phú của các tầng bài toán lời và các công cụ giải sẵn có đã giúp tối ưu hóa lời trở thành một xu hướng chủ đạo trong thiết kế hạ tầng mạng không dây hiện đại [43]. Trong bối cảnh UAV-VANET, tối ưu hóa lời thường được sử dụng để xây dựng bài toán xấp xỉ hoặc ràng buộc cho các bài toán gốc vốn mang tính không lời, từ đó cung cấp các nghiệm heuristic có chất lượng tốt và mang lại cái nhìn sâu sắc về cấu trúc nghiệm tối ưu. Ngược lại, đối với các bài toán không lời thuần túy, việc thiếu các công cụ giải hiệu quả khiến khó có thể đảm bảo tìm được nghiệm tối ưu toàn cục, đặc biệt trong các kịch bản mạng động và quy mô lớn như UAV-VANET.

2) Phương pháp tối ưu dựa trên các kỹ thuật heuristic

Trong các hệ thống UAV-VANET thực tế, nhiều bài toán tối ưu mang tính phi tuyến, không lời và có không gian tìm kiếm rất lớn, do sự kết hợp giữa các biến rời rạc (quyết định lưu trữ, gán nội dung) và các biến liên tục (vị trí, quỹ đạo UAV). Các thuật toán tối ưu lặp truyền thống thường không hiệu quả khi áp dụng cho các bài toán này, do độ phức tạp tính toán tăng nhanh theo quy mô hệ thống [44]. Vì vậy, các phương pháp tối ưu heuristic, mô phỏng hành vi tự nhiên của sinh vật, đã được áp dụng rộng rãi trong UAV-VANET.

Các thuật toán như tối ưu bầy đàn (PSO – Particle Swarm Optimization), giải thuật di truyền (GA – Genetic Algorithm), hay thuật toán tối ưu cá voi (WOA – Whale Optimization Algorithm) cho phép tìm kiếm nghiệm gần tối ưu trong không gian nghiệm lớn với chi phí tính toán chấp nhận được [45]. Trong UAV-VANET, các phương

pháp này đặc biệt phù hợp để giải các bài toán như tối ưu vị trí và quỹ đạo UAV, phân phối nội dung, hoặc tối ưu hóa đa mục tiêu giữa độ trễ, năng lượng và vùng phủ, nơi các ràng buộc thực tế khiến việc áp dụng tối ưu hóa lỗi trở nên khó khăn.

3) Phương pháp tối ưu dựa trên học máy và trí tuệ nhân tạo.

Gần đây, trí tuệ nhân tạo (AI - Artificial Intelligence) và học máy (ML - Machine Learning) đã nổi lên như một hướng tiếp cận đầy tiềm năng cho các bài toán tối ưu trong mạng không dây, đặc biệt là trong các hệ thống phức tạp và có tính động như UAV-VANET [18]. Thông qua khả năng học từ dữ liệu lớn, các mô hình ML có thể dự đoán trạng thái mạng, nhu cầu nội dung và hành vi di chuyển, từ đó hỗ trợ đưa ra các quyết định tối ưu mà không cần xây dựng mô hình toán học chính xác cho toàn bộ hệ thống.

Trong bối cảnh UAV-VANET, nơi thông tin hệ thống thường không hoàn hảo và thay đổi nhanh theo thời gian, AI và ML cung cấp một hướng tiếp cận mới để thiết kế và vận hành mạng, bổ sung hoặc thay thế các phương pháp tối ưu truyền thống. Các kỹ thuật này được kỳ vọng sẽ đóng vai trò quan trọng trong các hệ thống UAV-VANET tương lai, cho phép đạt được hiệu năng vượt trội thông qua việc kết hợp tối ưu hóa dựa trên mô hình và tối ưu hóa dựa trên dữ liệu, đặc biệt trong các bài toán phân phối video và đảm bảo QoE.

1.3.2. So sánh các phương pháp và giải pháp GA

Các phương pháp tối ưu trong UAV-VANET có thể được phân thành ba nhóm chính: tối ưu lỗi, phương pháp heuristic và phương pháp dựa trên học máy/trí tuệ nhân tạo. Mỗi nhóm phương pháp có ưu điểm và hạn chế riêng, tùy thuộc vào cấu trúc bài toán, loại biến tối ưu và yêu cầu tính toán.

Tối ưu lỗi có ưu điểm là cho phép tìm nghiệm tối ưu toàn cục với cơ sở toán học chặt chẽ và chi phí tính toán thấp khi bài toán có cấu trúc lỗi rõ ràng [43]. Tuy nhiên, các bài toán lưu trữ và phân phối video trong UAV-VANET thường có tính phi tuyến, không lồi và bao gồm nhiều biến rời rạc như quyết định lưu trữ nội dung, lựa chọn nút phục vụ hoặc vị trí dừng UAV. Vì vậy, tối ưu lỗi khó áp dụng trực tiếp cho bài toán gốc, mà thường chỉ phù hợp cho các bài toán con hoặc các mô hình xấp xỉ.

Các phương pháp học máy và trí tuệ nhân tạo có khả năng thích nghi với môi trường mạng động, đặc biệt khi hệ thống có dữ liệu lớn và thông tin trạng thái thay đổi theo thời gian [18]. Tuy nhiên, nhóm phương pháp này thường yêu cầu dữ liệu huấn luyện đủ lớn, quá trình huấn luyện phức tạp và khó bảo đảm tính ổn định cũng như khả năng giải thích nghiệm trong mọi kịch bản. Do đó, chúng phù hợp hơn với các hệ thống có dữ liệu vận hành thực tế phong phú hoặc các bài toán điều khiển thích nghi dài hạn.

Trong khi đó, các phương pháp heuristic như PSO, GA hoặc WOA cho phép tìm nghiệm gần tối ưu trong không gian nghiệm lớn với chi phí tính toán chấp nhận được [44], [45]. Nhóm phương pháp này phù hợp với các bài toán hỗn hợp nguyên–thực hoặc nguyên–nhị phân, nơi phương pháp tối ưu chính xác hoặc vét cạn không khả thi do số lượng tổ hợp nghiệm tăng nhanh theo quy mô hệ thống.

Trong luận án này, GA được lựa chọn vì phù hợp với đặc tính của hai bài toán tối ưu DPC và CAHO. Đối với DPC, nghiệm cần mã hóa đồng thời biến nguyên cho bán kính lưu trữ tại RSU và biến thực cho xác suất lưu trữ tại phương tiện. Đối với CAHO, nghiệm tiếp tục được mở rộng với các biến nhị phân biểu diễn quyết định lưu trữ tại UAV và vị trí dừng của UAV. GA có khả năng xử lý linh hoạt các dạng biến này thông qua mã hóa nhiễm sắc thể, đồng thời cho phép tích hợp các ràng buộc tài nguyên bằng hàm phạt.

Mặc dù GA không bảo đảm luôn tìm được nghiệm tối ưu toàn cục chính xác và có thể yêu cầu nhiều thế hệ để hội tụ, phương pháp này phù hợp với mục tiêu của luận án là tìm nghiệm gần tối ưu cho các bài toán phi tuyến, không lồi và có không gian tìm kiếm lớn. Trong các chương tiếp theo, GA được sử dụng để giải hai bài toán tối ưu DPC và CAHO, đồng thời hiệu quả hội tụ và chất lượng nghiệm được đánh giá thông qua mô phỏng.

1.4. LỘ TRÌNH NGHIÊN CỨU

1.4.1. Bài toán của luận án

Mặc dù đã có nhiều nghiên cứu về lưu trữ nội dung trong mạng VANET và UAV-VANET, vẫn còn tồn tại một số hạn chế đáng chú ý, cụ thể như sau.

Trước hết, nhiều công trình trước đây thường tập trung vào một tầng lưu trữ riêng lẻ, chẳng hạn như chỉ tối ưu chiến lược lưu trữ tại các RSU hoặc chỉ khai thác khả năng lưu trữ của các phương tiện di động. Cách tiếp cận này chưa phản ánh đầy đủ bản chất

hợp tác của mạng giao thông thông minh, nơi nhiều loại nút khác nhau có thể cùng tham gia vào quá trình phân phối nội dung.

Thứ hai, các nghiên cứu hiện có thường đánh giá hiệu năng hệ thống thông qua các chỉ tiêu trung gian như tỉ lệ truy xuất nội dung lưu trữ thành công (cache hit ratio), thông lượng hoặc mức sử dụng tài nguyên, trong khi các đại lượng này chưa phản ánh trực tiếp trải nghiệm dịch vụ của người dùng cuối, đặc biệt đối với các ứng dụng truyền video có yêu cầu độ trễ nghiêm ngặt. Bên cạnh đó, trong các nghiên cứu về UAV-VANET, UAV thường được sử dụng chủ yếu như các nút chuyển tiếp hoặc các trạm phát tạm thời, trong khi mối quan hệ giữa chiến lược lưu trữ nội dung và vị trí hoạt động của UAV vẫn chưa được xem xét một cách đầy đủ. Những hạn chế này cho thấy vẫn còn thiếu một khung tiếp cận thống nhất cho bài toán lưu trữ và phân phối nội dung trong mạng giao thông thông minh.

Xuất phát từ những vấn đề nêu trên, luận án đặt ra bài toán nghiên cứu là xây dựng các mô hình tối ưu lưu trữ nội dung nhằm giảm thời gian phục vụ dịch vụ video trong mạng VANET và UAV-VANET, trong đó các quyết định lưu trữ được xem xét đồng thời với cấu trúc mạng và điều kiện tài nguyên của hệ thống. Khác với nhiều nghiên cứu trước, luận án lựa chọn thời gian phục vụ video làm tiêu chí tối ưu trung tâm. Đại lượng này phản ánh trực tiếp khoảng thời gian mà một phương tiện cần để nhận được nội dung video yêu cầu, gắn liền với cả chất lượng dịch vụ của hệ thống (QoS) và trải nghiệm người dùng (QoE) trong các dịch vụ nhạy cảm với trì hoãn. Trên cơ sở đó, bài toán được mô hình hóa dưới dạng một bài toán tối ưu với các thông số về băng thông truyền dẫn, đặc tính kênh truyền và phân bố phương tiện trong mạng giao thông dưới ràng buộc về dung lượng lưu trữ. Việc giải bài toán đòi hỏi một cách tiếp cận linh hoạt có khả năng xử lý không gian nghiệm phức tạp, trong đó, các quyết định lưu trữ và cấu hình hệ thống có mối quan hệ phụ thuộc lẫn nhau.

Trên nền tảng này, luận án từng bước phát triển các mô hình nghiên cứu theo một trình tự logic từ môi trường VANET hai chiều đến UAV-VANET ba chiều. Trước hết, bài toán lưu trữ nội dung trong mạng VANET với hạ tầng RSU được xây dựng nhằm xác định chiến lược lưu trữ nội dung tại các RSU theo mục tiêu giảm thời gian phục vụ video cho các phương tiện đang di chuyển trên đường. Mô hình này đóng vai trò như một bước nền tảng, giúp phân tích ảnh hưởng của vị trí lưu trữ nội dung tại hạ tầng mặt

đạt đến hiệu năng phân phối video trong mạng giao thông. Để mô hình thực tế hơn, bài toán được mở rộng bằng cách khai thác khả năng lưu trữ của các phương tiện di động thông qua mô hình Deterministic and Probabilistic Caching (DPC). Trong mô hình DPC, chiến lược lưu trữ xác định được thực hiện tại RSU và chiến lược lưu trữ xác suất được thực hiện trên phương tiện tích hợp trong cùng một khung tối ưu thống nhất. Cách tiếp cận này cho phép khai thác đồng thời tính ổn định của hạ tầng RSU và tính linh hoạt của các phương tiện di động, từ đó cải thiện khả năng phân phối nội dung thông qua truyền thông Vehicle-to-Vehicle (V2V). Việc kết hợp hai tầng lưu trữ trong cùng một mô hình tối ưu giúp hệ thống tận dụng hiệu quả hơn các tài nguyên sẵn có trong môi trường giao thông đông đúc.

Trên cơ sở mô hình DPC, luận án tiếp tục khảo sát phạm vi nghiên cứu với môi trường UAV-VANET, nơi các UAV được đưa vào hệ thống như các nút mạng trên không có khả năng lưu trữ và truyền dữ liệu linh hoạt. Đóng góp quan trọng tiếp theo của luận án là mô hình Caching and Hovering Optimization (CAHO). Trong CAHO, UAV được định vị như một nút lưu trữ di động bù đắp các khoảng trống về vùng phủ và năng lực lưu trữ giữa các RSU mặt đất. Mô hình CAHO tối ưu đồng thời ba biến: bán kính lưu trữ tại RSU, quyết định lưu trữ tại UAV và vị trí dừng UAV trong không gian — phù hợp với các khu vực có tỷ lệ phương tiện sẵn sàng chia sẻ tài nguyên thấp hoặc hạ tầng RSU chưa được triển khai đầy đủ.

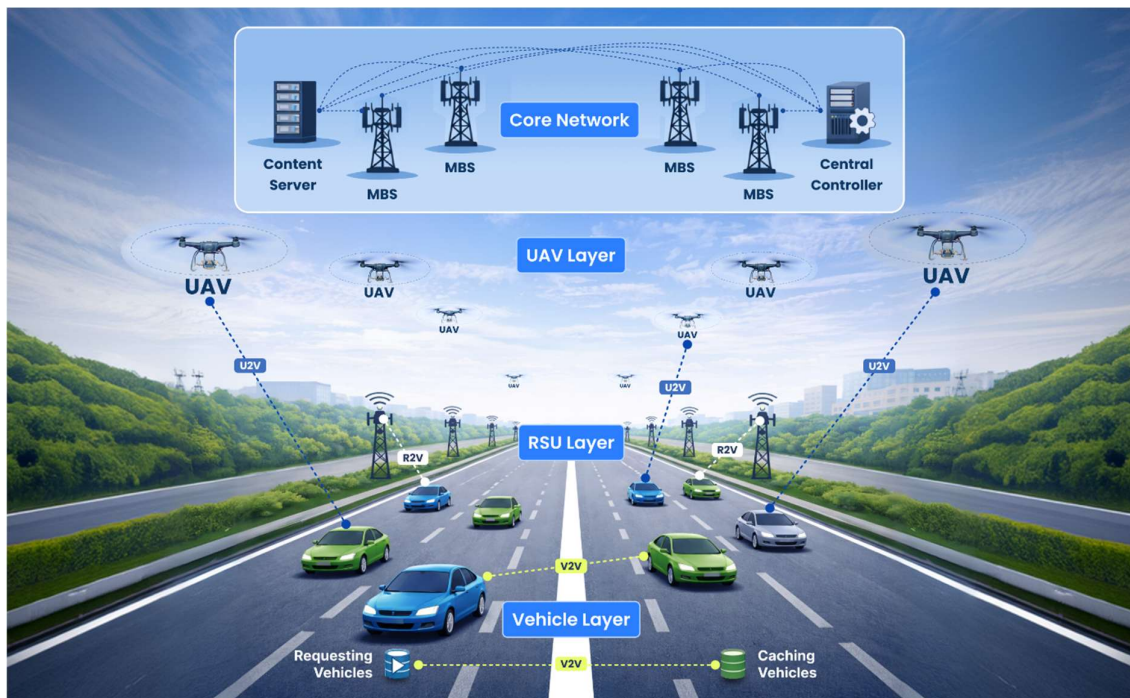
Các kết quả mô phỏng được thực hiện trong luận án cho thấy các mô hình đề xuất có thể giảm đáng kể thời gian phục vụ video so với các chiến lược lưu trữ truyền thống. Đồng thời, việc triển khai mô hình từ cơ bản đến DPC và CAHO, cho thấy một quá trình phát triển bài toán một cách logic: từ việc tối ưu lưu trữ tại RSU, đến việc khai thác khả năng lưu trữ của các phương tiện di động, và cuối cùng là tầng mạng trên không với UAV. Cách tiếp cận này không chỉ giúp cải thiện hiệu năng hệ thống trong các kịch bản giao thông khác nhau, phù hợp theo từng hiện trạng mạng khác nhau, mà còn góp phần xây dựng một khung tối ưu thống nhất cho bài toán phân phối nội dung video trong mạng giao thông thông minh thế hệ tiếp theo.

1.4.2. Mô hình nghiên cứu

Hệ thống được xem xét trong nghiên cứu này là một mô hình truyền video hợp tác trong mạng VANET và UAV-VANET, được xây dựng nhằm tối ưu việc lưu trữ và

phân phối nội dung video cho các phương tiện đang di chuyển trên đường. Như minh họa trong hình 1.3, hệ thống bao gồm bốn thành phần chính: tầng mạng lõi (MBS – Macro Base Station), tầng thiết bị bay không người lái (UAV), tầng trạm phát ven đường (RSU) và tầng phương tiện giao thông (xe).

Ở tầng trên cùng, máy chủ lưu trữ nội dung các video và kết nối với mạng lõi thông qua trạm gốc (MBS) hoặc hạ tầng mạng tương đương. Thành phần này đóng vai trò thu thập thông tin hệ thống như mức độ phổ biến của video, trạng thái lưu trữ của các nút trong mạng và thông tin lưu lượng giao thông. Dựa trên các thông tin này, bộ điều khiển trung tâm thực hiện quá trình tối ưu hóa và phân phối quyết định lưu trữ nội dung đến các nút trong mạng.



Hình 1.3: Mô hình nghiên cứu

Tầng thứ hai là tầng thiết bị bay không người lái (UAV), bao gồm các UAV hoạt động như các nút mạng trên không. Các UAV có khả năng lưu trữ nội dung video và cung cấp dịch vụ truyền dữ liệu cho các phương tiện dưới mặt đất thông qua liên kết UAV-to-Vehicle (U2V). Ngoài ra, UAV còn có thể đóng vai trò là các nút chuyển tiếp hoặc mở rộng vùng phủ sóng trong những khu vực mà hạ tầng RSU không đáp ứng đủ

nhu cầu. Trong trường hợp cần thiết, vị trí dừng của các UAV có thể được điều chỉnh nhằm tối ưu khả năng phục vụ cho các phương tiện trong vùng.

Bên dưới tầng UAV là tầng hạ tầng ven đường (RSU). Ở tầng này, các RSU được bố trí dọc theo tuyến đường. Mỗi RSU có một vùng phủ sóng nhất định và được trang bị bộ nhớ để lưu trữ các video. Các RSU có thể truyền dữ liệu trực tiếp cho các phương tiện thông qua liên kết RSU-to-Vehicle (R2V) và cũng có thể phối hợp với các RSU lân cận hoặc với UAV để phối hợp cung cấp video cho người dừng.

Dưới cùng là tầng phương tiện giao thông di chuyển trên tuyến đường. Trong tầng này, các phương tiện có thể được chia thành hai loại: phương tiện yêu cầu (RV - Requesting Vehicle), là các xe yêu cầu video, và phương tiện lưu trữ (CV - Caching Vehicle), là các xe có khả năng lưu trữ và phân phối video cho các xe khác thông qua kết nối Vehicle-to-Vehicle (V2V). Nhờ cơ chế truyền dữ liệu hợp tác giữa các phương tiện, các video có thể được phân phối hiệu quả hơn ngay cả khi phương tiện rời khỏi vùng phủ sóng của RSU.

Thông qua cấu trúc phân tầng này, hệ thống có thể linh hoạt khai thác đồng thời hạ tầng cố định (RSU), các nút di động (phương tiện giao thông) và các nút trên không (UAV), từ đó cải thiện hiệu quả khả năng lưu trữ và phân phối video trong các kịch bản giao thông thông minh. Trên mô hình hệ thống tổng quát này, các chiến lược lưu trữ được nghiên cứu trong luận án được xem như các trường hợp đặc biệt của cùng một kiến trúc mạng, tùy thuộc vào đặc điểm của môi trường mạng và điều kiện hạ tầng triển khai. Trong luận án, các video luôn được lưu trữ tối ưu tại các RSU đồng thời kết hợp lưu trữ hợp tác và phân phối với các thành phần khác trong mạng.

Mô hình DPC (Deterministic and Probabilistic Caching) phù hợp với các môi trường giao thông có mật độ phương tiện cao và ổn định, nơi các phương tiện có khả năng lưu trữ và tham gia vào quá trình truyền dữ liệu hợp tác. Trong kịch bản này, các RSU đóng vai trò là các nút lưu trữ ổn định với chiến lược lưu trữ xác định (DC), trong khi các phương tiện có thể lưu trữ nội dung thông qua cơ chế lưu trữ xác suất (PC). Nhờ mật độ phương tiện lớn và khả năng kết nối V2V thường xuyên, các phương tiện có thể trao đổi nội dung video trực tiếp với nhau, từ đó giảm tải cho hạ tầng RSU và cải thiện hiệu quả phân phối nội dung. Do đó, mô hình DPC đặc biệt phù hợp với các khu vực đô thị đông đúc, các tuyến đường có lưu lượng giao thông lớn hoặc các kịch bản

giao thông thông minh nơi các phương tiện được trang bị đầy đủ khả năng tính toán và lưu trữ mạnh.

Ngược lại, mô hình CAHO (CAching and Hovering Optimization) được thiết kế cho các kịch bản mà hạ tầng mặt đất hoặc mật độ phương tiện có khả năng lưu trữ không đủ (cơ hội truyền dữ liệu V2V bị hạn chế) để đảm bảo phục vụ liên tục. Trong các trường hợp này, việc bổ sung các UAV vào hệ thống cho phép mở rộng vùng phủ sóng và tăng tính linh hoạt của hệ thống. UAV có thể hoạt động như các nút lưu trữ và truyền dữ liệu trên không, đồng thời vị trí dừng của chúng có thể được tối ưu nhằm bù đắp các khoảng trống lưu trữ giữa các RSU. Ngoài ra, CAHO cũng phù hợp với các tình huống tạm thời hoặc đột xuất, chẳng hạn như các sự kiện tập trung đông người, các khu vực xảy ra sự cố giao thông hoặc các kịch bản khẩn cấp, nơi việc triển khai nhanh các UAV có thể cải thiện đáng kể khả năng cung cấp dịch vụ.

DPC và CAHO phù hợp với hai ngữ cảnh triển khai khác nhau. DPC khai thác tài nguyên lưu trữ tại phương tiện qua kênh V2V, phù hợp với khu vực đô thị có mật độ phương tiện cao và đủ phương tiện sẵn sàng tham gia lưu trữ. CAHO khai thác UAV như nút lưu trữ trên không qua kênh U2V, phù hợp với khu vực có hạ tầng RSU thưa hoặc phân bố phương tiện không đều. Mô hình DPC và CAHO có thể được xem là hai hướng mở rộng bổ sung của cùng một kiến trúc lưu trữ hợp tác, trong đó, DPC khai thác khả năng lưu trữ của phương tiện mặt đất, còn CAHO tận dụng tính linh hoạt không gian của UAV để nâng cao hiệu quả phân phối video trong mạng VANET.

1.4.3. Các giả định của bài toán

Trong quá trình xây dựng mô hình và thực hiện đánh giá hiệu năng của các phương pháp đề xuất, luận án đưa ra một số giả định nhằm đơn giản hóa việc mô hình hóa hệ thống và tập trung vào các yếu tố chủ yếu của bài toán. Những giả định này chủ yếu nhằm tạo ra một môi trường phân tích rõ ràng, từ đó làm nổi bật hiệu quả của các chiến lược lưu trữ và phân phối nội dung video trong mạng VANET và UAV-VANET.

Trước hết, về mặt hạ tầng mạng, mô hình giả định rằng các RSU và UAV luôn sẵn sàng hoạt động và được triển khai tương đối đồng đều trong khu vực nghiên cứu. Trong phạm vi mô hình hiện tại, các nút hạ tầng này được xem là hoạt động ổn định trong suốt quá trình phục vụ nội dung. Các vấn đề có thể xảy ra trong thực tế như gián

đoạn kết nối, hồng hóc thiết bị, hoặc những hạn chế trong việc triển khai hạ tầng do điều kiện địa hình và chi phí chưa được xét đến. Giả định này giúp đơn giản hóa quá trình mô hình hóa, đồng thời cho phép tập trung vào việc đánh giá hiệu quả của các cơ chế lưu trữ và chiến lược phân phối nội dung.

Bên cạnh đó, luận án cũng chưa xét đến yếu tố năng lượng của hệ thống, đặc biệt là đối với các UAV. Trong nghiên cứu này, hàm mục tiêu được xây dựng chủ yếu nhằm tối thiểu hóa thời gian phục vụ video và tối ưu việc sử dụng tài nguyên lưu trữ trong hệ thống. Tuy nhiên, trong các hệ thống UAV thực tế, năng lượng là một yếu tố quan trọng vì nó ảnh hưởng trực tiếp đến thời gian hoạt động cũng như khả năng duy trì dịch vụ của UAV. Do đó, việc đưa thêm các ràng buộc liên quan đến năng lượng hoặc xem xét các mô hình tối ưu đa mục tiêu kết hợp giữa độ trễ và tiêu thụ năng lượng là một hướng mở rộng cần được nghiên cứu trong các công trình tiếp theo.

Ngoài ra, mô hình truyền thông được sử dụng trong luận án cũng được xây dựng ở mức độ trừu tượng nhất định để thuận tiện cho việc phân tích và mô phỏng. Cụ thể, độ phổ biến của các video được giả định tuân theo phân bố Zipf-like, một mô hình thường được sử dụng để mô tả đặc tính truy cập nội dung trong các hệ thống phân phối video. Tuy nhiên, mô hình này chưa phản ánh đầy đủ sự biến động của nhu cầu nội dung theo thời gian hoặc theo từng khu vực trong môi trường giao thông thực tế. Đồng thời, các yếu tố phức tạp của kênh truyền như nhiễu giao thoa giữa các liên kết, fading đa đường, hiệu ứng shadowing hay sự thay đổi điều kiện kênh theo môi trường truyền dẫn cũng chưa được mô hình hóa đầy đủ.

Trong luận án hiện tại, các RSU được giả định phân bố đều dọc theo tuyến đường, điều này giúp mô hình hệ thống trở nên đơn giản hơn và cho phép tập trung phân tích hiệu quả của các chiến lược lưu trữ và phân phối nội dung. Tuy nhiên, trong thực tế, việc triển khai hạ tầng RSU thường không hoàn toàn đồng đều, mà phụ thuộc vào nhiều yếu tố như mật độ giao thông, điều kiện địa hình, chi phí triển khai và nhu cầu dịch vụ tại từng khu vực. Do đó, khoảng cách giữa các RSU có thể khác nhau, dẫn đến sự tồn tại của các đoạn đường có vùng phủ tốt và các đoạn khác có vùng phủ hạn chế.

Khi các RSU không được phân bố đều, hiệu năng của hệ thống có thể bị ảnh hưởng theo hai hướng. Thứ nhất, tại các khu vực có mật độ RSU cao, phương tiện có

nhiều khả năng truy cập nội dung trực tiếp từ RSU, từ đó giảm thời gian phục vụ video. Ngược lại, tại các đoạn đường có khoảng cách giữa các RSU lớn, khả năng truy cập nội dung từ hạ tầng mặt đất sẽ bị hạn chế, làm tăng độ trễ khi truyền dữ liệu. Trong trường hợp này, vai trò của các cơ chế lưu trữ hợp tác và các nút trung gian như phương tiện hoặc UAV trở nên quan trọng hơn, vì chúng có thể giúp bù đắp các khoảng trống về vùng phủ giữa các RSU.

Đối với mô hình DPC, khi RSU không phân bố đều, chiến lược lưu trữ nội dung có thể cần được điều chỉnh để ưu tiên lưu trữ nhiều nội dung hơn tại các RSU nằm ở những khu vực có khoảng cách lớn hoặc mật độ xe cao. Điều này giúp tăng khả năng phục vụ nội dung tại các điểm nút quan trọng trong mạng. Trong khi đó, đối với mô hình CAHO, sự xuất hiện của UAV mang lại lợi thế đáng kể trong các kịch bản hạ tầng không đồng đều. Các UAV có thể được triển khai linh hoạt tại những khu vực có vùng phủ yếu nhằm hỗ trợ phân phối nội dung và giảm thời gian phục vụ video cho các phương tiện trong khu vực đó.

Do đó, mặc dù luận án giả định các RSU được phân bố đều để thuận tiện cho việc phân tích và mô phỏng, các mô hình đề xuất vẫn có khả năng thích nghi với các kịch bản triển khai hạ tầng không đồng đều. Trong các nghiên cứu tiếp theo, việc mở rộng mô hình sang các kịch bản RSU phân bố không đồng đều hoặc mật độ RSU thay đổi theo không gian sẽ giúp đánh giá toàn diện hơn hiệu năng của các chiến lược lưu trữ và phân phối nội dung trong các hệ thống giao thông thông minh thực tế.

Các giả định trên giúp xây dựng một mô hình đủ đơn giản để có thể phân tích và đánh giá hiệu quả của các phương pháp đề xuất. Đồng thời, chúng cũng cho thấy vẫn còn nhiều yếu tố của hệ thống thực tế chưa được xem xét đầy đủ, qua đó mở ra những hướng nghiên cứu tiếp theo nhằm hoàn thiện mô hình và tiến gần hơn tới điều kiện triển khai thực tế của các hệ thống VANET và UAV-VANET.

1.4.4. Lộ trình nghiên cứu

Các đóng góp của luận án được xây dựng theo một lộ trình nghiên cứu nhất quán và tăng tiến, trong đó mỗi nghiên cứu kế tiếp kế thừa trực tiếp những kết quả và hạn chế của nghiên cứu trước, đồng thời mở rộng không gian mô hình và mức độ tối ưu, nhằm

từng bước hoàn thiện khung tối ưu cho dịch vụ truyền video trong hệ thống UAV-VANET.

Nghiên cứu thứ nhất: Lưu trữ tại các RSU - nền tảng của bài toán tối ưu thời gian phục vụ.

Luận án bắt đầu từ kịch bản cơ bản, trong đó nội dung video được lưu trữ tại các RSU. Thời gian phục vụ trung bình (Average Service Time) được lựa chọn làm mục tiêu đánh giá, phản ánh trực tiếp chất lượng dịch vụ video trong môi trường VANET.

Nghiên cứu này chứng minh rằng việc đưa lưu trữ nội dung xuống RSU có thể giảm đáng kể thời gian phục vụ video so với mô hình truyền thống phụ thuộc hoàn toàn vào máy chủ trung tâm. Kết quả này đặt nền móng cho việc xem thời gian phục vụ như là tối ưu quan trọng trong các nghiên cứu tiếp theo.

Nghiên cứu thứ hai: Mở rộng sang lưu trữ kết hợp RSU và phương tiện – tăng cường khả năng thích ứng trong môi trường động.

Dựa trên những kết quả và đồng thời nhận diện hạn chế của mô hình lưu trữ đơn tầng tại RSU, nghiên cứu thứ hai mở rộng không gian lưu trữ sang các phương tiện giao thông. Việc chuyển từ mô hình chỉ triển khai tại RSU sang mô hình RSU và phương tiện cho phép khai thác tài nguyên lưu trữ phân tán và tính linh hoạt vốn có của mạng VANET.

Mối liên hệ và tăng tiến: Trong khi nghiên cứu thứ nhất tập trung vào cải thiện thời gian phục vụ thông qua hạ tầng cố định, nghiên cứu thứ hai kế thừa tiêu chí tối ưu thời gian phục vụ nhưng mở rộng không gian quyết định lưu trữ, từ đó phản ánh sát hơn đặc tính động của mạng VANET. Nghiên cứu này chứng minh rằng mô hình lưu trữ kết hợp RSU và phương tiện vượt trội hơn so với lưu trữ chỉ tại RSU, đặc biệt trong các kịch bản lưu lượng cao và kết nối không ổn định.

Nghiên cứu thứ ba: Tích hợp UAV - chuyển dịch từ không gian hai chiều sang ba chiều.

Mặc dù mô hình RSU và phương tiện mang lại cải thiện đáng kể, nó vẫn chịu hạn chế trong các kịch bản mật độ RSU thưa hoặc hạ tầng mặt đất không đồng đều.

Trên cơ sở đó, nghiên cứu thứ ba tiếp tục mở rộng mô hình sang không gian ba chiều, bằng cách tích hợp lưu trữ nội dung tại các UAV.

Mối liên hệ và tăng tiến: Nghiên cứu này kế thừa khung tối ưu thời gian phục vụ và mô hình lưu trữ đa tầng từ nghiên cứu thứ hai, đồng thời mở rộng không gian lưu trữ từ hai chiều (RSU và phương tiện) sang ba chiều (RSU và UAV). Kết quả cho thấy lưu trữ hỗ trợ bởi UAV có khả năng bù đắp hiệu quả các hạn chế về hạ tầng trong kịch bản RSU thưa, đồng thời cải thiện thời gian phục vụ và vùng phủ cho dịch vụ video.

Ba nghiên cứu không tồn tại độc lập mà tạo thành một chuỗi phát triển logic, trong đó mỗi bước đều mở rộng không gian mô hình và nâng cao mức độ tối ưu: từ lưu trữ đơn tầng tại RSU, sang lưu trữ đa tầng RSU kết hợp với phương tiện, và cuối cùng là lưu trữ ba chiều với sự hỗ trợ của UAV. Chuỗi nghiên cứu này hình thành một khung tối ưu toàn diện cho dịch vụ truyền video trong UAV-VANET, đặt trọng tâm vào việc giảm thời gian phục vụ và nâng cao hiệu quả sử dụng tài nguyên mạng trong các hệ thống giao thông thông minh thế hệ mới.

1.5. CÁC KHÁI NIỆM VÀ MÔ HÌNH NỀN TẢNG

1.5.1. Thời gian phục vụ

Khác với nhiều nghiên cứu trước đây thường sử dụng các tiêu chí trung gian như tỉ lệ truy suất nội dung lưu trữ thành công hoặc thông lượng hệ thống để đánh giá hiệu năng, luận án lựa chọn thời gian phục vụ làm tiêu chí chính. Lý do của lựa chọn này xuất phát từ việc các tiêu chí truyền thống chỉ phản ánh một khía cạnh cục bộ của hệ thống, nhưng chưa phản ánh trực tiếp chất lượng dịch vụ mà người dùng thực sự nhận được. Một hệ thống có tỉ lệ truy suất nội dung lưu trữ thành công cao không đảm bảo người dùng nhận được nội dung thành công trong thời gian cho phép, bởi dữ liệu có thể vẫn phải truyền qua các liên kết có băng thông giới hạn hoặc chịu ảnh hưởng của điều kiện kênh truyền. Tương tự, thông lượng hệ thống lớn cũng không đồng nghĩa với việc mỗi yêu cầu dịch vụ được đáp ứng nhanh, vì thông lượng chỉ phản ánh tổng lượng dữ liệu truyền trong mạng mà không trực tiếp biểu diễn độ trễ phục vụ cho từng yêu cầu cụ thể. Do đó, nếu chỉ tối ưu các chỉ tiêu này, hệ thống có thể đạt hiệu năng tốt ở mức tổng thể nhưng vẫn chưa đảm bảo trải nghiệm dịch vụ tốt cho từng người dùng.

Thời gian phục vụ được định nghĩa là khoảng thời gian cần thiết để một phương tiện nhận được đầy đủ nội dung video kể từ thời điểm phát sinh yêu cầu. Đại lượng này phản ánh trực tiếp hiệu quả cung cấp dịch vụ của hệ thống, bởi nó bao gồm đồng thời nhiều yếu tố quan trọng như độ trễ truy cập nội dung, thời gian truyền dữ liệu qua các liên kết không dây và khả năng khai thác các nguồn nội dung sẵn cũng như tài nguyên trong mạng.

Từ góc độ hệ thống, việc giảm thời gian phục vụ đồng nghĩa với việc cải thiện các chỉ tiêu chất lượng dịch vụ (QoS) như độ trễ truyền dữ liệu, khả năng đáp ứng yêu cầu trong thời gian thực và hiệu quả sử dụng tài nguyên mạng. Đồng thời, đối với người dùng cuối, thời gian phục vụ cũng ảnh hưởng trực tiếp đến chất lượng trải nghiệm (QoE), vì thời gian chờ đợi ngắn hơn giúp quá trình phát video diễn ra mượt mà hơn, giảm hiện tượng gián đoạn hoặc trễ khi truy cập nội dung. Do đó, tối ưu hóa thời gian phục vụ không chỉ cải thiện hiệu năng kỹ thuật của hệ thống mà còn nâng cao trải nghiệm dịch vụ của người dùng, đặc biệt là trong các dịch vụ và ứng dụng truyền video vốn rất nhạy cảm với trì hoãn.

Nhờ đặc điểm này, thời gian phục vụ có thể được xem như một đại lượng tổng hợp, kết nối giữa các yếu tố kỹ thuật của mạng và cảm nhận thực tế của người sử dụng. Việc lựa chọn thời gian phục vụ làm tiêu chí tối ưu sẽ cho phép xây dựng một khung mô hình thống nhất, trong đó, các quyết định về lưu trữ nội dung, cấu trúc mạng và cơ chế truyền dữ liệu đều hướng tới mục tiêu cuối cùng là cung cấp nội dung video nhanh hơn và ổn định hơn cho các phương tiện trong mạng VANET và UAV-VANET.

Trong các mô hình tối ưu được phát triển xuyên suốt nghiên cứu, thời gian t_i được sử dụng như đại lượng cơ bản phản ánh hiệu năng phục vụ của hệ thống đối với nội dung video i . t_i biểu diễn thời gian cần thiết để một phương tiện hoàn tất việc nhận nội dung video i thông qua hạ tầng lưu trữ hiện tại của hệ thống.

Giá trị t_i phụ thuộc trực tiếp vào:

- Cấu hình lưu trữ nội dung (tại RSU, phương tiện hoặc UAV);
- Khoảng cách giữa phương tiện và nút lưu trữ gần nhất;

- Tốc độ truyền dữ liệu của liên kết tương ứng;
- Và trong các mô hình mở rộng, còn phụ thuộc vào vị trí hoặc quỹ đạo của UAV trong không gian ba chiều.

Trong nghiên cứu thứ nhất (chỉ lưu trữ tại RSU), t_i chủ yếu bị chi phối bởi vị trí các RSU lưu trữ nội dung i và khoảng cách truyền trên mặt đất. Khi mở rộng sang mô hình lưu trữ đa tầng (RSU và phương tiện), t_i được cải thiện nhờ khả năng truy xuất nội dung từ các nút lưu trữ gần hơn trong mạng động. Ở mô hình tích hợp UAV, t_i còn chịu ảnh hưởng của cấu trúc không gian ba chiều và vị trí triển khai UAV, cho phép rút ngắn khoảng cách truyền và tăng xác suất liên kết LoS.

Do đó, thay vì tối ưu trực tiếp các tham số truyền thông riêng lẻ, cả ba nghiên cứu đều sử dụng t_i làm đơn vị đánh giá, và xây dựng hàm mục tiêu toàn cục dựa trên tập các giá trị t_i , thường dưới dạng tổng có trọng số theo phân bố phổ biến nội dung. Cách tiếp cận này cho phép liên kết trực tiếp cấu hình lưu trữ và vị trí hệ thống với chỉ tiêu hiệu năng thực tế của dịch vụ truyền video.

1.5.2. Độ nổi tiếng của video

Do nhu cầu truy cập nội dung trong các hệ thống video thực tế thường không phân bố đồng đều, trong triển khai thực tế, độ phổ biến nội dung có thể được ước lượng từ lịch sử truy cập tại RSU/MEC, yêu cầu video, dữ liệu nội dung hoặc thông kê truy cập trong các cửa sổ thời gian gần nhất. Vì vậy, độ phổ biến p_i trong mô hình được xem là tham số đầu vào có thể được cập nhật định kỳ hoặc dự đoán. Trong luận án, độ phổ biến của video i được mô hình hoá bằng phân bố Zipf-like [46], cho bởi:

$$p_i = \frac{i^{-\alpha}}{\sum_{l=1}^I l^{-\alpha}}, \quad (1.1)$$

Trong đó, I là tổng số video trong hệ thống. $\alpha \geq 0$ là thông số đặc trưng cho mức chênh lệch về phân bố độ phổ biến giữa các video. Khi $\alpha = 0$, tất cả các video có cùng độ phổ biến $p_i = 1/I$. Khi α tăng, mức độ phổ biến tập trung nhiều hơn vào một số video nhất định.

1.5.3. Mô hình đường đi

Trong luận án, NCS áp dụng mô hình truyền R2V được trình bày trong [34]. Trong mô hình này, phạm vi phủ sóng d của một RSU được chia thành K vùng. Các tham số của vùng k , với $k = 1, 2, \dots, K$ và $K = 7$, được liệt kê trong bảng 1.1.

Bảng 1.1: Vùng phủ của một RSU

Vùng	1	2	3	4	5	6	7
d_k (m)	25	30	40	60	40	30	25
R_k (Mbps)	1	2	5.5	11	5.5	2	1

Như vậy mỗi RSU có khả năng phủ sóng một đoạn đường dài khoảng 250 m, hệ thống gồm 25 RSU trong luận án tương ứng với một đoạn đường khảo sát có chiều dài xấp xỉ 6.25 km. Đây là một quy mô đủ lớn để phản ánh đặc trưng của các trục giao thông đô thị chính hoặc các tuyến đường vành đai trong thành phố, nơi mật độ phương tiện và nhu cầu truy cập dịch vụ video thường ở mức cao. Việc lựa chọn chiều dài tuyến như vậy giúp đánh giá hiệu năng hệ thống không chỉ ở mức cục bộ mà còn trên một không gian đủ rộng để quan sát tương tác giữa các vùng phủ RSU lân cận và cơ chế phối hợp lưu trữ.

Bên cạnh đó, tốc độ phương tiện được thiết lập ngẫu nhiên trong khoảng từ 30 km/h đến 60 km/h, tương ứng với điều kiện lưu thông phổ biến trong môi trường đô thị. Khoảng tốc độ này đảm bảo phản ánh đúng đặc tính di chuyển thực tế của phương tiện trong thành phố (không quá thấp như tình huống ùn tắc kéo dài), cũng không quá cao như cao tốc liên tỉnh, qua đó tạo điều kiện kiểm chứng hiệu quả của cơ chế lưu trữ và phân phối video trong bối cảnh người dùng có nhu cầu giải trí hoặc truy cập nội dung khi đang di chuyển.

Việc kết hợp chiều dài tuyến khảo sát đủ lớn và dải tốc độ phù hợp với môi trường thành thị cho thấy các kịch bản mô phỏng trong luận án không mang tính lý tưởng hóa quá mức, mà được thiết kế nhằm tiệm cận điều kiện vận hành thực tế. Điều

này góp phần củng cố khả năng ứng dụng của mô hình trong các hệ thống giao thông thông minh và mạng UAV-VANET triển khai trong đô thị.

1.5.4. Lưu trữ tại các RSU

Trong mạng VANET, ngoài chức năng là điểm truy cập hạ tầng cho phương tiện, các RSU còn được sử dụng như nút lưu trữ nội dung tại biên mạng. Lưu trữ video phổ biến ngay tại RSU giảm độ trễ truy cập trong các ứng dụng truyền video so với việc chuyển tiếp yêu cầu về máy chủ ở mạng lõi, vốn phát sinh thêm độ trễ truyền dẫn và tải mạng. Vì vậy, việc đặt nội dung video ngay tại các RSU giúp rút ngắn đường đi của dữ liệu, từ đó giảm đáng kể thời gian cần thiết để phục vụ yêu cầu của phương tiện. Chi tiết về nghiên cứu này được trình bày trong [C1].

Từ góc độ mô hình hóa, thời gian phục vụ video trong mạng VANET có thể được xem là sự kết hợp của hai thành phần chính: thời gian di chuyển của phương tiện để tiếp cận RSU có lưu trữ nội dung và thời gian truyền dữ liệu từ RSU tới phương tiện thông qua liên kết R2V. Khi nội dung được lưu trữ hợp lý tại các RSU dọc tuyến đường, khoảng cách trung bình từ phương tiện đến RSU gần nhất chứa video giảm xuống, kéo theo thành phần thời gian di chuyển trong tổng thời gian phục vụ giảm tương ứng.

Mặt khác, việc lưu trữ nội dung tại RSU cho phép khai thác đặc tính phân bố truy cập nội dung của người dùng. Trong thực tế, nhu cầu truy cập video thường không phân bố đồng đều mà tuân theo quy luật Zipf-like, nghĩa là một số lượng nhỏ các video phổ biến chiếm phần lớn số lần truy cập. Khi các video có độ phổ biến cao được lưu trữ tại RSU, các yêu cầu từ phương tiện có thể được phục vụ ngay tại tầng hạ tầng biên mà không cần truy cập tới mạng lõi. Điều này không chỉ làm giảm thời gian phục vụ trung bình mà còn giúp giảm tải cho hệ thống mạng phía sau, từ đó nâng cao hiệu năng tổng thể của hệ thống.

Ngoài ra, việc bố trí nội dung video tại các RSU theo một bán kính lưu trữ tối ưu còn giúp cân bằng giữa hiệu quả phục vụ và dung lượng lưu trữ của hệ thống. Nếu mỗi video được lưu tại tất cả các RSU, hệ thống có thể đạt thời gian phục vụ rất nhỏ nhưng lại tiêu tốn dung lượng lưu trữ lớn. Ngược lại, nếu nội dung được lưu quá thưa, phương tiện có thể phải di chuyển qua nhiều đoạn đường mới tiếp cận được RSU có chứa video,

làm tăng đáng kể thời gian phục vụ. Do đó, bài toán tối ưu vị trí lưu trữ nội dung tại các RSU thực chất là bài toán tìm ra cấu hình phân bố nội dung phù hợp để giảm thiểu thời gian phục vụ trong điều kiện tài nguyên lưu trữ có hạn.

Từ các phân tích trên có thể thấy rằng lưu trữ nội dung tại RSU đóng vai trò nền tảng trong việc giảm thời gian phục vụ video trong mạng VANET. Việc khai thác hợp lý tài nguyên lưu trữ tại các RSU không chỉ giúp rút ngắn khoảng cách truyền dữ liệu, mà còn tận dụng được đặc tính phân bố truy cập nội dung của người dùng. Đây cũng chính là cơ sở để các mô hình tối ưu trong các nghiên cứu tiếp theo, như DPC và CAHO, tiếp tục mở rộng bằng cách kết hợp thêm các tầng lưu trữ khác trong mạng nhằm cải thiện hơn nữa hiệu năng cung cấp dịch vụ.

Đối với lưu trữ tại RSU, bán kính lưu trữ h_i của video i được xác định rằng video này sẽ được lưu trữ sau mỗi h_i RSU. Khi đó, các RSU lưu trữ video i được xác định theo chỉ số:

$$n_{i,j} = h_i(j - 1) + 1, j = 1, 2, \dots, J_i,$$

trong đó J_i là số RSU lưu trữ video i , được tính bởi:

$$J_i = \left\lfloor \frac{N-1}{h_i} \right\rfloor + 1, \quad (1.2)$$

với:

N : tổng số RSU được triển khai dọc theo tuyến đường;

h_i : bán kính lưu trữ của video i , được biểu diễn bằng một số nguyên dương. Trong mô hình, h_i là số RSU giữa hai RSU liên tiếp cùng lưu trữ video i ;

J_i : tổng số RSU lưu trữ video i ;

$\lfloor x \rfloor$ hàm lấy phần nguyên dưới, tức là số nguyên lớn nhất không vượt quá x .

1.6. KẾT LUẬN CHƯƠNG 1

Chương 1 đã trình bày tổng quan về mạng UAV-VANET, bao gồm kiến trúc hệ thống, các ứng dụng tiêu biểu và những thách thức kỹ thuật liên quan đến truyền thông, lưu trữ và phân phối nội dung trong môi trường giao thông động. Trên cơ sở đó, chương này khảo sát các nhóm nghiên cứu liên quan đến truyền video, lưu trữ tại biên, lưu trữ hợp tác, lưu trữ xác suất và các mô hình VANET có UAV hỗ trợ.

Từ quá trình tổng quan, luận án xác định ba khoảng trống nghiên cứu chính: thiếu mô hình tối ưu tích hợp giữa các tầng lưu trữ RSU–phương tiện–UAV; chưa khai thác đầy đủ vai trò của UAV như một nút lưu trữ và phân phối nội dung linh hoạt; và chưa đặt thời gian phục vụ video làm hàm mục tiêu tối ưu trực tiếp. Đây là cơ sở để luận án xây dựng các mô hình tối ưu nhằm giảm thời gian phục vụ và nâng cao hiệu quả phân phối video trong VANET/UAV-VANET. Chương 2 trước hết xây dựng mô hình DPC nhằm tối ưu kết hợp lưu trữ xác định tại RSU và lưu trữ xác suất tại phương tiện trong VANET. Trên cơ sở kết quả và những giới hạn của mô hình mặt đất, Chương 3 tiếp tục mở rộng nghiên cứu sang UAV-VANET thông qua mô hình CAHO, tối ưu đồng thời lưu trữ nội dung và vị trí dừng của UAV nhằm giảm thời gian phục vụ video.

Ngoài ra, chương này cũng giới thiệu các mô hình nền tảng phục vụ cho các chương tiếp theo, gồm độ phổ biến nội dung, phân bố phương tiện, cơ chế lưu trữ tại RSU và tiêu chí thời gian phục vụ. Các nội dung này là cơ sở cho việc đề xuất mô hình DPC trong Chương 2 và mở rộng sang mô hình CAHO trong Chương 3.

CHƯƠNG 2: KẾT HỢP LƯU TRỮ XÁC ĐỊNH VÀ LƯU TRỮ XÁC SUẤT TRONG VANET

Tóm tắt: Trong Chương 2, NCS tập trung nghiên cứu bài toán kết hợp lưu trữ theo vị trí xác định và lưu trữ xác suất (DPC – Deterministic and Probabilistic Caching) cho dịch vụ truyền video hợp tác trong mạng VANET (Vehicular Ad-hoc Network). Xuất phát từ thực tế nhu cầu giải trí trên xe ngày càng gia tăng, trong khi tài nguyên lưu trữ và truyền dẫn tại các nút mạng là hữu hạn, chương này đặt mục tiêu xây dựng một mô hình lưu trữ có khả năng giảm thời gian phục vụ thông qua việc khai thác hiệu quả cả hạ tầng RSU (Roadside Unit) và các phương tiện giao thông.

Khác với các phương pháp lưu trữ truyền thống chỉ xem xét lưu trữ độc lập tại từng nút, mô hình DPC đề xuất kết hợp hai chiến lược lưu trữ bổ trợ lẫn nhau. Cụ thể, lưu trữ xác định (DC – Deterministic Caching) được áp dụng tại các RSU, trong đó, video được phân bổ theo bán kính lưu trữ nhằm đảm bảo khả năng truy cập nội dung ổn định dọc theo tuyến đường. Song song đó, lưu trữ xác suất (PC – Probabilistic Caching) được triển khai tại các phương tiện, cho phép các xe lưu trữ video với xác suất phụ thuộc vào độ phổ biến nội dung, từ đó tăng khả năng chia sẻ video thông qua truyền dẫn hợp tác giữa RSU và các phương tiện trong quá trình di chuyển.

Trên cơ sở mô hình hóa sự xuất hiện của phương tiện bằng quá trình Poisson một chiều, kết hợp với các mô hình truyền RSU-to-Vehicle (R2V) và Vehicle-to-Vehicle (V2V), bài toán tối ưu DPC được xây dựng nhằm cực tiểu hóa thời gian phục vụ trung bình của các video được yêu cầu. Bài toán này đồng thời chịu các ràng buộc nghiêm ngặt về dung lượng lưu trữ tại RSU và phương tiện, cũng như đặc tính động của mạng VANET.

Do bài toán DPC mang tính phi tuyến, rời rạc và có không gian tìm kiếm lớn, NCS áp dụng GA kết hợp với hàm phạt để xử lý các ràng buộc về tài nguyên lưu trữ. Cách tiếp cận này cho phép tối ưu đồng thời các biến nguyên và liên tục, đồng thời đảm bảo khả năng hội tụ về nghiệm gần tối ưu trong thời gian tính toán chấp nhận được.

Các kết quả mô phỏng cho thấy mô hình DPC giúp giảm đáng kể thời gian phục vụ so với các phương pháp lưu trữ so sánh khác, đồng thời sử dụng tài nguyên lưu trữ

một cách hiệu quả và cân bằng giữa RSU và phương tiện. Những kết quả này không chỉ khẳng định tính hiệu quả của chiến lược lưu trữ kết hợp, mà còn đóng vai trò là nền tảng quan trọng cho việc mở rộng trong các nghiên cứu tiếp theo. Kết quả nghiên cứu này được công bố trong bài báo [C2].

2.1. GIỚI THIỆU

Mạng phương tiện giao thông tự tổ chức (VANET – Vehicular Ad-hoc Network) đã trở thành thành phần không thể thiếu trong các hệ thống giao thông thông minh hiện đại, cho phép triển khai đa dạng các ứng dụng và dịch vụ truyền thông giữa người dùng trên xe (VU – Vehicle User) và hạ tầng ven đường trong mạng 6G [47]. Trên cơ sở đó, các nhà sản xuất ô tô đã phát triển nhiều loại ứng dụng và dịch vụ trên các phương tiện, như hệ thống giải trí, lái xe tự động, cập nhật phần mềm qua mạng, và các hệ thống hỗ trợ lái xe nâng cao [48]. Vai trò ngày càng quan trọng của các ứng dụng và dịch vụ trên xe đã thúc đẩy quá trình phân phối dữ liệu mạnh mẽ trong VANET, kéo theo sự gia tăng đáng kể tải truyền trên mạng lõi của các trạm gốc MBS (Macro Base Station), các trạm gốc tế bào nhỏ SBS (Small Base Station), cũng như các trạm phát ven đường (RSU – Roadside Unit) [34], [49], [50], [51], [52]. Trong môi trường VANET, thách thức này càng trở nên nghiêm trọng hơn do đặc tính di động của phương tiện, hiện tượng gián đoạn kết nối xảy ra thường xuyên, và đặc biệt là nhu cầu rất lớn về tốc độ dữ liệu cao đối với các dịch vụ nhạy cảm về độ trễ như dịch vụ truyền video (VAS – Video Applications and Services). Nhằm giải quyết các thách thức nêu trên, thay vì triển khai cơ sở hạ tầng mạng lõi tốc độ cao với chi phí lớn, các kỹ thuật biên dựa trên phần mềm có thể được áp dụng như một giải pháp thay thế hiệu quả [53].

Một trong những kỹ thuật biên dựa trên phần mềm đầy hứa hẹn nhất, thu hút được sự quan tâm đáng kể trong nghiên cứu VANET thế hệ 6G, là kỹ thuật lưu trữ [54], [55], [56]. Trong các mạng 6G, sự phát triển của dữ liệu lớn, học máy và tính toán hiệu năng cao cho phép cải thiện độ chính xác của các mô hình dự đoán độ phổ biến nội dung và mô hình di động phương tiện, là hai yếu tố đầu vào quan trọng cho các bài toán lưu trữ trong VANET. Một lợi ích quan trọng của kỹ thuật lưu trữ là khả năng sử dụng tài nguyên một cách hiệu quả, đồng thời giúp giảm chi phí đầu tư hạ tầng cho các nhà cung cấp dịch vụ Internet và các nhà cung cấp nội dung, trong khi vẫn cải thiện chất

lượng trải nghiệm cho người dùng di động. Theo hướng này, nhiều phương pháp lưu trữ mới đã được đề xuất cho cả các mạng mặt đất và các mạng không gian động, bao gồm từ VANET đến các hệ thống phân phối nội dung, truyền thông thiết bị–thiết bị (D2D – Device to Device), Internet vạn vật (IoT – Internet of Things), phương tiện bay không người lái (UAV – Unmanned Aerial Vehicle), và các mạng vệ tinh [53], [57], [58], [59], [60], [61].

Thực tế, lưu trữ trong VANET khác biệt so với các loại mạng khác do sự đa dạng của các yếu tố liên quan đến ứng dụng và dịch vụ, các đặc tính phức tạp của VU, môi trường có tính động cao, cũng như sự thiếu phổ cập của hạ tầng ven đường [52], [62]. Do đó, nhằm nâng cao hiệu năng của VANET, việc lưu trữ cần xem xét nhiều loại nội dung, tác vụ và dịch vụ (CTS – Content, Tasks, and Services), trong đó mỗi loại đều gắn liền với các mô hình phổ biến riêng biệt. Bên cạnh đó, kỹ thuật lưu trữ còn cần tính đến các yếu tố như tính di động, sự phân bố, các mối quan hệ xã hội, cơ chế khuyến khích và nguồn tài nguyên sẵn có của các VU. Các phương pháp lưu trữ được phân thành hai nhóm chính là lưu trữ xác định (DC – Deterministic Caching) và lưu trữ xác suất (PC – Probabilistic Caching). Các phương pháp này có thể được kết hợp với nhau và được hỗ trợ bởi các kỹ thuật khác, ví dụ như: phân cụm (clustering) [63], [64], truyền hợp tác (CoT – Cooperative Transmission) [34], [50], [51], [65], [66], [67], cũng như lập lịch truyền quảng bá (TrS – Transmission Scheduling) và thích nghi tốc độ truyền (TrA – Transmission Adaptation) [27] nhằm thích ứng với môi trường có tính động cao và hạ tầng ven đường không phổ cập.

Thông qua việc lưu trữ, các CTS được truy cập thường xuyên sẽ được lưu tại biên của VANET, gần với người dùng cuối, nhằm tạo điều kiện thuận lợi cho việc cung cấp các ứng dụng và dịch vụ cho phương tiện. Cách tiếp cận này giúp giảm tải cho mạng lõi thông qua cơ chế chuyển tải (offloading), cải thiện chất lượng dịch vụ (QoS – Quality of Service) bằng cách rút ngắn khoảng cách truyền dẫn, đồng thời cung cấp khả năng truy cập cao hơn tới nhiều CTS đã được lưu trữ. Bên cạnh đó, phương pháp này còn cho phép sử dụng hiệu quả các tài nguyên năng lượng, phổ tần và lưu trữ. Trong các phương pháp DC, các CTS được đặt tại các RSU cố định [64] nhằm tận dụng truyền thông từ hạ tầng đến phương tiện cũng như các tài nguyên lưu trữ đáng tin cậy dọc theo

tuyến đường [27], [34], [50], [51], [64], [65], [68], [69], [70]. Cách tiếp cận này cho phép việc tối ưu hóa vị trí lưu trữ nội dung được thực hiện theo kế hoạch từ trước và một cách có hệ thống. Tuy nhiên, hạ tầng ven đường còn hạn chế (chủ yếu do chi phí triển khai cao [52]), ngăn cản cung cấp các ứng dụng và dịch vụ liên tục, đặc biệt là các dịch vụ video yêu cầu tốc độ dữ liệu cao và độ trễ thấp được người dùng cuối yêu cầu trong các VANET có mật độ lớn. Ngược lại, phương pháp PC cho phép lưu trữ nội dung tại các phương tiện di động [67], [71], [72], [73], [74], [75]. Trong bối cảnh này, do các VU có tính di động và chỉ kết nối gián đoạn trong khi cùng di chuyển trong một khoảng thời gian nhất định, chiến lược PC được sử dụng rộng rãi nhằm thích ứng với các mô hình di động của VU. Quan trọng hơn, phương pháp PC có thể bù đắp hiệu quả cho những hạn chế của phương pháp DC khi mật độ VU cao tạo ra nhiều cơ hội lưu trữ hơn, từ đó cho phép cung cấp các ứng dụng và dịch vụ liên mạch hơn, ngay cả khi các VU tham gia hoặc rời khỏi mạng một cách ngẫu nhiên [58].

Để làm rõ các ưu điểm và hạn chế của các phương pháp DC và PC, một bảng tổng hợp so sánh các đặc trưng chính của hai phương pháp này được trình bày trong bảng 2.1. Cụ thể, có thể khẳng định rằng các phương pháp hiện có đều khai thác tính di động của các VU và mức độ phổ biến của các CTS để tối ưu hóa việc lưu trữ tại các trạm gốc MBS, các RSU và/hoặc các VU. Một số phương pháp còn xem xét kết hợp thêm các kỹ thuật khác (như phân cụm, CoT, TrS và TrA) nhằm cải thiện hiệu năng lưu trữ. Tuy nhiên, các nghiên cứu này dựa trên giả định chỉ có một số lượng nhỏ các RSU, điều này là không thực tế. Hơn nữa, mặc dù đã có các thuật toán và lời giải tối ưu được đề xuất, các đặc tính liên quan đến khả năng hội tụ, độ ổn định và độ chính xác của chúng vẫn chưa được nghiên cứu một cách đầy đủ. Quan trọng hơn, các phương pháp lưu trữ thường được triển khai một cách tách biệt, điều này ngăn cản việc khai thác đầy đủ các lợi ích kết hợp của lưu trữ xác định và lưu trữ xác suất nhằm nâng cao hiệu năng của các ứng dụng và dịch vụ trong VANET.

Bảng 2.1: Đặc điểm của kỹ thuật lưu trữ DC và PC

Đặc điểm TLTK	DC	PC	Nơi lưu trữ	Tính linh động	Độ phổ biến	Số lượng RSU	Độ hội tụ	Độ ổn định	Độ chính xác	Kỹ thuật hỗ trợ
[68]	Có	Không	RSU	Có	Có	3	N/A	N/A	Khá tốt	Không
[70]	Có	Không	RSU	Có	Có	6	N/A	N/A	N/A	Không
[63]	Có	Không	RSU	Có	Có	10	Khá tốt	Khá	Khá tốt	Phân cụm
[64]	Có	Không	RSU	Có	Có	30	N/A	N/A	N/A	CoT
[65]	Có	Không	RSU	Có	Có	10	N/A	N/A	N/A	TrS, TrA
[27]	Có	Không	RSU	Có	Có	7	N/A	N/A	N/A	CoT
[34]	Có	Không	RSU, MBS	Có	Có	2	Tốt	N/A	N/A	CoT
[50], [51]	Có	Không	RSU, MBS	Có	Có	30, 2	Khá tốt	N/A	Khá tốt	Không
[69]	Có	Không	RSU, MBS	Có	Có	10	Khá tốt	N/A	Khá tốt	CoT
[66]	Có	Không	RSU, VU	Có	Có	4	N/A	N/A	N/A	CoT
[76]	Có	Không	RSU, VU	Không	Có	4	Khá tốt	N/A	N/A	Phân cụm
[77]	Có	Không	RSU, VU	Có	Có	2	N/A	N/A	Khá tốt	Không
[78]	Có	Không	RSU, VU	Có	Có	6	N/A	N/A	N/A	Không
[79]	Có	Không	VU	Có	Có	0	N/A	N/A	N/A	Không
[80]	Có	Không	VU	Có	Có	0	Khá tốt	N/A	Khá tốt	Không
[81]	Có	Không	VU	Có	Có	0	Khá tốt	N/A	Khá tốt	Không
[71]	Không	Có	VU	Có	Có	0	N/A	N/A	N/A	Không
[72]	Không	Có	VU	Có	Có	0	N/A	N/A	N/A	Không
[67]	Không	Có	VU	Có	Có	0	N/A	N/A	N/A	CoT
[75]	Không	Có	VU	Có	Có	30	N/A	N/A	N/A	Không
[74]	Không	Có	VU	Có	Có	0	N/A	N/A	N/A	Không
[73]	Không	Có	VU	Có	Không	0	N/A	N/A	N/A	Không
DPC	Có	Có	RSU, VU	Có	Có	25	Tốt	Tốt	Tốt	CoT

Xuất phát từ những phân tích nêu trên, trong chương này, NCS đề xuất phương pháp lưu trữ biên kết hợp giữa DC và PC có tên gọi là DPC (Deterministic and Probabilistic Caching), nhằm khai thác toàn diện lưu trữ xác định tại các RSU và lưu trữ xác suất tại các VU cho truyền video hợp tác trong VANET. Mục tiêu của phương pháp là tối thiểu hóa thời gian phục vụ và sử dụng hiệu quả tài nguyên lưu trữ. Mô hình DPC khai thác tài nguyên lưu trữ tại cả RSU và VU để hỗ trợ truyền video hợp tác thông qua các liên kết truyền từ RSU đến VU (R2V) và từ VU đến VU (V2V), qua đó giảm thời gian phục vụ cho các VU truy cập đồng thời các dịch vụ video trong khi đang di chuyển. Trong mô hình DPC, các phương pháp DC và PC lần lượt được áp dụng cho các RSU cố định và các VU di động. Bài toán tối ưu DPC được xây dựng để xác định:

- (i) vị trí lưu trữ nội dung video tại các RSU và
- (ii) xác suất lưu trữ nội dung video tại các VU.

Để đảm bảo tính thực tiễn, nghiên cứu xem xét nhiều yếu tố của các dịch vụ video trong VANET, bao gồm tính di động của VU, mức độ phổ biến của video, số lượng lớn các RSU, và kỹ thuật truyền hợp tác CoT. Ngoài ra, do tính phức tạp của các biến tối ưu, GA được điều chỉnh để giải bài toán tối ưu DPC với hai loại biến khác nhau, bao gồm biến nguyên cho lưu trữ xác định và biến thực cho lưu trữ xác suất.

2.2. CÁC NGHIÊN CỨU LIÊN QUAN

2.2.1. Các ứng dụng và dịch vụ được hỗ trợ bởi lưu trữ trong VANET

Khi các tài nguyên trong VANET được khai thác một cách đầy đủ, mô hình phương tiện-như là-một dịch vụ (Vehicle-as-a-Service) có thể thiết lập như một nền tảng cung cấp đa dạng các ứng dụng và dịch vụ cho các VU, đặc biệt là trong các đô thị thông minh [47]. Bằng cách tận dụng khả năng lưu trữ của VANET, kỹ thuật lưu trữ được thiết kế như một phương tiện nhằm hỗ trợ việc phân phối khối lượng dữ liệu lớn được tạo ra bởi các ứng dụng và dịch vụ, chẳng hạn như an toàn công cộng, hỗ trợ lái xe nâng cao, lái xe tự động, cập nhật phần mềm qua mạng, quản lý chuyển giao, hệ thống giải trí, bản sao số, và thậm chí cả các ứng dụng mới như metaverse [35], [50], [54], [82], [83], [84], [85], [86], [87], [88]. Mỗi loại ứng dụng và dịch vụ đều có vai trò

riêng, phù hợp của kỹ thuật lưu trữ, và vai trò này sẽ được mô tả chi tiết trong các phần tiếp theo.

Đối với các ứng dụng an toàn công cộng và hỗ trợ lái xe nâng cao trong VANET [47], [82], các phản hồi đối với các yêu cầu chỉ dẫn an toàn thường xuyên có thể được lưu trữ càng nhiều càng tốt tại các VU nhằm giảm tải cho mạng lõi. Trong các hệ thống lái xe tự động có hỗ trợ cập nhật qua mạng [48], [50], [54], [83], kỹ thuật lưu trữ được tích hợp với tính toán tại biên để chuyển tải các nội dung, dịch vụ có khả năng tái sử dụng cao cũng như các tác vụ tính toán chuyên sâu về phía các máy chủ tại biên. Cách tiếp cận này đồng thời cải thiện cả chất lượng trải nghiệm (QoE – Quality of Experience) và tính năng vận hành của các hệ thống phương tiện tự hành được kết nối.

Đối với các VAS, kỹ thuật lưu trữ được triển khai tại các VU nhằm hỗ trợ quản lý chuyển giao với dung lượng lưu trữ tối thiểu, đồng thời tối đa hóa tốc độ truyền trung bình để đảm bảo quá trình truyền dẫn liên tục [84]. Để bảo đảm chất lượng cao cho truyền video trong mạng VANET có đặc tính biến thiên theo thời gian, một video có thể được chia thành nhiều phiên bản với các mức chất lượng khác nhau và được lưu trữ tại trạm gốc MBS nhằm hỗ trợ phân phối thích ứng đến các phương tiện [35].

Đối với các ứng dụng bản sao số (DT – Digital Twin), do khối lượng lớn nội dung và dịch vụ cần được phân phối, lưu trữ trong VANET đóng vai trò quan trọng trong việc hỗ trợ truyền nội dung có ràng buộc độ trễ và cải thiện tỷ lệ truy cập nội dung lưu trữ thành công (cache hit rate). Các ứng dụng DT có hỗ trợ lưu trữ có khả năng mô hình hóa mối quan hệ xã hội giữa các phương tiện, dự đoán các mô hình luồng yêu cầu dịch vụ và phương tiện, đồng thời điều phối một cách thông minh các tài nguyên truyền thông và lưu trữ nhằm bảo đảm việc phân phối nội dung liên tục và hiệu quả [85], [86], [87]. Trong bối cảnh các phương tiện tự hành và hệ thống giao thông thông minh được dẫn dắt bởi metaverse, bài toán lưu trữ ngày càng trở nên phức tạp, đòi hỏi các quyết định cân trọng về nội dung cần lưu trữ, vị trí lưu trữ, thời điểm lưu trữ và phương thức lưu trữ, nhằm đạt được tỷ lệ truy cập nội dung và truyền dẫn thành công cao [88].

2.2.2. Phương pháp DC

Được nghiên cứu rộng rãi trong các tài liệu, các phương pháp DC được xem là một chiến lược quan trọng để xây dựng nền tảng lưu trữ nhằm cung cấp các ứng dụng và dịch vụ trong VANET [27], [34], [50], [51], [63], [64], [65], [66], [68], [69], [70], [76], [77], [78], [79], [80], [81]. Trong [68], các tác giả đã giới thiệu một mô-đun dự đoán di động có khả năng ước lượng trước các kết nối giữa các VU và các RSU. Sau đó, một thuật toán dựa trên học máy được áp dụng nhằm xác định những nội dung phổ biến nào cần được lưu trữ, cũng như thời điểm và vị trí lưu trữ chúng tại các RSU, qua đó đạt được hiệu quả sử dụng tài nguyên tối ưu, đặc biệt là giảm mức tiêu thụ băng thông.

Tương tự, bằng cách dự đoán tính di động của các VU và mức độ phổ biến của nội dung, các tác giả trong [70] đã đề xuất một phương pháp linh hoạt cho phép các RSU có thể phối hợp với nhau trong quá trình lưu trữ. Do tính phức tạp của bài toán tối ưu lưu trữ, các thuật toán thực tiễn thường được xây dựng dưới dạng bài toán cái túi và các phép nối lỏng cận tối ưu nhằm đưa ra các quyết định lưu trữ gần tối ưu.

Việc lưu trữ tại RSU được cải thiện khi kết hợp với các kỹ thuật phân cụm [63], [64]. Trong [63], các VU có quỹ đạo tương tự được phân cụm để phục vụ cho mục đích lưu trữ. Cách tiếp cận được đề xuất xây dựng một bài toán tối ưu lưu trữ dựa trên phân cụm có xét đến độ trễ, và giải bài toán này bằng cách sử dụng mạng DQN (Deep Q-Network), với mục tiêu giảm thiểu mức tiêu thụ năng lượng trong khi vẫn bảo đảm độ trễ thấp. Như được trình bày trong [64], việc phân cụm các RSU dựa trên mô hình di động của các VU, thời hạn tải xuống, và tài nguyên lưu trữ sẽ dẫn đến hiệu quả lưu trữ được cải thiện. Bằng cách áp dụng thuật toán phân cụm Markov dựa trên xác suất truy cập cục đại cho lời giải tối ưu kết hợp giữa lưu trữ và phân cụm, tỷ lệ truy cập nội dung lưu trữ thành công được gia tăng đáng kể, đồng thời vẫn thỏa mãn ràng buộc về thời hạn tải xuống.

Hiệu năng lưu trữ tại các RSU có thể được cải thiện thông qua kỹ thuật truyền hợp tác giữa hai chế độ VU-MBS và VU-RSU, như đã được nghiên cứu trong [65]. Điều này có nghĩa là quyết định lưu trữ tại các RSU phụ thuộc vào chất lượng kênh truyền trong cả hai chế độ, với mục tiêu đạt được tỷ lệ truy cập trùng cao và độ trễ thấp.

Cách tiếp cận lưu trữ hợp tác tập trung này, khi được chuyển đổi sang dạng bài toán cái túi đa lựa chọn, sẽ được giải bằng thuật toán tham lam để thu được các kết quả tối ưu xấp xỉ. Trong [27], việc lưu trữ tại các RSU với sự hỗ trợ của các kỹ thuật TrS và TrA được tối ưu hóa dưới các ràng buộc về dung lượng lưu trữ và tốc độ truyền, thông qua việc sử dụng thuật toán heuristic. Cách tiếp cận này có thể nâng cao hiệu quả thông lượng của hệ thống một cách hiệu quả. Các kết quả cho thấy phương pháp lưu trữ hợp tác được đề xuất tại các RSU đạt được độ trễ thấp nhất, do đó có thể được áp dụng cho các ứng dụng nhạy cảm về độ trễ.

Một cách đáng chú ý, nghiên cứu [34] đã xem xét việc lưu trữ tại trạm gốc MBS để phối hợp với lưu trữ tại các RSU. Dựa trên tính di động của VU và mức độ phổ biến của nội dung, một nội dung bất kỳ có thể được lưu trữ đồng thời tại cả MBS và RSU, chỉ tại MBS, hoặc chỉ tại RSU, nhằm đạt được độ trễ truyền dẫn thấp và chi phí dịch vụ nhỏ. Đây là bài toán cái túi đa lựa chọn đa mục tiêu và đa chiều, một thuật toán dựa trên tối ưu hóa đàn kiến được sử dụng để tìm nghiệm gần tối ưu. Tương tự, việc lưu trữ tại các trạm gốc MBS và các trạm phát ven đường cũng được đề xuất trong các nghiên cứu [50], [51], [69] nhằm tối ưu hóa độ trễ, thời gian trễ và tỷ lệ truy cập nội dung lưu trữ thành công. Để giải quyết các thách thức phát sinh từ không gian trạng thái có số chiều lớn và không gian hành động lai liên tục-rời rạc, một phương pháp dựa trên gradient chính sách xác định sâu (DDPG – Deep Deterministic Policy Gradient) được áp dụng để giải bài toán tối ưu độ trễ [50]. Trong khi đó, sự kết hợp giữa học liên kết không đồng bộ và học tăng cường sâu (DRL – Deep Reinforcement Learning) được sử dụng nhằm tăng cường bảo vệ quyền riêng tư của người dùng đầu cuối và cải thiện tỷ lệ truy cập trúng [51]. Ngoài ra, trong [69], để xử lý tính phức tạp phát sinh từ việc dự đoán tính di động và mức độ phổ biến nội dung, bảo mật dữ liệu, cũng như lưu trữ hợp tác động, các tác giả đã đồng thời sử dụng mạng nhớ ngắn-dài hạn (LSTM – Long Short-Term Memory), học liên kết phân cấp (FL – Federated Learning), và thuật toán giảm dốc gradient thích nghi nhằm nâng cao hiệu năng lưu trữ xét theo các chỉ tiêu tỷ lệ truy cập trúng và độ trễ.

Hơn nữa, các phương pháp lưu trữ xác định có thể được triển khai đồng thời tại cả các trạm phát ven đường và người dùng đầu cuối, hoặc chỉ triển khai tại các VU [66],

[76], [77], [78], [79], [80], [81]. Trong [66], các tác giả đã phân tích các đặc trưng của nội dung (mức độ phổ biến và kích thước), tính di động, cũng như mật độ lưu thông để tối ưu hóa việc lưu trữ tại các RSU và VU bằng cách áp dụng thuật toán entropy chéo. Mục tiêu là cải thiện tỷ lệ truy cập trúng, giảm độ trễ, đồng thời duy trì chi phí xử lý thấp. Trong khi đó, các tác giả trong [76] đã khai thác mối quan hệ xã hội giữa các VU nhằm tối thiểu hóa độ trễ truy cập thông qua DRL, và các tác giả trong [77] đã sử dụng kỹ thuật phân cụm để tối đa hóa tỷ lệ truy cập trúng và giảm độ trễ bằng cách áp dụng LSTM. Tương tự, nghiên cứu [78] đặt mục tiêu cải thiện tỷ lệ truy cập trúng và giảm độ trễ, đồng thời vẫn duy trì chi phí thấp. Một thuật toán tham lam dựa trên hợp tác cũng được đề xuất để giải quyết bài toán lưu trữ kết hợp tại các RSU và VU trong các hệ thống quy mô lớn với độ phức tạp thời gian thấp. Khi việc lưu trữ được thực hiện hoàn toàn tại các VU, các chính sách khuyến khích, bảo vệ quyền riêng tư, và độ chính xác cao trở thành những yêu cầu nghiêm ngặt nhằm nâng cao hiệu năng lưu trữ. Các vấn đề này có thể được giải quyết thông qua mô hình trò chơi Stackelberg [79] và các phương pháp DQN dựa trên học liên kết [80], [81].

Bảng 2.1 tổng hợp các phương pháp lưu trữ xác định nêu trên, triển khai theo nhiều cấu hình khác nhau, bao gồm: chỉ tại RSU, tại RSU và trạm gốc MBS, tại RSU và VU, hoặc chỉ tại VU. Tất cả các phương pháp này đều khai thác tính di động, mức độ phổ biến của nội dung và các kỹ thuật hỗ trợ khác nhằm nâng cao hiệu năng phân phối nội dung. Tuy nhiên, vấn đề đặt ra là các phương pháp này không khai thác các tài nguyên lưu trữ sẵn có của VANET để lưu trữ nội dung tại các VU, cũng như không tận dụng được các lợi ích của các phương pháp lưu trữ xác suất. Ngoài ra, số lượng các RSU được triển khai thường tương đối nhỏ, trong khi các đặc tính về khả năng hội tụ, độ ổn định và độ chính xác của lời giải chưa được nghiên cứu đầy đủ (ngoại trừ [63]). Do đó, các phương pháp DC hiện tại chưa thể sử dụng hiệu quả tài nguyên lưu trữ cho lưu trữ nội dung hoặc cải thiện đáng kể hiệu năng của hệ thống.

2.2.3. Phương pháp PC

Trái ngược với các phương pháp DC dựa trên việc xác định trước vị trí lưu trữ tại các trạm gốc MBS, các RSU và VU, phương pháp PC áp dụng các chiến lược động tại các VU thông qua các mô hình xác suất [67], [71], [72], [73], [74], [75]. Theo đó,

phương pháp PC cho phép các VU tham gia hoặc rời khỏi hệ thống lưu trữ một cách ngẫu nhiên, nhằm thích ứng với các điều kiện thời gian thực và tính động của VANET. Đặc biệt, trong [71], các tác giả đã đề xuất một phương pháp lưu trữ xác suất phân tán bằng cách xem xét nhu cầu và mức độ quan trọng của người dùng đầu cuối, mức độ phổ biến của nội dung, cũng như chuyển động tương đối. Trình mô phỏng NS-3 được sử dụng để chứng minh rằng phương pháp đề xuất có thể cải thiện tỷ lệ truy cập nội dung lưu trữ thành công trong khi vẫn duy trì độ trễ thấp và hiệu quả sử dụng dung lượng lưu trữ cao. Tương tự, nghiên cứu [72] tiếp tục xem xét các mức đánh giá quyền riêng tư của các VU nhằm đạt được không chỉ độ trễ thấp và tỷ lệ truy cập trúng cao, mà còn giảm mức độ dư thừa lưu trữ, thông qua các mô phỏng được thực hiện trong môi trường OMNeT++.

Phương pháp PC cũng có thể được sử dụng để phục vụ người dùng di động thông thường (MU – Mobile User), như đã được nghiên cứu trong [67]. Để xây dựng bài toán tối ưu lưu trữ PC theo hướng hiệu quả về năng lượng, nghiên cứu này sử dụng quá trình Markov để mô hình hóa sự tương tác giữa các VU và các MU. Mức độ phổ biến của nội dung cũng được xem xét nhằm nâng cao hiệu quả của lời giải. Kỹ thuật truyền hợp tác từ các trạm gốc MBS và các VU đến các MU được khai thác nhằm cải thiện hiệu năng tổng thể của hệ thống. Các phương pháp lập trình phân đoạn phi tuyến và lý thuyết tối ưu Lyapunov được sử dụng để giải bài toán xác định xác suất lưu trữ tối ưu, đạt được tỷ lệ truy cập trúng cao và hiệu quả năng lượng. Trong [75], mối quan hệ xã hội giữa các VU được đưa vào xem xét bên cạnh tính di động của các VU và mức độ phổ biến của nội dung. Các trạm phát ven đường chịu trách nhiệm thu thập thông tin chuyển động của các VU và cập nhật nội dung mới cho các VU thông qua mô hình Markov ẩn. Trình mô phỏng môi trường mạng cơ hội được triển khai nhằm chứng minh lợi ích của phương pháp PC đề xuất có xét đến mối quan hệ xã hội, thông qua các chỉ tiêu như tỷ lệ truy cập trúng, độ trễ truy cập trung bình, số lượng chặng trung bình, và mức sử dụng bộ nhớ lưu trữ trung bình.

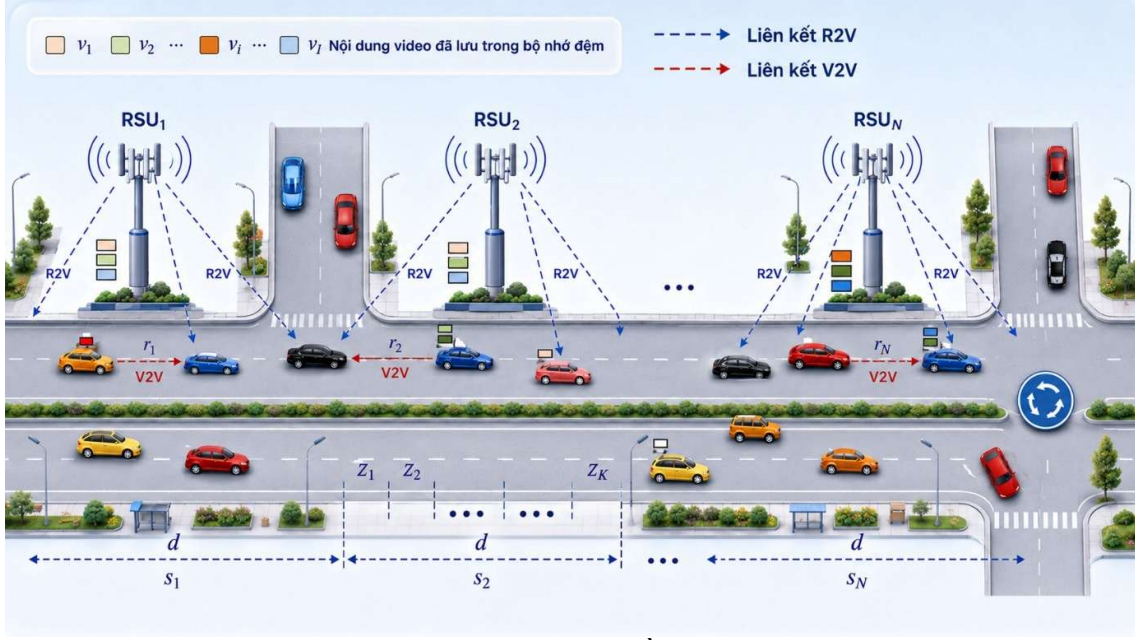
Giải pháp kết hợp giữa mạng dữ liệu định danh (NDN – Named Data Networking) và lưu trữ xác suất không gian được nghiên cứu trong [74] nhằm đề xuất một cơ chế truyền thông cho việc phân phối nội dung trong VANET. Mạng nơ-ron tích

chập (CNN – Convolutional Neural Network) được áp dụng để nắm bắt mối quan hệ phức tạp giữa tính di động của các VU và các yêu cầu nội dung. Phương pháp PC dựa trên CNN được xây dựng nhằm tối ưu hóa xác suất lưu trữ, qua đó, đạt được tỷ lệ truy cập nội dung thành công và độ trễ tốt hơn so với các phương pháp không sử dụng CNN. Một hướng mở rộng đầy hứa hẹn khác của PC là việc tích hợp cả các VU đang di chuyển và các VU có khả năng lưu trữ, như được khảo sát trong [73]. Nghiên cứu này chú trọng đến mật độ giao thông, mức độ trung tâm của tuyến đường, và mức độ phổ biến của nội dung để đưa ra quyết định về xác suất lưu trữ nội dung tại các VU. Các kết quả mô phỏng cho thấy tính hiệu quả của lời giải được đề xuất thông qua tỷ lệ truy cập nội dung thành công cao và độ trễ truy xuất nội dung thấp.

Đối với các phương pháp lưu trữ xác suất đã đề cập ở trên [67], [71], [72], [73], [74], [75], có thể nhận thấy từ bảng 2.1 rằng, các phương pháp này được triển khai trong các hệ thống đơn giản bằng cách chỉ thực hiện lưu trữ tại người dùng đầu cuối. Trong các phương pháp này, các trạm gốc MBS và các trạm phát ven đường chủ yếu đảm nhiệm vai trò truyền nội dung một cách hợp tác [67] hoặc thu thập thông tin di động của các VU [75]. Hơn nữa, các đặc tính về khả năng hội tụ, độ ổn định và độ chính xác của các thuật toán và lời giải được đề xuất vẫn chưa được nghiên cứu đầy đủ, do đó các kết quả tối ưu thu được chưa thực sự thuyết phục. Mặc dù mối quan hệ xã hội giữa các VU và cách tiếp cận dựa trên CNN đã được khai thác trong [74], [75], việc chưa tận dụng được đầy đủ tài nguyên lưu trữ biên mạnh mẽ và khả năng truyền thông không dây ổn định của các RSU trong VANET khiến cho các phương pháp PC truyền thống trở nên kém hiệu quả.

2.3. MÔ HÌNH HỆ THỐNG

Mô hình hệ thống tối ưu lưu trữ DPC được miêu tả như hình 2.1. Mô hình gồm N RSU và I video. Các VU được phân bố trong không gian theo quá trình Poisson đồng nhất một chiều (1D-PPP – One-dimensional homogeneous Poisson point process) Φ với mật độ λ [89], [90]. Mỗi RSU bao phủ một đoạn đường có chiều dài d mét, được chia thành K vùng. Các video có thể được lưu trữ tại cả RSU và phương tiện. Các phương tiện đi vào đoạn đường n , với $n = 1, 2, \dots, N$, và di chuyển với vận tốc trung bình là s_n km/h.



Hình 2.1: Mô hình DPC cho truyền video trong VANET

Có hai loại phương tiện: phương tiện lưu trữ (CV – Caching VU) sẵn sàng lưu trữ video, và các phương tiện yêu cầu (RV – Requesting VU) sẽ gửi yêu cầu truy xuất video. Để thuận lợi cho việc triển khai mô hình đề xuất nhưng không làm mất tính tổng quát, nghiên cứu giả định rằng cả các RSU và các phương tiện đều có tài nguyên hạn chế dành cho việc lưu trữ; do đó, một ràng buộc được đưa ra nhằm bảo đảm việc sử dụng tài nguyên một cách hiệu quả. Bên cạnh đó, một cơ chế khuyến khích được áp dụng nhằm thúc đẩy các xe lưu trữ các video phổ biến [79]. Về thông tin trạng thái kênh (CSI – Channel State Information), NCS giả định rằng thông tin thống kê của CSI có thể được các RSU ước lượng theo chu kỳ [32], [33].

Trong mô hình này, nhằm cải thiện QoS của VAS trong VANET, một bài toán tối ưu lưu trữ kết hợp DPC được xây dựng và giải nhằm xác định vị trí lưu trữ video tối ưu tại các RSU theo phương pháp DC và tối ưu xác suất lưu trữ video được triển khai tại các CV theo phương pháp PC. Mục tiêu của bài toán là tối thiểu thời gian phục vụ trong khi vẫn khai thác hiệu quả tài nguyên lưu trữ của các RSU và các CV. Các thông số hệ thống được mô tả trong bảng 2.2. Thời gian phục vụ được định nghĩa là khoảng thời gian trung bình để các xe yêu cầu (RV) nhận được các video đã yêu cầu từ các RSU và các CV trong khi đang di chuyển. Sau khi triển khai phương pháp DPC, giả sử một RV

yêu cầu video i , với $i = 1, 2, \dots, I$, đi vào một đoạn đường được bao phủ bởi RSU n , thì RV đó sẽ được phục vụ theo một trong hai trường hợp sau:

Bảng 2.2: Bảng thông số hệ thống mô hình DPC

Ký hiệu	Thông số
N	Số lượng RSU
K	Số lượng vùng được phủ bởi RSU
I	Số lượng video
d	Vùng phủ của một RSU
d_k	Độ dài của vùng k trong một đoạn đường, $k = 1, 2, \dots, K$
R_k	Tốc độ truyền trong vùng k
S_i	Kích thước video i đo bằng bits, $i = 1, 2, \dots, I$
p_i	Tốc độ truy cập (độ phổ biến) của video i
ρ_i	Xác suất lưu trữ video i
λ	Mật độ xe trung bình trên toàn tuyến đường
μ_n	Tốc độ phục vụ bởi RSU n
s_n	Tốc độ trung bình của xe trong vùng phủ của RSU n

- Nếu RSU n không lưu trữ video i , RV sẽ nhận video i từ bất kỳ CV nào đang lưu trữ video i trong một phạm vi truyền xác định r_n thông qua các kết nối V2V.
- Nếu RSU n lưu trữ video i , RSU đó phục vụ RV thông qua các kết nối R2V với sự phối hợp của các CV.

Nếu video i có kích thước lớn không thể được truyền hoàn toàn sau khi các RV đã đi qua toàn bộ N RSU, thì quá trình truyền sẽ được tiếp tục bởi một tập các RSU tiếp theo, gắn với một trạm gốc MBS khác. Kịch bản này nằm ngoài phạm vi của luận án và sẽ được thực hiện trong các nghiên cứu trong tương lai.

2.3.1. Quá trình Poisson đồng nhất cho các VU

Đối với các xe, không làm mất tính tổng quát, nghiên cứu áp dụng quá trình Poisson đồng nhất một chiều độc lập (T1D-PPP) để mô hình hóa phân bố không gian

của các CV phục vụ truyền dữ liệu và các RV thực hiện truy xuất trên các tuyến đường một chiều [89], [90]. τ và ρ_i lần lượt là tỷ lệ các phương tiện đóng vai trò là CV và xác suất lưu trữ video i , khi đó phân bố của các CV và RV cũng tuân theo các T1D-PPP Φ_i^{CV} và Φ_i^{RV} với các mật độ λ_i^{CV} và λ_i^{RV} , được biểu diễn như sau:

$$\lambda_i^{CV} = \tau \rho_i \lambda, \quad (2.1)$$

$$\lambda_i^{RV} = (1 - \tau \rho_i) \lambda. \quad (2.2)$$

2.3.2. DC và PC

Trong phương pháp DC, dọc theo tuyến đường gồm N trạm phát ven đường, vị trí lưu trữ của video i được xác định thông qua bán kính lưu trữ h_i , được đo theo số chặng (hop) và tỉ lệ nghịch với mật độ lưu trữ [91]. Điều này có nghĩa là video i được lưu trữ tại các RSU cách nhau mỗi h_i chặng. Nếu $h_i = 1$, video i sẽ được lưu trữ tại tất cả các RSU. Trong khi đó, đối với phương pháp PC, xác suất ρ_i để xác định khả năng lưu trữ video i tại các CV. Trong quá trình di chuyển, các RV được phục vụ bởi cả các RSU và các CV thông qua phương pháp lưu trữ kết hợp DPC, trong đó có xét đến mô hình phổ biến của video i , tuân theo phân bố Zipf-like [46], được biểu diễn như công thức (1.1).

2.3.3. Kênh truyền V2V

1) Dung lượng V2V

Nếu video i được một RV bất kỳ yêu cầu, trong trường hợp xấu nhất, RV đó cần một khoảng thời gian di chuyển đi qua $h_i - 1$ chặng để đến được RSU lưu trữ video i . Trong khoảng thời gian di chuyển này, RV được phục vụ bởi các CV. Trong mô hình T1D-PPP, xác suất để một CV diễn hình lưu trữ và một RV diễn hình yêu cầu video i trong phạm vi truyền r_n tại đoạn đường thứ n lần lượt được cho bởi [90]:

$$p_{i,n}^{CV} = 1 - e^{-2r_n \lambda_i^{CV}}, \quad (2.3)$$

$$p_{i,n}^{RV} = 1 - e^{-2r_n \lambda_i^{RV}}. \quad (2.4)$$

Kênh truyền V2V giữa các CV và RV được định nghĩa theo mô hình suy hao ngẫu nhiên quy mô nhỏ có thành phần truyền thẳng dựa trên phân bố Rician [92], [93],

[94]. Do đó, dung lượng kênh giữa một CV a (với $a \in \Phi_i^{\text{CV}}$) và một RV b (với $b \in \Phi_i^{\text{RV}}$) được xác định như sau:

$$C_{a,b} = W \log_2 \left(1 + \frac{P|h|^2 d_{a,b}^{-\eta}}{N_0} \right), \quad (2.5)$$

trong đó W là băng thông hệ thống, P là công suất truyền, η là hệ số suy hao đường truyền, và $d_{a,b}$ là khoảng cách từ CV a đến RV b . h biểu diễn hệ số kênh, được tính như sau:

$$h = \sqrt{\frac{\kappa}{1+\kappa}} e^{j\theta} + \sqrt{\frac{1}{1+\kappa}} \omega, \quad (2.6)$$

trong đó κ là hệ số Rician, θ là biến ngẫu nhiên phân bố đều trong khoảng $[0, 2\pi]$, và ω là biến ngẫu nhiên Gauss phức có kỳ vọng bằng 0 và phương sai bằng 1, $\mathcal{CN}(0,1)$.

Nghiên cứu này giả định rằng, trong phạm vi truyền dẫn r_n đã cho, chỉ có một CV lưu trữ video i chịu trách nhiệm phục vụ RV tương ứng thông qua các kết nối từ thiết bị đến thiết bị (device to device).

Dựa trên công thức (2.5) và trong trường hợp xấu nhất, dung lượng truyền video i đến một RV bất kỳ trong phạm vi truyền cực đại $d_{a,b} = r_n$ được xác định bởi:

$$C_{i,n} = p_{i,n}^{\text{CV}} p_{i,n}^{\text{RV}} W \log_2 \left(1 + \frac{P|h|^2 r_n^{-\eta}}{N_0} \right). \quad (2.7)$$

2) Thời gian truyền V2V

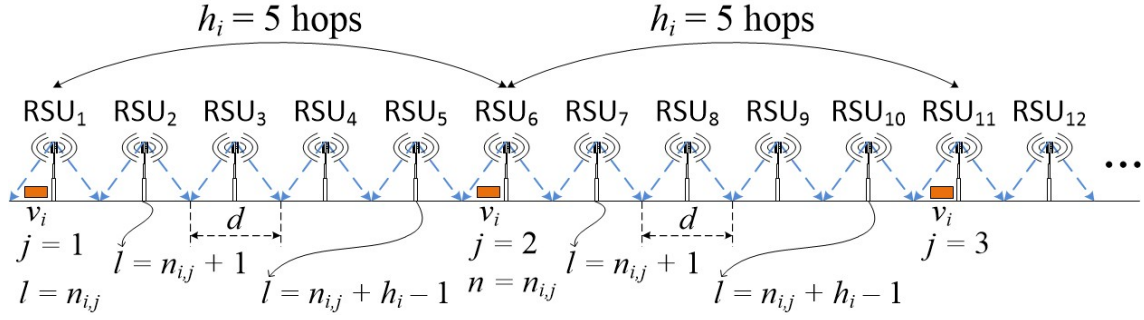
Cho trước một bán kính lưu trữ h_i , có thể dễ dàng xác định được số lượng các RSU lưu trữ video i theo công thức (1.2):

$$J_i = \left\lfloor \frac{N-1}{h_i} \right\rfloor + 1,$$

trong đó, toán tử $\lfloor \cdot \rfloor$ được sử dụng để lấy phần nguyên lớn nhất không vượt quá một số thực cho trước.

Như vậy, video i được lưu trữ tại RSU $n_{i,j} = h_i(j-1) + 1$, với $j = 1, 2, \dots, J_i$. Khi các RV di chuyển ra khỏi vùng phủ của RSU $n_{i,j}$, thời gian để nhận video i từ các CV trong đoạn đường $m_{i,j}$, với $m_{i,j} = n_{i,j} + 1, n_{i,j} + 2, \dots, n_{i,j} + h_i - 1$, được xác định như sau:

$$t_{i,m_{i,j}} = \frac{d}{s_{m_{i,j}}}. \quad (2.8)$$



Hình 2.2: Ví dụ mô hình DPC với $h_i = 5$

Nghiên cứu giả định rằng không tồn tại khoảng trống truyền dẫn giữa hai RSU liền kề. Tuy nhiên, nếu bán kính lưu trữ $h_i > 1$, sẽ tồn tại một khoảng trống lưu trữ giữa hai RSU liền kề cùng lưu trữ video i .

2.3.4. Kênh truyền R2V

1) Dung lượng R2V

Để xác định dung lượng R2V, NCS áp dụng mô hình truyền R2V được trình bày trong [34]. Trong mô hình này, phạm vi phủ sóng d của một RSU được chia thành K vùng. Các tham số của vùng k , với $k = 1, 2, \dots, K$ và $K = 7$, được liệt kê trong bảng 1.1, cụ thể như sau:

Vùng	1	2	3	4	5	6	7
d_k (m)	25	30	40	60	40	30	25
R_k (Mbps)	1	2	5.5	11	5.5	2	1

Dung lượng R2V cho một RV yêu cầu video i tại vùng k và được phục vụ bởi RSU n được xác định như sau:

$$R_{i,n}^{(k)} = \begin{cases} 0, & \text{nếu } n \neq n_{i,j}, \\ \frac{(1-p_{i,n}^{\text{CV}}) p_{i,n}^{\text{RV}} R_k}{V_{i,k}}, & \text{nếu } n = n_{i,j}. \end{cases} \quad (2.9)$$

trong đó $p_{i,k}^{RV}$ và $V_{i,k}$ lần lượt là xác suất để một RV tại vùng k yêu cầu video i từ RSU n , và số lượng các RV yêu cầu video i trong vùng k , được biểu diễn như sau:

$$p_{i,k}^{RV} = 1 - e^{-d_k \lambda_i^{RV}}, \quad (2.10)$$

$$V_{i,k} = d_k \lambda_i^{RV}. \quad (2.11)$$

2) Thời gian truyền R2V

Thời gian truyền của RSU n để phục vụ các RV tại vùng k , tức là thời gian trung bình để di chuyển xuyên suốt vùng k được bao phủ bởi RSU n , được tính như sau:

$$t_{n,k} = \frac{d_k}{s_n}. \quad (2.12)$$

2.3.5. Thời gian phục vụ trung bình

Thời gian phục vụ trung bình tổng thể đối với I video, đây cũng chính là hàm mục tiêu cần được tối thiểu hóa thông qua việc tìm h_i và ρ_i , được tính như sau:

$$\bar{T} = \sum_{i=1}^I \rho_i t_i. \quad (2.13)$$

trong đó t_i là thời gian trung bình để nhận video i một cách hợp tác từ các CV và các RSU, như được trình bày trong thuật toán 2.1.

Thuật toán 2.1 mô tả quá trình phân phối video trong mô hình DPC, trong đó, quá trình truyền nội dung được thực hiện thông qua sự phối hợp giữa các RSU và các CV trong mạng VANET. Thuật toán nhận đầu vào là kích thước của video S_i cần truyền đến phương tiện yêu cầu, đồng thời khởi tạo thời gian phục vụ ban đầu $t_i = 0$. Mục tiêu của thuật toán là xác định tổng thời gian cần thiết để truyền toàn bộ video i đến RV, trên cơ sở khai thác đồng thời khả năng truyền dữ liệu từ hạ tầng RSU và từ các CV lân cận có lưu trữ nội dung.

Trước hết, thuật toán xét các cụm RSU có khả năng tham gia phục vụ video i . Vòng lặp ngoài cùng được thực hiện theo chỉ số j từ 1 đến J_i , trong đó J_i biểu diễn số cụm có thể tham gia vào quá trình phân phối video. Với mỗi giá trị j , RSU đầu tiên trong cụm được xác định bởi biểu thức $n_{i,j} = h_i(j - 1) + 1$. Sau khi xác định RSU

tương ứng, thuật toán lần lượt xét các phương tiện trong vùng phủ của RSU thông qua vòng lặp theo chỉ số k .

Thuật toán 2.1: Phân phối video DPC

Input: S_i
 $t_i = 0$
Output: t_i

```

1: for  $j = 1 : J_i$  do
2:    $n_{i,j} = h_i(j - 1) + 1$ 
3:   for  $k = 1 : K$  do
4:     if  $S_i > 0$  then
5:        $S_i = S_i - t_{k,n_{i,j}}(C_{i,n_{i,j}} + R_{i,n_{i,j}}^{(k)})$ 
6:        $t_i = t_i + t_{n_{i,j},k}$ 
7:     else
8:       break
9:     end if
10:  end for
11:  for  $m_{i,j} = n_{i,j} + 1 : n_{i,j} + h_i - 1$  do
12:    if  $S_i > 0$  and  $m_{i,j} \leq N$  then
13:       $S_i = S_i - t_{i,m_{i,j}}C_{i,m_{i,j}}$ 
14:       $t_i = t_i + t_{i,m_{i,j}}$ 
15:    else
16:      break
17:    end if
18:  end for
19: end for

```

Tại mỗi lần lặp, nếu phần dữ liệu video còn lại S_i vẫn lớn hơn 0, hệ thống sẽ thực hiện truyền dữ liệu đến RV thông qua khả năng phục vụ của RSU và sự hỗ trợ từ các CV lân cận. Cụ thể, kích thước dữ liệu còn lại được cập nhật theo biểu thức tại dòng 5, trong đó, $C_{i,n_{i,j}}$ và $R_{i,n_{i,j}}^{(k)}$ được tính từ công thức (2.7) và (2.9) phản ánh phần tốc độ truyền bổ sung nhờ cơ chế hợp tác với các phương tiện lân cận. Đồng thời, thời gian phục vụ t_i được cộng thêm một lượng tương ứng với thời gian truyền $t_{n_{i,j},k}$. Quá trình này được lặp lại cho đến khi toàn bộ dữ liệu video đã được truyền hết hoặc không còn khả năng truyền thêm dữ liệu trong bước hiện tại.

Sau khi khai thác khả năng cung cấp nội dung từ RSU và các phương tiện trong vùng phủ trực tiếp, thuật toán tiếp tục xét các nút thuộc cùng cụm phục vụ, được biểu diễn qua chỉ số $m_{i,j}$ từ $n_{i,j} + 1$ đến $n_{i,j} + h_i - 1$. Bước này nhằm tính đến khả năng tiếp tục phân phối phần dữ liệu còn lại thông qua các nút kế tiếp trong cụm. Nếu tại thời điểm đó dữ liệu video vẫn còn chưa được truyền hết và chỉ số $m_{i,j}$ không vượt quá tổng số nút N trong hệ thống, dữ liệu còn lại sẽ tiếp tục được cập nhật theo lượng dữ liệu truyền được từ nút tương ứng (dòng 13), đồng thời thời gian phục vụ t_i cũng được cộng thêm lượng thời gian truyền tương ứng (dòng 14). Nếu video đã được truyền xong hoặc chỉ số xét vượt ra ngoài phạm vi hệ thống, vòng lặp sẽ dừng lại.

Có thể thấy rằng, thuật toán được tổ chức theo hai bước kế tiếp nhau. Ở bước thứ nhất, hệ thống ưu tiên khai thác khả năng phân phối nội dung từ RSU kết hợp với hỗ trợ truyền hợp tác từ các phương tiện lân cận. Ở bước thứ hai, nếu dữ liệu video vẫn chưa được phục vụ đầy đủ, thuật toán tiếp tục mở rộng phạm vi xét sang các nút còn lại trong cùng cụm phục vụ để hoàn tất quá trình truyền. Cách tổ chức này phản ánh đúng bản chất của mô hình DPC, trong đó, phương pháp DC tại RSU và PC tại CV được kết hợp trong cùng một cơ chế phân phối nội dung nhằm rút ngắn thời gian phục vụ video.

Kết quả cuối cùng của thuật toán là giá trị t_i , biểu diễn tổng thời gian cần thiết để phục vụ video i cho RV trong cấu hình hệ thống đã cho. Thông qua thuật toán này, mô hình DPC cho thấy khả năng khai thác đồng thời tính ổn định của hạ tầng RSU và tính linh hoạt của các CV, từ đó nâng cao hiệu quả phân phối nội dung trong mạng VANET.

2.4. THIẾT KẾ TỐI ƯU DPC

2.4.1. Bài toán tối ưu DPC

Dựa trên hàm mục tiêu (2.13) và đồng thời xét thêm ràng buộc về tài nguyên lưu trữ của các RSU và các CV, bài toán tối ưu DPC được xây dựng như sau:

$$\min \bar{T} \quad (2.14a)$$

$$h_i, \rho_i \quad (2.14b)$$

$$\text{s.t.} \quad 1 \leq h_i \leq N - 1, \forall i \quad (2.14c)$$

$$S^{\text{CR}} + S^{\text{CV}} \leq \delta N \sum_{i=1}^I S_i$$

Với h_i : bán kính lưu trữ và ρ_i : xác suất lưu trữ và trong đó ràng buộc (2.14c) được sử dụng nhằm giới hạn tổng mức tiêu thụ tài nguyên cho việc lưu trữ tại cả các trạm phát ven đường và các xe lưu trữ, với $0 < \delta < 1$.

Ở đây, S^{CR} và S^{CV} lần lượt là lượng tài nguyên lưu trữ được tiêu thụ bởi các RSU và các CV, được tính như sau:

$$S^{\text{CR}} = \sum_{i=1}^I J_i S_i, \quad (2.15)$$

$$S^{\text{CV}} = \tau \lambda d(N-1) \sum_{i=1}^I \rho_i S_i. \quad (2.16)$$

Biểu thức (2.14) mô tả bài toán tối ưu của mô hình DPC, trong đó, mục tiêu là xác định chiến lược lưu trữ phù hợp nhằm cực tiểu thời gian phục vụ video dưới các ràng buộc tài nguyên lưu trữ của hệ thống. Đây là một bài toán tối ưu hỗn hợp nguyên-thực, trong đó hệ thống phải đồng thời ra hai quyết định :

- *Quyết định lưu trữ xác định tại RSU* thông qua biến nguyên h_i – bán kính lưu trữ video i , xác định vị trí lưu trữ của video đó trên các RSU dọc tuyến đường. Giá trị h_i nhỏ tương ứng với việc video i được lưu trữ tại nhiều RSU (mật độ dày), trong khi h_i lớn nghĩa là video chỉ được lưu trữ ở một số ít RSU cách xa nhau.
- *Quyết định lưu trữ xác suất tại phương tiện* thông qua biến thực $\rho_i \in [0,1]$, xác suất mà mỗi phương tiện tham gia lưu trữ video i . Đây là biến thể hiện mật độ phương tiện lưu trữ (CV) video i trong mạng.

Hai biến này có mối liên kết chặt chẽ: khi RSU lưu trữ nhiều (h_i nhỏ), nhu cầu lưu trữ tại phương tiện giảm và ngược lại. Bài toán DPC tìm điểm cân bằng tối ưu giữa hai tầng lưu trữ này dưới điều kiện tài nguyên hữu hạn.

Ngoài ra, trong bài toán DPC, ràng buộc (2.14b) xác định miền giá trị của biến nguyên h_i . Cận dưới $h_i \geq 1$ đảm bảo rằng mỗi video i phải được lưu trữ tại ít nhất một RSU, nếu không, hệ thống sẽ không có nguồn cung cấp video xác định. Cận trên $h_i \leq N-1$ phản ánh thực tế rằng với N RSU dọc tuyến đường, bán kính lưu trữ không thể

vượt quá $N - 1$ vì khi đó toàn bộ hệ thống không có RSU nào lưu video i , đây là trường hợp không thể xảy ra trong lưu trữ theo vị trí xác định. Đối với ràng buộc (2.14c), đây là ràng buộc nhằm phân bổ và tận dụng hiệu quả tổng dung lượng lưu trữ tiêu thụ trên toàn hệ thống. Đại lượng $N \sum_{i=1}^I S_i$ là dung lượng tối đa lý thuyết nếu mỗi RSU lưu trữ toàn bộ video. Hệ số $\delta \in (0,1)$ được dùng để hiệu chỉnh dung lượng lưu trữ thực tế được phép sử dụng so với mức tối đa lý thuyết. Việc hiệu chỉnh này nhằm đảm bảo sự chia sẻ dung lượng lưu trữ của hệ thống với các ứng dụng khác và thuận tiện trong việc khảo sát và đánh giá giữa các cơ chế lưu trữ khác nhau, cũng như cho phép lựa chọn cấu hình hệ thống phù hợp với điều kiện triển khai cụ thể.

2.4.2. Giải bài toán bằng GA

GA đã cho thấy khả năng mạnh mẽ trong việc giải các bài toán tối ưu phức tạp trên nhiều lĩnh vực khác nhau [95], trải dài từ các hệ thống truyền thông [38], [96], [97] đến các VAS [39], [40], [41], [98]. Những ưu điểm chính của GA nằm ở nền tảng dựa trên các nguyên lý tiến hóa của chọn lọc tự nhiên và biến dị di truyền. Điều này cho phép GA linh hoạt cung cấp các nghiệm tối ưu toàn cục chính xác hoặc gần tối ưu, bất kể không gian tìm kiếm là đơn đỉnh hay đa đỉnh. Nói cách khác, GA đặc biệt hiệu quả trong việc tránh các điểm tối ưu cục bộ bằng cách đồng thời khám phá nhiều đỉnh trong không gian tìm kiếm [39], [40]. Nhờ vào tính vững chắc, khả năng thích nghi và năng lực tìm kiếm toàn cục mạnh mẽ, GA đã trở thành một phương pháp được áp dụng rộng rãi để giải quyết các bài toán tối ưu phức tạp trong thực tế.

Nghiên cứu này áp dụng GA [99] để giải bài toán (2.14). Tuy nhiên, vấn đề đặt ra là GA chỉ hỗ trợ các ràng buộc đơn giản dưới dạng cận trên và cận dưới, chẳng hạn như (2.14b), mà không thể trực tiếp xử lý các ràng buộc phức tạp hơn như (2.14c). Vấn đề này được giải quyết bằng cách sử dụng phương pháp hàm phạt [39]. Theo đó, bài toán tối ưu có ràng buộc (2.14) được chuyển đổi thành một bài toán tối ưu không ràng buộc bằng cách biến đổi ràng buộc (2.15c) thành:

$$\Delta S = \delta N \sum_{i=1}^I S_i - (S^{\text{CR}} + S^{\text{CV}}) \geq 0. \quad (2.17)$$

Tiếp theo, hàm phạt được xây dựng như sau:

$$F = \rho(\min\{0, \Delta S\})^2, \quad (2.18)$$

trong đó ρ là hệ số mức độ vi phạm ràng buộc, được sử dụng để điều chỉnh mức độ trừng phạt đối với các cá thể trong GA khi chúng vi phạm các ràng buộc.

Cuối cùng, GA có khả năng giải bài toán tối ưu DPC không ràng buộc sau đây:

$$\min_{h_i, \rho_i} T_F = \bar{T} + F. \quad (2.19)$$

GA chi tiết được sử dụng để giải bài toán (2.19) được trình bày trong thuật toán 2.2. Trong thuật toán 2.2, hai biến tối ưu dạng số nguyên và số thực không thể được biểu diễn bằng cùng một nhiễm sắc thể (chuỗi) trong một cá thể, cũng không thể được xử lý bằng cùng một sơ đồ lai ghép và đột biến. Do đó, chúng được xử lý riêng biệt thông qua hai chuỗi con $X_1^{(z)}$ và $X_2^{(z)}$ của cá thể z . Cụ thể, đối với $X_1^{(z)}$, một phép toán cơ số $(N-1)$ [99] được sử dụng để sinh ra các giá trị số nguyên trong khoảng từ 0 đến $(N-2)$, sau đó cộng thêm 1 để thỏa mãn ràng buộc (2.14b). Ngoài ra, đối với $X_2^{(z)}$, chuỗi này được chuyển đổi thành các biến tối ưu dạng thực $X_{2,R}^{(z)}$ thông qua phép toán b2r (binary-to-real – b2r) [99].

GA được triển khai thông qua một chuỗi các toán tử chính bao gồm lựa chọn, lai ghép, và đột biến, và được lặp lại cho đến khi thỏa mãn một trong hai điều kiện dừng sau:

- Giá trị trung bình của hàm phạt trên mỗi cá thể $\bar{F}$, được suy ra từ (2.18), nhỏ hơn 10^{-3} trong 10 thế hệ liên tiếp.
- Số thế hệ Gen đạt đến một giá trị xác định trước ($N_G = 100$).

Vì GA là thuật toán tìm kiếm heuristic thích nghi, độ phức tạp của thuật toán 2.2 chịu ảnh hưởng trực tiếp bởi kích thước không gian tìm kiếm N_P và số lần lặp N_G . Trong khi đó, các giá trị được lựa chọn cho N_P và N_G có thể lớn hoặc nhỏ, phụ thuộc vào:

- quy mô của hệ thống (N, I) ,
- độ phức tạp tính toán của hàm mục tiêu, các ràng buộc, và các biến tối ưu trong (2.14), và
- yêu cầu về độ chính xác đối với VAS.

Ngoài ra, độ phức tạp của thuật toán 2.2 còn phụ thuộc vào các toán tử lựa chọn, lai ghép, và đột biến. Do đó, độ phức tạp tổng thể của thuật toán này nhìn chung có bậc $O(N_p N_G)$ [40], [96].

Thông thường, thuật toán 2.2 phù hợp hơn khi được thực thi bởi một trạm MBS duy nhất, sau đó áp dụng phương pháp DPC cho toàn bộ các RSU và các CV. Tuy nhiên, trong các hệ thống có quy mô lớn hơn với các giá trị N và I lớn, những thách thức chính trong việc triển khai thực tế phương pháp DPC xuất hiện ở hai khía cạnh:

- Thứ nhất, việc thu thập thông tin hệ thống, chẳng hạn như CSI, để xây dựng bài toán tối ưu DPC trở nên khó khăn hơn.
- Thứ hai, độ phức tạp của thuật toán 2.2 gia tăng đáng kể.

Để giải quyết các thách thức này, số lượng các RSU được quản lý bởi một MBS cần được xác định một cách cẩn trọng, dựa trên quy mô hệ thống và năng lực xử lý của MBS. Ngoài ra, cần lưu ý rằng, nếu các video được yêu cầu có kích thước quá lớn, nhiều MBS có thể phối hợp với nhau để phục vụ liên tục các RV [69].

2.4.3. Triển khai giải pháp DPC

Để hoàn thiện quá trình thiết kế giải pháp DPC, các bước triển khai tại các trạm gốc được mô tả như sau:

- Bước 1: Một trạm gốc chủ được bầu chọn từ các trạm gốc liên quan đến tuyến đường đang xét.
- Bước 2: Trạm gốc chủ yêu cầu các trạm gốc thành viên thu thập thông tin của các RSU và phương tiện trên các đoạn đường được bao phủ liên quan.
- Bước 3: Dựa vào thông tin thu thập, trạm gốc chủ xây dựng và giải bài toán tối ưu để tìm ra h_i và ρ_i .
- Bước 4: Triển khai kết quả tối ưu vào tuyến đường bằng cách: (i) với h_i , trạm nền chủ chỉ định video i được lưu vào N RSU của tuyến với bán kính lưu trữ (ví dụ $h_i = 5$) như được mô tả trong hình 2.2; (ii) với ρ_i , trạm nền chủ sẽ lần lượt lưu trữ video i tại từng phương tiện với xác suất lưu là ρ_i , khi đó, tại mỗi phương tiện, trạm nền chủ sẽ tạo ngẫu nhiên một số $x \in [0, 1]$, nếu $x \leq \rho_i$, thì video i được lưu vào phương tiện đang xét.

Thuật toán 2.2: GA cho DPC**Đầu vào:** Thông số trong bảng 2.3, $Gen = 1$ **Đầu ra:** $T_F^*(h_i^*, \rho_i^*)$

1. Khởi tạo ngẫu nhiên

- N_p chuỗi con $\{X_1^{(z)}\}$, $z = 1, 2, \dots, N_p$, mỗi chuỗi gồm I biến tối ưu nguyên $\{h_i\}$
- N_p chuỗi con $\{X_{2,R}^{(z)}\}$, mỗi chuỗi gồm $(I \times P_R)$ bit để biểu diễn biến tối ưu thực $\{\rho_i\}$, trong đó P_R là số bit dùng để biểu diễn độ chính xác của một biến tối ưu thực, với giá trị được lượng tử hóa thành 2^{P_R} mức.

2. Tính giá trị hàm thích nghi (fitness) theo công thức (2.20) để có $\{T_F^{(z)}\}$, tức là: $T_F^{(z)} = T_F(X_1^{(z)}, X_{2,R}^{(z)})$

3. Khi điều kiện dừng TC chưa thỏa mãn, thực hiện:

4. Đặt $\{X^{(z)}\} = [\{X_1^{(z)}\}\{X_{2,R}^{(z)}\}]$ cùng giá trị hàm phạt $T_F^{(z)}$ vào tập xếp hạng để lai tạo.

5. Lựa chọn $N_{PG} = N_p \times N_G$ cá thể tốt nhất có giá trị fitness nhỏ nhất bằng toán tử lấy mẫu phổ quá ngẫu nhiên [18] để lai ghép, được tập:

$$\{X^{(t)}\} = [\{X_1^{(t)}\}\{X_{2,R}^{(t)}\}], t = 1, 2, \dots, N_{PG}$$

6. Chọn một cặp cá thể bố mẹ để tạo cá thể con bằng lai ghép một điểm với các xác suất P_{ci} và P_{cr} , lần lượt áp dụng cho $X_1^{(t)}$ và $X_{2,R}^{(t)}$.

7. Thực hiện đột biến các cá thể con $\{X_1^{(t)}\}$ và $\{X_{2,R}^{(t)}\}$ tương ứng với các xác suất P_{mi} và P_{mr} bằng phép đảo bit, nhằm khôi phục các đặc trưng di truyền tích cực có thể đã bị mất trong các bước trước để thu được $\{X^{(t),*}\} = [\{X_1^{(t),*}\}\{X_{2,R}^{(t),*}\}]$

8. Lặp lại bước 2 để có $T_F^{(t),*}(X_1^{(t),*}, X_{2,R}^{(t),*})$.

9. Chèn $\{X^{(t),*}\}$ và $T_F^{(t),*}$ vào quần thể hiện tại để có quần thể mới $\{X^{(z)}\}$ và $T_F^{(z)}$.

10. Cập nhật $Gen = Gen + 1$

11. Kết thúc vòng lặp

12. Tìm giá trị hàm phạt tốt nhất $T_F^*(h_i^*, \rho_i^*) \in \{T_F^{(z)}(X_1^{(z)}, X_{2,R}^{(z)})\}$

Quá trình trên được thực hiện trước khi các phương tiện được dự đoán sẽ di chuyển vào tuyến đường đang xét kết hợp với thông tin hiện có trên tuyến đường này (khoảng thời gian này đủ để giải bài toán tối ưu bằng GA và triển khai lưu trữ). Quá trình sẽ được thực hiện định kỳ hoặc khi có sự thay đổi lớn tác động lên các đặc trưng thống kê như khung giờ, thời tiết,... Thực tế, các đặc trưng về mật độ và tốc độ phương tiện cũng như độ nổi tiếng của video có đặc trưng thống kê thay đổi theo giờ. Riêng đặc trưng kênh truyền thay đổi nhanh, giải pháp mô hình phân vùng (zone) và thiết kế bài toán tối ưu dưới dạng “cực đại tình huống xấu nhất” được áp dụng để giảm thiểu ảnh hưởng của sự biến đổi nhanh này – nghĩa là các kết quả tối ưu không trở nên “lỗi thời” khi triển khai.

2.5. ĐÁNH GIÁ HIỆU SUẤT

2.5.1. Thiết lập thông số hệ thống

Các thông số của hệ thống và của GA được trình bày chi tiết trong bảng 2.3. Để đánh giá hiệu năng của phương pháp DPC, DPC được so sánh với các phương pháp khác, bao gồm: chỉ lưu trữ tại các VU (Only Caching in VUs – OCV) [67], chỉ lưu trữ tại các RSU (Only Caching in RSUs – OCR) [65], DPC với cá thể trung bình (DPC with Average Individuals – DAV), và DPC với cá thể kém nhất (DPC with Worst individuals – DWO).

Bốn phương án đối sánh được xây dựng để đánh giá hiệu năng của DPC. OCV [65] và OCR [67] triển khai độc lập lưu trữ xác suất tại VU và lưu trữ xác định tại RSU, kết hợp kỹ thuật CoT (bảng 2.1). DAV và DWO áp dụng cùng GA giải bài toán DPC, nhưng thay vì chọn cá thể tốt nhất như thuật toán 2.2, hai phương án này tính giá trị $\bar{T}$ trong (2.13) trên toàn bộ cá thể thỏa mãn ràng buộc (2.14b) và (2.14c): DAV lấy giá trị trung bình, DWO lấy giá trị lớn nhất. Để đảm bảo công bằng trong so sánh, cả bốn phương án đều bị ràng buộc sử dụng cùng mức tài nguyên lưu trữ như DPC.

Bảng 2.3: Thiết lập thông số cho DPC

Ký hiệu	Giá trị	Đơn vị
λ	0.025	phương tiện/m
N	25	RSU
I	10	video
K	7	vùng/RSU
s_n	Ngẫu nhiên trong khoảng 30 đến 60	km/h
r	Ngẫu nhiên trong khoảng 50 đến 200	m
S_i	Ngẫu nhiên trong khoảng 10 đến 200	Mbits
d_k	[25,30,40,60,40,30,25]	m
d	250	m
R_k	[1,2,5.5,11,5.5,2,1]	Mbps
α	1	
η	4	
τ	0.2	
δ	0.5	
κ	4	
P	0.1	W
ρ	10^{-2}	
N_0	10^{-12}	W
W	10	W
N_P	3000	cá thể
N_G	100	thế hệ
P_G	0.9	
P_R	5 (sử dụng 2^5 bits để lượng tử hóa biến thực ρ_i trong khoảng [0,1])	bits
P_{ci}, P_{cr}	0.6, 0.8	
P_{mi}, P_{mr}	10^{-12} , 10^{-10}	

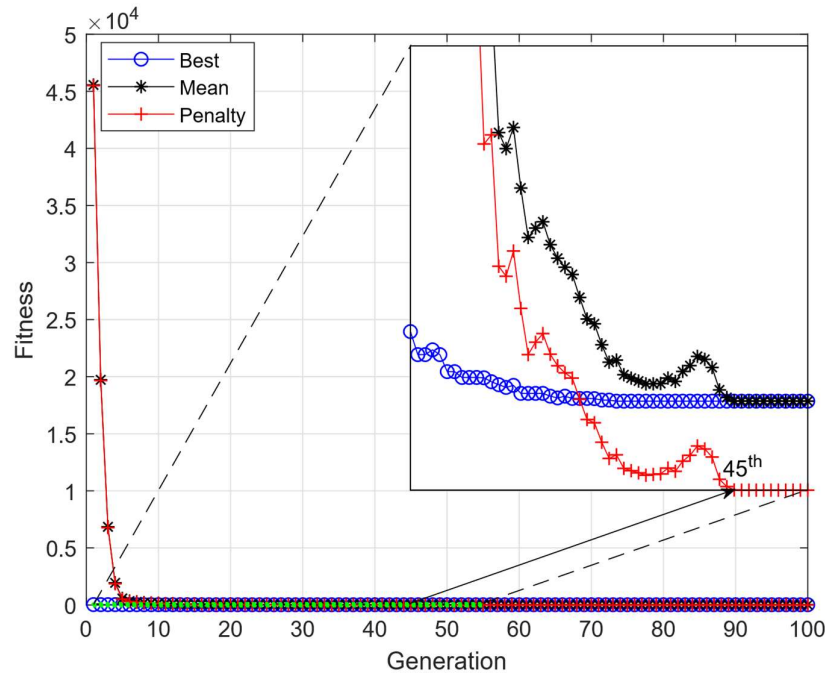
Trong GA, toán tử đột biến đóng vai trò duy trì tính đa dạng của quần thể và mở rộng khả năng khám phá không gian nghiệm. Trong quá trình tiến hóa của GA, toán tử đột biến được sử dụng với xác suất rất nhỏ $p_m = 10^{-12}$, thậm chí có thể bỏ qua quá trình đột biến. Với giá trị xác suất rất nhỏ này, kết hợp với xác suất lai đủ lớn, từ bất kỳ cá thể nào trong quần thể hiện tại, sự kết hợp của phép lai ghép và đột biến, mọi cấu hình nghiệm trong không gian tìm kiếm vẫn có khả năng được tạo ra trong quá trình tiến hóa của quần thể. Hơn nữa, quá trình tiến hóa được lặp lại qua nhiều thế hệ cũng sẽ làm xuất hiện những cá thể có tính đa dạng cao. Do đó, đối với mô hình DPC, việc thiết lập xác suất đột biến ở mức 10^{-12} không làm mất khả năng tiếp cận nghiệm tối ưu của thuật toán về mặt lý thuyết. Ngược lại, việc sử dụng xác suất đột biến rất nhỏ còn giúp hạn chế sự phá vỡ cấu trúc của các cá thể tốt đã hình thành trong quá trình tiến hóa ở thế hệ trước đó, góp phần duy trì tính ổn định của quá trình hội tụ khi tìm kiếm nghiệm tối ưu cho bài toán tối ưu trong mạng UAV-VANET.

Các thông số mô phỏng trong mô hình DPC được lựa chọn cẩn thận nhằm phản ánh các đặc trưng cơ bản của môi trường VANET, bao gồm mật độ phương tiện, tốc độ di chuyển, phạm vi phủ sóng của RSU, kích thước nội dung video, dung lượng lưu trữ và điều kiện kênh truyền R2V/V2V. Các thông số này cho phép tạo ra các kịch bản đại diện để đánh giá xu hướng và hiệu năng của mô hình đề xuất dưới nhiều điều kiện khác nhau. Đặc biệt, các thông số như độ lệch phân bố phổ biến nội dung α , mật độ phương tiện λ và tỷ lệ phương tiện có khả năng lưu trữ τ được thay đổi trong quá trình đánh giá nhằm kiểm tra phản ứng của hệ thống và tính ổn định của phương pháp.

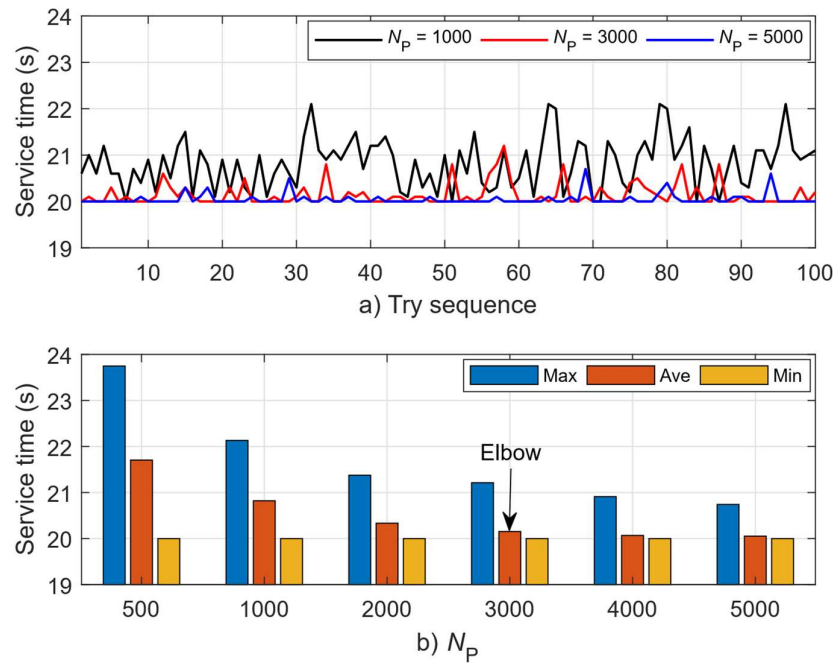
2.5.2. Đánh giá sự hội tụ của GA

Hình 2.3 thể hiện tốc độ hội tụ của GA thông qua việc tính toán các giá trị Best, Mean, và Penalty tại mỗi thế hệ. Trong đó, Best là giá trị hàm thích nghi (fitness) nhỏ nhất gắn với cá thể tốt nhất, trong khi Mean và Penalty lần lượt đại diện cho giá trị trung bình của giá trị hàm thích nghi và giá trị trung bình của hàm phạt trên mỗi cá thể. Các kết quả trong hình 2.3 cho thấy GA đạt trạng thái hội tụ kể từ thế hệ thứ 45. Trong kết quả hội tụ của GA có xuất hiện các vị trí bất thường vì ở giai đoạn đầu, quần thể có mức độ đa dạng cao (giá trị Best và Mean chênh lệch lớn). GA đang trong quá trình thực hiện khai phá không gian nghiệm (trên tập 3000 cá thể), đây là đặc tính được mong

muốn để tránh rơi sớm vào nghiệm cục bộ khi các toán tử lai ghép (crossover) và đột biến (mutation) trong GA được chọn với xác suất phù hợp đã hoạt động hiệu quả.



Hình 2.3: Sự hội tụ của GA trong phương pháp DPC



Hình 2.4: Độ ổn định và chính xác của GA trong phương pháp DPC

Ở trạng thái hội tụ, giá trị Mean trùng với Best, cho thấy toàn bộ các cá thể đều mang đặc tính di truyền tốt nhất. Đồng thời, giá trị Penalty giảm về 0, bảo đảm thỏa mãn các ràng buộc.

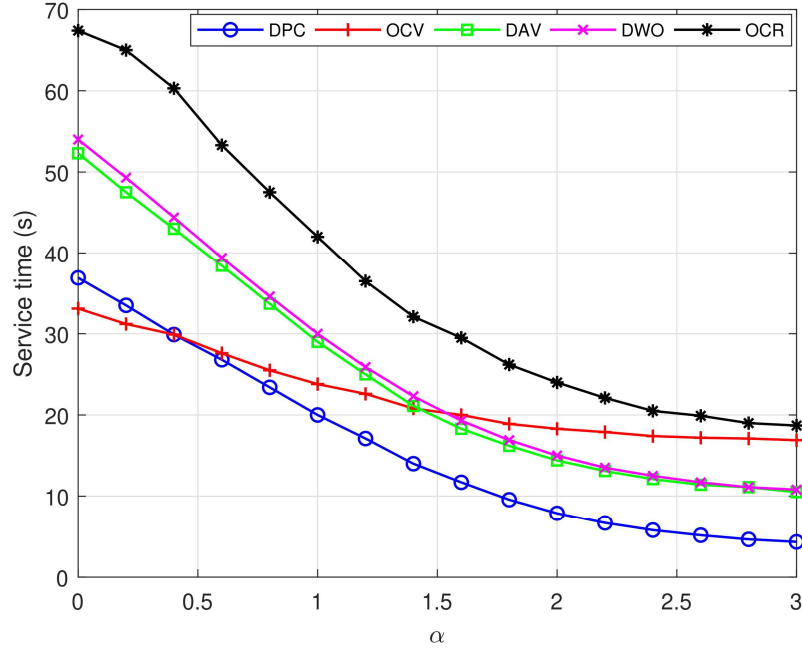
Trong hình 2.4, NCS tiếp tục khảo sát tính ổn định và độ chính xác của GA bằng cách thực hiện GA 100 lần với các kích thước quần thể khác nhau (N_p). Tính ổn định của GA được thể hiện trong hình 2.4(a) khi tăng N_p từ 1000 lên 3000 và 5000. Giá trị N_p càng lớn thì GA đạt được mức độ ổn định càng cao. Mức độ ổn định cao hơn mang lại nhiều khả năng xuất hiện các kết quả chính xác hơn, cụ thể là đạt $\bar{T} = 20s$ trong phạm vi 100 lần.

Độ chính xác của GA được phân tích trong hình 2.4(b) thông qua việc tính toán các giá trị lớn nhất (Max), trung bình (Ave), và nhỏ nhất (Min) trong 100 lần thực hiện GA đối với mỗi giá trị N_p . Có thể nhận thấy rằng giá trị Min ($\bar{T} = 20s$), tức nghiệm chính xác của GA, luôn xuất hiện trong 100 lần ngay cả khi kích thước quần thể rất nhỏ ($N_p = 500$). Việc tăng N_p giúp cải thiện độ chính xác của kết quả, qua đó làm giảm các giá trị Ave và Max. Tuy nhiên, GA không thể triển khai với kích thước quần thể quá lớn do độ phức tạp cao về thời gian và bộ nhớ. Trong mô hình DPC, giá trị N_p phù hợp được lựa chọn là 3000 tại điểm “khuyết” (elbow point) của đường Ave, nhờ đó GA không chỉ đạt được tính ổn định và độ chính xác cao mà còn duy trì được mức độ phức tạp tính toán hợp lý.

2.5.3. *Đánh giá hiệu suất DPC*

Trong hệ thống này, OCV đóng vai trò quan trọng hơn trong giảm thời gian phục vụ và hiệu quả hơn so với phương án OCR. Khi α quá nhỏ ($\alpha \leq 0.4$), tức là tất cả các video có mức độ phổ biến giống nhau hoặc gần tương đương, phương pháp DPC (là sự kết hợp giữa OCV và OCR) khó có thể vượt trội hơn phương án OCV thuần túy, do hiệu năng của OCR trong trường hợp này là rất kém. Thậm chí, DPC không thể đạt hiệu năng tương đương với OCV, bởi như trong hình 2.9, GA gần như loại bỏ hoàn toàn vai trò của OCR kém hiệu quả này khỏi DPC bằng cách ép mức dung lượng lưu trữ của DPC-CR (S^{CR}) về rất thấp và tăng mức lưu trữ của DPC-CV (S^{CV}) lên gần bằng với OCV. Do đó, để thuận tiện cho việc triển khai, các nhà thiết kế hệ thống có xu hướng

ưu tiên OCV hơn DPC khi α ở mức thấp. Tuy nhiên, khi α đủ lớn ($\alpha > 0.4$), sự kết hợp giữa OCV và OCR hoạt động hiệu quả, giúp DPC trở nên hiệu quả hơn và vượt trội hơn so với các phương án khác. Nhờ các ưu điểm của DPC, hai phương án DAV và DWO đều đạt được thời gian phục vụ thấp hơn so với OCV và OCR.

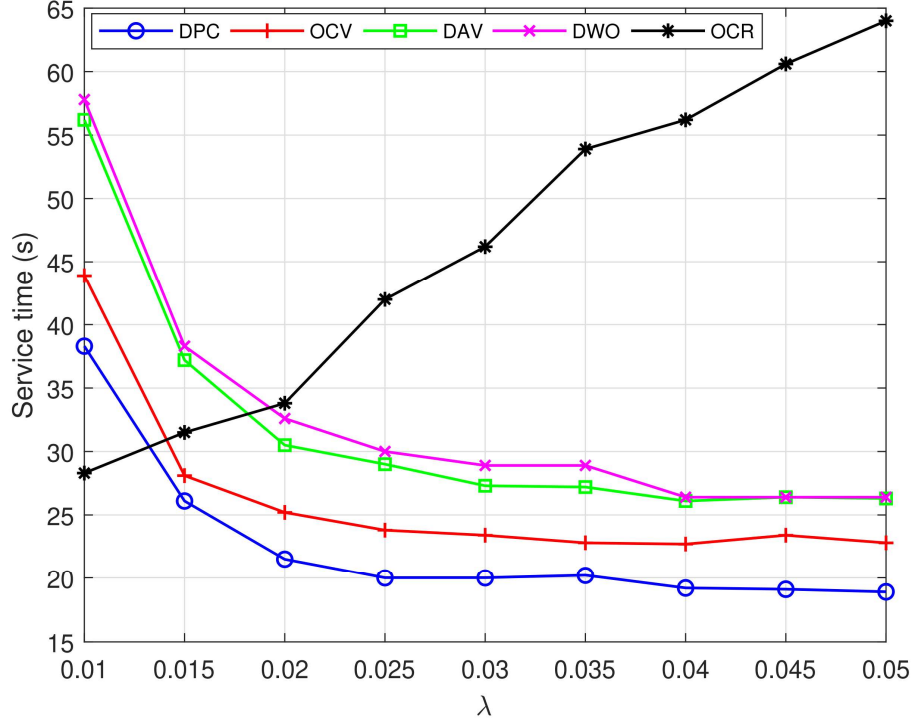


Hình 2.5: Thời gian phục vụ theo α

Cơ chế cải thiện hiệu năng của DPC nằm ở việc GA cân đối tỷ lệ S^{CR}/S^{CV} theo độ lệch phân bố Zipf α . Khi α nhỏ ($\alpha \leq 0.4$), các video có độ phổ biến gần đều nhau, GA phân bổ phần lớn dung lượng lưu trữ cho tầng CV (DPC tiệm cận OCV). Khi α lớn ($\alpha > 0.4$), một số video phổ biến nổi trội rõ rệt, GA tăng đầu tư vào tầng RSU để phục vụ những video này qua kênh R2V có dung lượng cao hơn V2V. Cơ chế cân đối này lý giải vì sao DPC chỉ vượt trội rõ rệt so với OCV và OCR khi $\alpha > 0.4$ (hình 2.5).

Trong hình 2.6, hiệu năng của hệ thống được khảo sát theo mật độ phương tiện λ . Đối với các phương án DPC, OCV, DAV và DWO, chúng phụ thuộc một phần hoặc hoàn toàn vào lưu trữ xác suất, do đó thời gian phục vụ giảm dần khi λ tăng. Nguyên nhân là do việc tăng λ cho phép hệ thống lưu trữ các video tại các xe lưu trữ hiệu quả hơn, từ đó phục vụ các RV nhanh hơn. Thời gian phục vụ giảm xuống đến một giá trị bão hòa nhất định do các giới hạn được xác định bởi (2.3) và (2.4). Khi λ thấp ($\lambda \leq 0.0125$), các phương pháp DPC, OCV, DAV và DWO dựa trên PC đều cho hiệu năng

kém, thậm chí còn thấp hơn so với OCR – một giải pháp thay thế phù hợp cho kịch bản có mật độ phương tiện thấp. Ngược lại, đối với OCR, thời gian phục vụ tăng lên khi λ tăng, do dung lượng kênh bị chia sẻ theo số lượng RV được tính trong (2.10). Khi λ đủ lớn, có thể kết luận rằng DPC luôn vượt trội hơn các phương án còn lại.



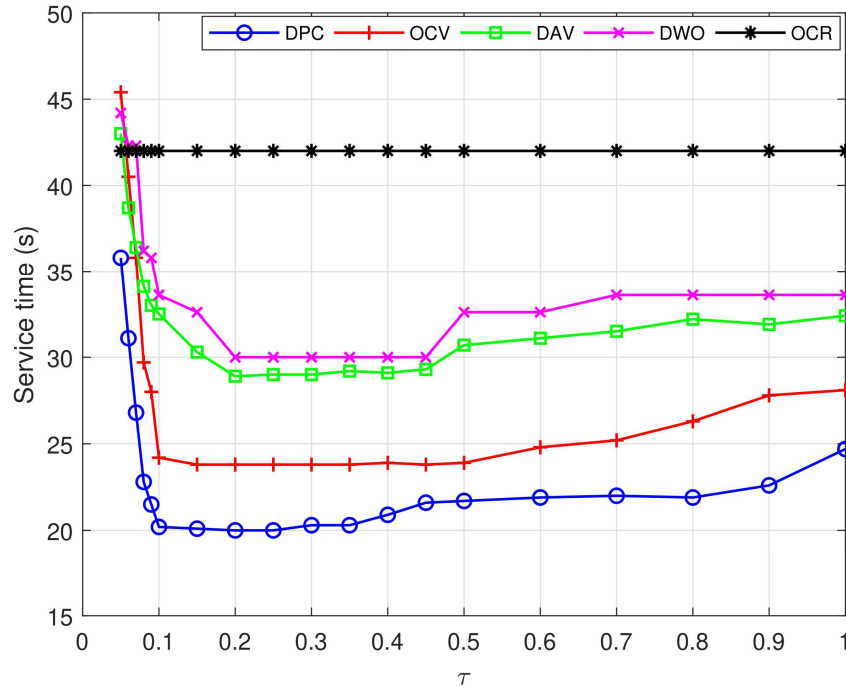
Hình 2.6: Thời gian phục vụ theo λ

Khi λ tăng, mật độ CV theo công thức $\lambda_i^{CV} = \tau \rho_i \lambda$ tăng tuyến tính, kéo theo xác suất tồn tại CV trong phạm vi truyền (công thức 2.3) tiệm cận đến 1. Cơ chế này lý giải vì sao thời gian phục vụ của các phương pháp dùng lưu trữ xác suất (OCV, DPC, DAV, DWO) giảm theo λ , trong khi OCR, chỉ dùng lưu trữ tại RSU, lại tăng theo λ do dung lượng R2V trên mỗi RV bị chia nhỏ.

Hình 2.7 thể hiện mối quan hệ giữa thời gian phục vụ và tham số τ , tức là tỷ lệ các phương tiện sẵn sàng đóng vai trò là CV. Có thể nhận thấy rằng OCR thuần túy dựa trên lưu trữ xác định giữ nguyên hiệu năng không đổi, do cả tử số và mẫu số trong (2.10) cùng giảm đồng thời khi τ tăng. Trong khi đó, các phương pháp khác dựa trên lưu trữ xác suất, bao gồm DPC, OCV, DAV và DWO, cho thấy rằng tồn tại một khoảng giá trị tối ưu của τ nhằm đạt được thời gian phục vụ nhỏ nhất. Khoảng giá trị tối ưu của τ được

xác định khi các biểu thức (2.3) và (2.4) được cân bằng một cách tối ưu để cực đại (2.7), hoặc tương đương là cực tiểu thời gian phục vụ. Phát hiện thú vị này cung cấp một phương pháp hữu ích để xác định tỷ lệ tối ưu các phương tiện đóng vai trò là CV, nhằm phục vụ các RV với thời gian phục vụ nhanh nhất. So sánh cho thấy, phương pháp DPC được đề xuất đạt hiệu năng tốt nhất về thời gian phục vụ khi τ đủ lớn ($\tau \geq 0.05$). Ngược lại, khi τ nhỏ ($\tau < 0.05$), DPC có xu hướng tiệm cận với OCR, trong khi OCV, DAV và DWO còn lại trở nên kém hơn OCR.

Sự tồn tại điểm tối ưu τ^* trong hình 2.7 bắt nguồn từ cấu trúc tích $p_{i,n}^{CV} \cdot p_{i,n}^{RV}$ trong công thức (2.7). Hai xác suất này biến đổi ngược chiều theo τ : khi τ tăng thì $p_{i,n}^{CV}$ tăng nhưng $p_{i,n}^{RV}$ giảm, vì một xe không thể đồng thời đóng vai trò CV và RV. Tích của hai hàm đơn điệu ngược chiều đạt cực đại tại một giá trị τ^* nội tại của bài toán. Với cấu hình tham số đã chọn, τ^* nằm trong khoảng $[0.05, 0.07]$ và DPC đạt hiệu năng tốt nhất tại vùng này. Khi $\tau < 0.05$, tầng CV gần như không đóng góp và DPC suy biến về OCR.

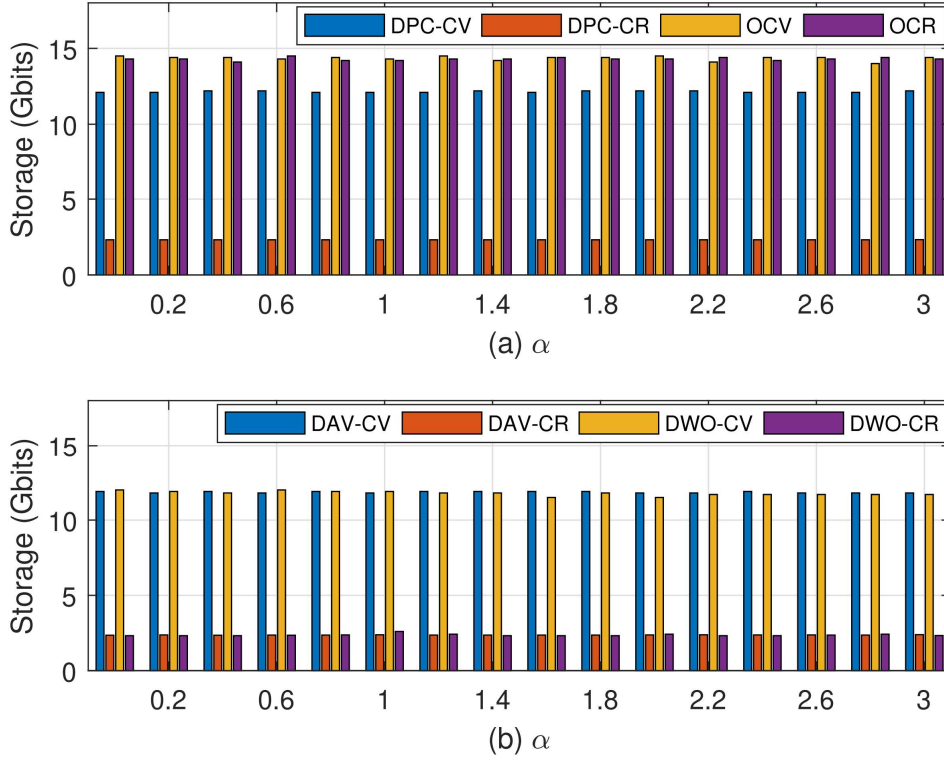


Hình 2.7: Thời gian phục vụ theo τ

2) Dung lượng lưu trữ

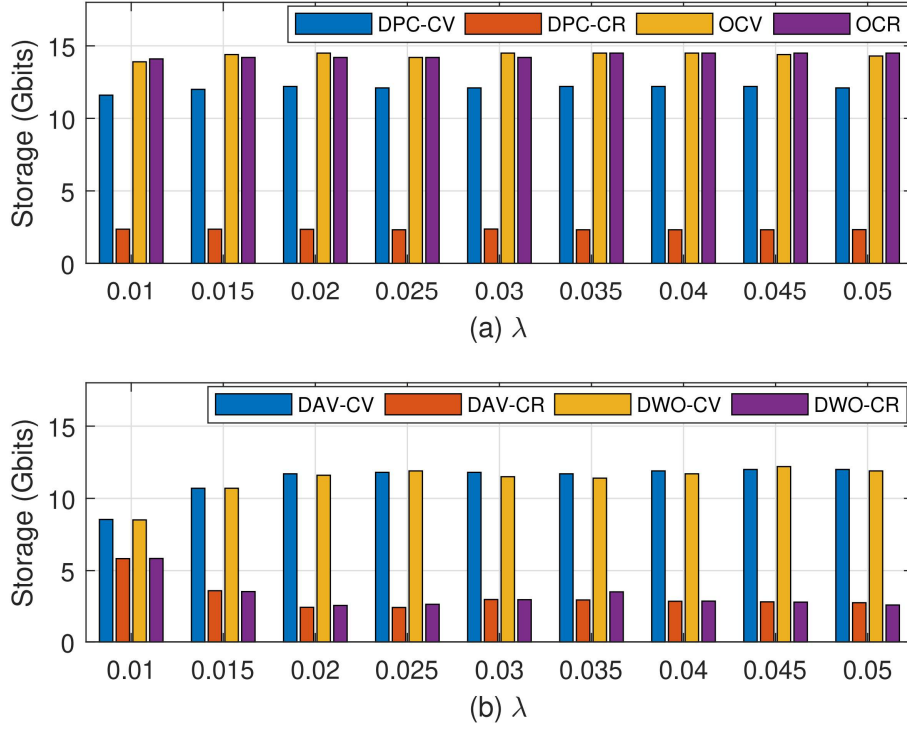
Để đánh giá dung lượng lưu trữ, tổng dung lượng lưu trữ được sử dụng bởi phương pháp DPC – bao gồm lưu trữ tại các xe lưu trữ (DPC-CV) và lưu trữ tại các

trạm phát ven đường (DPC-CR), được các phương pháp OCV, OCR, DAV và DWO sử dụng tương đương nhằm bảo đảm tính công bằng trong so sánh. Cần lưu ý rằng, do các phương án DAV và DWO được xây dựng dựa trên phương pháp DPC, tổng mức tiêu thụ lưu trữ của DAV (hoặc DWO) cũng tương tự, bao gồm lưu trữ tại DAV-CV và DAV-CR (hoặc DWO-CV và DWO-CR).



Hình 2.8: Dung lượng lưu trữ theo α

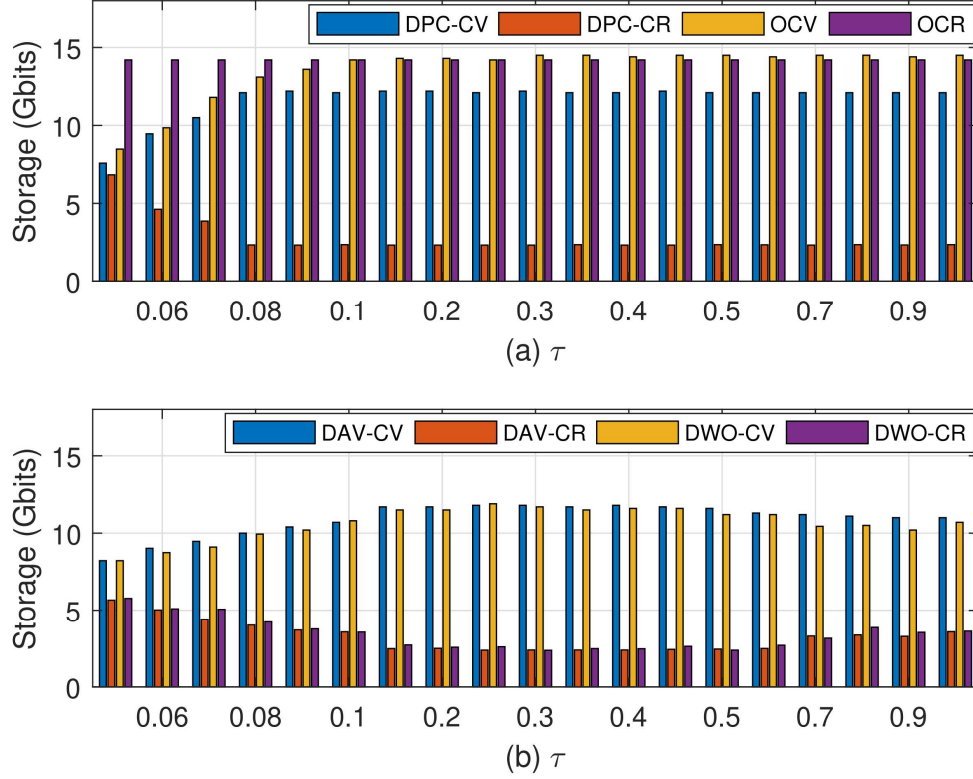
Hình 2.8 thể hiện dung lượng lưu trữ theo hệ số α . Trong hình 2.8(a), dung lượng lưu trữ được sử dụng bởi DPC (bao gồm DPC-CV và DPC-CR), OCV và OCR là tương đương nhau. Điều này cũng xảy ra đối với DAV và DWO theo cùng cách thức như DPC, như được thể hiện trong hình 2.8(b). Trong bối cảnh này, lợi thế của DPC không đến từ việc sử dụng nhiều dung lượng lưu trữ hơn, mà đến từ cách phân bổ dung lượng lưu trữ giữa các tầng lưu trữ khác nhau trong hệ thống. Nhờ cơ chế tối ưu đồng thời giữa lưu trữ tại RSU và lưu trữ xác suất trên các phương tiện, DPC có thể khai thác hiệu quả hơn tài nguyên lưu trữ sẵn có để rút ngắn thời gian phục vụ. Điều này cho thấy đóng góp quan trọng của DPC nằm ở chiến lược phân bổ lưu trữ hợp lý, thay vì chỉ đơn thuần tăng dung lượng lưu trữ của hệ thống.



Hình 2.9: Dung lượng lưu trữ theo λ

Tương tự, hình 2.9 thể hiện dung lượng lưu trữ theo mật độ xe λ . Có thể nhận thấy rằng dung lượng lưu trữ được sử dụng bởi tất cả các phương pháp DPC, OCV, OCR, DAV và DWO vẫn là tương đương, nhằm bảo đảm tính công bằng trong so sánh. Đáng chú ý, trong hình 2.9(a), DPC phân bổ tối ưu một tỷ lệ nhất quán của dung lượng lưu trữ giữa DPC-CV và DPC-CR, bất chấp λ tăng lên, nhằm đạt được thời gian phục vụ nhanh nhất. Trong khi đó, ở hình 2.9, dung lượng lưu trữ được DAV và DWO sử dụng thay đổi một cách tự nhiên mà không có bất kỳ cơ chế phân bổ lưu trữ tối ưu nào. DAV và DWO có thể khai thác nhiều dung lượng lưu trữ sẵn có hơn tại các CV khi λ tăng, nhằm cải thiện thời gian phục vụ bằng cách tăng mức tiêu thụ lưu trữ tại DAV-CV và DWO-CV, đồng thời giảm mức tiêu thụ tại DAV-CR và DWO-CR.

DPC giữ tỷ lệ S^{CR}/S^{CV} gần như không đổi khi λ tăng (hình 2.9(a)), trong khi DAV và DWO tăng S^{CV} và giảm S^{CR} theo λ (hình 2.9(b)). Khác biệt này phản ánh nguyên tắc lựa chọn nghiệm: DPC chọn cá thể có thời gian phục vụ nhỏ nhất nên cấu hình tối ưu được duy trì, còn DAV/DWO chọn cá thể trung bình/xấu nhất nên dung lượng lưu trữ trôi theo phân bố ngẫu nhiên của quần thể GA.



Hình 2.10: Dung lượng lưu trữ theo τ

Hình 2.10 cho thấy cách DPC phân bổ dung lượng lưu trữ giữa hai tầng CV và RSU theo τ . Trong hình 2.10(a), DPC tăng S^{CV} và giảm S^{CR} tương ứng khi τ tăng, để duy trì tổng tài nguyên không đổi theo ràng buộc (2.14c). Cơ chế này phản ánh chiến lược của GA: khi $\tau < 0.05$, mật độ CV không đủ để khai thác qua kênh V2V, GA dồn dung lượng lưu trữ về RSU, làm cho dung lượng tại lớp CV gần như bằng 0 và DPC suy biến về OCR; khi $0.05 \leq \tau \leq 0.07$, đây là vùng tối ưu τ^* , GA cân đối tỷ lệ S^{CV}/S^{CR} sao cho thời gian phục vụ đạt giá trị nhỏ nhất; khi $\tau > 0.07$, mật độ CV đủ lớn, GA chuyển dần lưu trữ sang S^{CV} để khai thác lợi thế của kênh V2V. Hình 2.10(b) cho thấy DAV và DWO có xu hướng phân bổ tương tự DPC nhưng với mức S^{CV} thấp hơn do hai phương pháp này không chọn cá thể tốt nhất từ quần thể GA. Khoảng cách giữa S^{CV} của DPC và của DAV/DWO mở rộng theo τ là minh chứng cho hiệu quả của chiến lược chọn nghiệm trong DPC.

2.5.4. Độ phức tạp của mô hình DPC

Trong mô hình DPC, độ phức tạp tính toán của bài toán tối ưu phụ thuộc chủ yếu vào quy mô của hệ thống, bao gồm số lượng RSU N và video I . Trong thuật toán 2.1,

giá trị của I quyết định số lần tính toán t_i và giá trị N quyết định độ lớn của J_i theo công thức (1.2). Do đó, độ phức tạp hay chi phí tính toán hàm mục tiêu C_0 của bài toán tỷ lệ với $I \times \max\{J_i\}$ và trong tình huống xấu nhất sẽ là $I \times N$.

Nếu giải bài toán bằng phương pháp vét cạn, việc giải sẽ trở nên không khả thi khi quy mô hệ thống lớn. Luận án sử dụng GA để tìm nghiệm gần tối ưu với chi phí tính toán chấp nhận được. Trong GA, độ phức tạp tính toán chủ yếu phụ thuộc vào kích thước quần thể N_p , số thế hệ N_G và C_0 . Như vậy, độ phức tạp tổng thể của GA trong mô hình DPC có thể được biểu diễn xấp xỉ theo bậc $O(N_p \times N_G \times C_0)$.

Tuy nhiên độ phức tạp của GA không chỉ phụ thuộc vào N_p , N_G và chi phí đánh giá hàm mục tiêu C_0 , mà còn phụ thuộc vào kích thước không gian nghiệm. Trong mô hình DPC, khi số lượng video I hoặc số lượng RSU N tăng, số biến quyết định h_i và ρ_i tăng tương ứng, đồng thời số khả năng lựa chọn vị trí lưu trữ tại RSU cũng mở rộng. Nếu xét mỗi video có xấp xỉ N khả năng lựa chọn cho biến h_i , và biến ρ_i được lượng tử hóa bằng P_R bit. Như vậy độ phức tạp của GA được xác định bởi công thức $O(N_p \times N_G \times I \times N)$ không gian nghiệm tăng theo số mũ khi số lượng video I tăng. Do đó, phương pháp vét cạn không khả thi đối với các hệ thống có quy mô lớn. GA được sử dụng nhằm tìm nghiệm gần tối ưu trong một số lượng thế hệ hữu hạn, thay vì duyệt toàn bộ không gian nghiệm. Tuy nhiên, GA vẫn có chi phí tính toán đáng kể, đặc biệt khi cần tăng kích thước quần thể N_p hoặc số thế hệ N_G để cải thiện độ ổn định và chất lượng nghiệm.

Mật độ phương tiện λ không làm tăng trực tiếp chiều dài nhiệm vụ sắc thể trong mô hình giải tích của DPC, vì λ được đưa vào các biểu thức xác suất thông qua mô hình Poisson. Tuy nhiên, khi λ thay đổi, xác suất tồn tại phương tiện lưu trữ và giá trị thời gian phục vụ cũng thay đổi, từ đó ảnh hưởng đến giá trị hàm mục tiêu và nghiệm tối ưu. Trong trường hợp triển khai mô phỏng chi tiết theo từng phương tiện hoặc sử dụng dữ liệu vết di chuyển thực tế, số phương tiện kỳ vọng trên tuyến đường sẽ tăng theo λ , làm tăng thêm chi phí cập nhật trạng thái và đánh giá hệ thống.

Vì vậy, mô hình DPC phù hợp hơn với cơ chế tối ưu ngoại tuyến hoặc tái tối ưu định kỳ tại MBS/MEC dựa trên thống kê lưu lượng, độ phổ biến nội dung và trạng thái

phương tiện, thay vì thực hiện tối ưu lại cho từng yêu cầu video riêng lẻ. Trong các kịch bản VANET quy mô lớn, chi phí tính toán có thể được giảm bằng cách phân cụm RSU theo khu vực, chỉ tối ưu nhóm video có độ phổ biến cao, khởi tạo GA bằng nghiệm heuristic hoặc song song hóa quá trình đánh giá hàm thích nghi của các cá thể trong quần thể.

Kết quả này cho thấy độ phức tạp của thuật toán tăng tuyến tính theo quy mô hệ thống. Mặc dù không gian tìm kiếm lý thuyết tăng theo hàm mũ, GA cho phép khám phá hiệu quả trong không gian nghiệm phù hợp và tìm được nghiệm toàn cục xấp xỉ tối ưu trong thời gian tính toán hợp lý. Mô hình DPC vẫn có khả năng mở rộng tốt khi áp dụng cho các hệ thống VANET với quy mô lớn.

2.6. KẾT LUẬN CHƯƠNG 2

Trong chương này, NCS đã đề xuất mô hình lưu trữ kết hợp xác định và xác suất (DPC) cho các dịch vụ video (VAS) trong VANET. Mô hình DPC khai thác các ưu điểm của lưu trữ xác định (DC) và lưu trữ xác suất (PC) để lưu trữ video lần lượt tại các trạm phát ven đường cố định và các xe có tính động (VU). Sự kết hợp giữa DC và PC cho phép thực hiện truyền video hợp tác từ các RSU và các phương tiện lưu trữ (CV) nhằm phục vụ các phương tiện yêu cầu (RV) một cách hiệu quả.

Bên cạnh đó, nhiều đặc trưng của RSU (số lượng), phương tiện (mật độ và vận tốc), video (mức độ phổ biến và kích thước), cũng như các kỹ thuật truyền hợp tác (CoT) và GA đã được xem xét để xây dựng bài toán tối ưu DPC nhằm xác định bán kính lưu trữ và xác suất lưu trữ tối ưu. Tổng dung lượng lưu trữ được sử dụng bởi các RSU và các CV được đưa vào như một ràng buộc chung trong bài toán tối ưu DPC, qua đó đảm bảo sự cân bằng hợp lý giữa lượng lưu trữ được tiêu thụ bởi DC và PC, tùy thuộc vào điều kiện cụ thể của VANET. Các phát hiện chính được thảo luận thêm nhằm cung cấp những hiểu biết sâu hơn cho việc thiết kế các kỹ thuật lưu trữ biên phục vụ các dịch vụ video trong VANET.

GA được điều chỉnh để xử lý tính phức tạp của các biến tối ưu dạng số nguyên và số thực trong bài toán DPC. Do đó, các kết quả mô phỏng được trình bày nhằm chứng minh khả năng hội tụ, tính ổn định và độ chính xác của GA, cũng như những lợi ích của

phương pháp DPC xét theo tiêu chí tối thiểu hóa thời gian phục vụ trong khi vẫn sử dụng hiệu quả tài nguyên lưu trữ.

Mặc dù DPC cải thiện thời gian phục vụ và khai thác hiệu quả tài nguyên lưu trữ tại RSU và phương tiện, mô hình vẫn giới hạn trong kiến trúc mặt đất và phụ thuộc vào mật độ, khả năng lưu trữ cũng như kết nối V2V của phương tiện. Đồng thời, các RSU có vị trí cố định nên chưa thể chủ động bù đắp các khoảng trống lưu trữ và vùng phủ. Vì vậy, Chương 3 tiếp tục nghiên cứu việc sử dụng UAV như một nút lưu trữ–truyền dẫn linh hoạt và tối ưu vị trí phục vụ của UAV.

CHƯƠNG 3: KẾT HỢP LƯU TRỮ VÀ XÁC ĐỊNH VỊ TRÍ PHỤC VỤ TỐI ƯU TRONG TRUYỀN VIDEO HỢP TÁC QUA UAV-VANET

Tóm tắt: Trong Chương 3, bài toán lưu trữ tại RSU (Roadside Unit) trong VANET (Vehicular Ad-hoc Network) được thực hiện tối ưu với sự hỗ trợ bởi UAV (Unmanned Aerial Vehicle) như các nút lưu trữ và truyền dẫn linh hoạt cho dịch vụ truyền video hợp tác trong mạng UAV-VANET. Cụ thể, mô hình tối ưu kết hợp giữa lưu trữ và vị trí dừng CAHO (CAching and Hovering Optimization) được đề xuất, trong đó, xem xét đồng thời ba biến tối ưu: (i) bán kính lưu trữ video tại các RSU (Roadside Unit), (ii) quyết định lưu trữ video tại UAV, và (iii) vị trí dừng của UAV, nhằm bù đắp các khoảng trống lưu trữ (caching gap) giữa các RSU trong môi trường VANET có tính linh động cao và phân bố phương tiện (VU – Vehicular User) không đồng đều. Trên cơ sở mô hình hóa phân bố phương tiện bằng quá trình Poisson một chiều, kết hợp các mô hình truyền R2V (RSU to Vehicle) và U2V (UAV to Vehicle) cùng với đặc tính độ phổ biến của video, bài toán tối ưu CAHO được xây dựng nhằm cực tiểu thời gian phục vụ trung bình, dưới ràng buộc nghiêm ngặt về tài nguyên lưu trữ và số lượng UAV khả dụng.

Để giải bài toán tối ưu có tính phi tuyến, rời rạc và đa ràng buộc này, NCS áp dụng GA kết hợp với hàm phạt và chiến lược chia để trị, cho phép xử lý hiệu quả đồng thời các biến tối ưu nguyên và nhị phân. Kết quả mô phỏng cho thấy CAHO đạt thời gian phục vụ thấp nhất trong nhóm năm phương pháp (CAHO, RAHO, RACR, RACU, NUAV) trên cả ba tham số α , M và λ , trong điều kiện sử dụng cùng mức tài nguyên lưu trữ. So với mô hình DPC ở Chương 2, CAHO bổ sung tầng lưu trữ UAV vào không gian ba chiều và đặt quyết định vị trí dừng UAV thành biến tối ưu thay vì cố định. Mô hình DPC đã được công bố trong bài báo [C3].

3.1. GIỚI THIỆU

Sự phát triển nhanh chóng của các ứng dụng và dịch vụ tiên tiến (A&S – Advanced Applications and Services) đặt ra những yêu cầu ngày càng khắt khe đối với hạ tầng mạng không dây, đặc biệt trong các hệ thống giao thông thông minh dựa trên

VANET (Vehicular Ad-hoc Network). Sự bùng nổ của Internet of Things (IoT), xe tự hành và các thiết bị thông minh, cùng với sự phổ biến của các A&S đa phương tiện, đã khiến lưu lượng dữ liệu di động tăng trưởng theo cấp số nhân. Một báo cáo thống kê trong [26] ước tính rằng khoảng 300 triệu A&S sẽ phục vụ 12.3 tỷ người dùng di động (MU – Mobile User), trong đó lưu lượng video chiếm tỷ trọng áp đảo, lên tới 79% tổng lưu lượng dữ liệu di động [25]. Do đó, nhu cầu về chất lượng trải nghiệm (QoE – Quality of Experience) của người dùng ngày càng gia tăng mạnh mẽ.

Trong tình huống này, việc tích hợp VANET với các kỹ thuật điện toán biên được xem là một hướng tiếp cận đầy hứa hẹn nhằm đưa nội dung và tài nguyên tính toán đến gần hơn với người dùng phương tiện (VU – Vehicular User), từ đó giảm tải cho mạng lõi và rút ngắn độ trễ đầu cuối. Các kỹ thuật lưu trữ nội dung tại biên đóng vai trò then chốt trong việc hỗ trợ các dịch vụ nhạy cảm với độ trễ, chẳng hạn như dịch vụ truyền video, bằng cách lưu trữ các nội dung phổ biến tại biên mạng để đồng thời giảm độ trễ phục vụ và mức tiêu thụ năng lượng [100]. Một nghiên cứu khác trong [101] xem xét môi trường có tính động cao, đặc trưng bởi phân bố không đồng đều và mật độ cao của VU, trong đó các kỹ thuật lưu trữ và offloading triển khai tại các RSU (Roadside Unit) gặp khó khăn trong việc đảm bảo chất lượng dịch vụ (QoS – Quality of Service) trước các yêu cầu nội dung đa dạng và quy mô lớn. Để giải quyết vấn đề này, các tác giả đã đề xuất một khung tối ưu kết hợp giảm tải (offloading) tính toán và lưu trữ nội dung, nhằm cực tiểu hóa độ trễ tổng thể của mạng trong khi vẫn thỏa mãn các ràng buộc năng lượng dài hạn.

Để nâng cao hơn nữa hiệu quả của các kỹ thuật điện toán biên trong VANET, nhiều nghiên cứu gần đây đã đề xuất các chiến lược hợp tác, khai thác đồng thời tài nguyên lưu trữ và truyền thông tại RSU, trạm gốc MBS (Macro Base Station) và các VU, nhằm giảm tắc nghẽn mạng và độ trễ truyền dẫn [27], [31], [34]. Cụ thể, trong [34], các tác giả đề xuất một chiến lược lưu trữ cho VANET dựa trên vị trí và độ phổ biến của nội dung, trong đó, mỗi nội dung có thể được lưu trữ tại cả MBS và RSU, chỉ tại MBS, hoặc chỉ tại RSU, với mục tiêu cực tiểu độ trễ truyền và chi phí dịch vụ tổng thể. Nghiên cứu trong [27] kết hợp lưu trữ tại RSU với lập lịch truyền quảng bá và điều chỉnh tốc độ truyền, nhằm tối đa lưu lượng được giảm tải. Trong [31], một chiến lược

lưu trữ xác định và xác suất kết hợp được đề xuất để triển khai đồng thời tại RSU và VU. Bên cạnh đó, một khung truyền thông hợp tác cho phép truyền video từ RSU và VU đến các VU khác cũng được phát triển, nhằm đạt được thời gian phục vụ tối thiểu trong khi vẫn sử dụng hiệu quả tài nguyên lưu trữ của cả RSU và VU.

Gần đây, UAV đã được khai thác như một thành phần hỗ trợ hiệu quả cho VANET, nhờ vào khả năng lựa chọn vị trí dừng linh hoạt với xác suất đường truyền nhìn thẳng (LoS – Line of Sight) cao, vùng phủ rộng, và khả năng triển khai nhanh chóng. UAV có thể hoạt động như trạm gốc trên không hoặc nút lưu trữ di động, hỗ trợ VANET trong việc phân phối nội dung tại các khu vực mà RSU bị quá tải hoặc hạ tầng mặt đất không đáp ứng được [7], [102], [103]. Theo hướng tiếp cận này, nghiên cứu trong [104] triển khai UAV với băng thông hạn chế để hỗ trợ VANET, trong đó, nội dung được phân phối giữa các VU thông qua lưu trữ hợp tác dựa trên mã hóa mạng, nhằm cải thiện hiệu quả sử dụng băng thông. Trong [105], UAV được sử dụng để lưu trữ nội dung và phục vụ các VU có tính động, với bài toán lập kế hoạch quỹ đạo và quản lý bộ nhớ đệm được giải quyết đồng thời nhằm nâng cao lượng dữ liệu tải về và hiệu quả năng lượng. Ngoài ra, nghiên cứu trong [106] xem xét lưu trữ nội dung tại cả UAV và VU, bằng cách kết hợp sở thích người dùng với các cơ chế lưu trữ và thay thế dựa trên khuyến khích, từ đó giúp giảm độ trễ yêu cầu và mức tiêu thụ năng lượng. Tuy nhiên, các nghiên cứu hiện có về UAV-VANET vẫn chưa giải quyết đầy đủ bài toán lưu trữ nội dung kết hợp tại cả RSU và UAV, cũng như xác định vị trí dừng của UAV tối ưu, nhằm hỗ trợ dịch vụ truyền video hợp tác với mục tiêu giảm thiểu thời gian phục vụ trong khi vẫn tiết kiệm tài nguyên lưu trữ.

Xuất phát từ những hạn chế trên, trong chương này, bằng cách mô hình hóa phân bố VU thông qua quá trình Poisson đồng nhất một chiều (T1D-PPP - T1D Poisson Point Process), kết hợp các mô hình truyền R2V (RSU to Vehicle) và U2V (UAV to Vehicle) cùng với đặc tính độ phổ biến của video, một thiết kế tối ưu lưu trữ và vị trí dừng – CAHO (CAching and Hovering Optimization) được đề xuất cho các dịch vụ truyền video hợp tác. Cụ thể, một bài toán tối ưu CAHO được xây dựng và giải để xác định những video được lưu trữ tại RSU và những video được lưu trữ tại UAV kèm theo vị trí dừng (hovering) tương ứng, nhằm bù đắp các khoảng trống lưu trữ giữa các RSU. Mục

tiêu là cực tiểu hóa thời gian phục vụ và đảm bảo các phiên truyền video liên tục trong điều kiện tài nguyên lưu trữ hạn chế tại cả RSU và UAV. Bài toán tối ưu CAHO được giải bằng GA (Genetic Algorithm) kết hợp với phương pháp hàm phạt và chiến lược chia để trị áp dụng cho từng cá thể được mã hóa bằng chuỗi nhiễm sắc thể. Các kết quả mô phỏng cho thấy phương pháp CAHO vượt trội so với các phương pháp so sánh khác trong nhiều kịch bản hệ thống khác nhau về thời gian phục vụ và hiệu quả sử dụng tài nguyên lưu trữ, qua đó khẳng định tính hiệu quả và tiềm năng ứng dụng thực tiễn của giải pháp cho các dịch vụ truyền video trong UAV-VANET.

Chương 3 kế thừa khung mô hình và tiêu chí tối thiểu hóa thời gian phục vụ đã được xây dựng trong Chương 2, đồng thời mở rộng từ VANET mặt đất sang UAV-VANET. Mô hình CAHO tiếp tục xem xét lưu trữ xác định tại RSU và phương pháp giải bằng GA, nhưng bổ sung quyết định lưu trữ tại UAV, vị trí dừng của UAV và liên kết U2V nhằm khắc phục những hạn chế của mô hình DPC.

3.2. MÔ HÌNH HỆ THỐNG

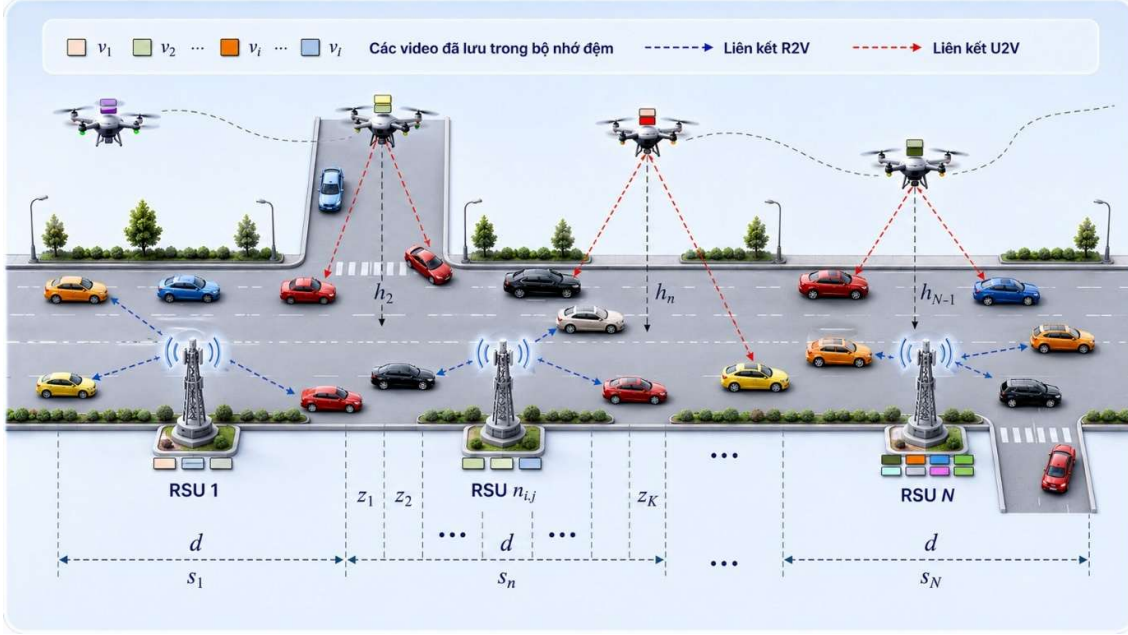
Trong nghiên cứu này, mô hình CAHO được triển khai cho dịch vụ truyền video hợp tác trong mạng UAV-VANET như trình bày trong hình 3.1. Các ký hiệu chính được liệt kê trong bảng 3.1. Mô hình hệ thống bao gồm N RSU được hỗ trợ bởi M UAV để phục vụ các ứng dụng và dịch vụ truyền video. Hệ thống xét I video, trong đó mỗi video i có độ phổ biến p_i và dung lượng S_i .

Các VU được phân bố không gian theo quá trình Poisson một chiều với mật độ λ . Mỗi RSU bao phủ một đoạn đường có chiều dài d , được chia thành K vùng. Nội dung video có thể được lưu trữ đồng thời tại RSU và UAV. Các VU di chuyển vào đoạn đường thứ n , với $n = 1, 2, \dots, N$, và có vận tốc trung bình là s_n km/h. Trong các VU, một tập con được xác định là xe yêu cầu (RV – Requesting Vehicle), tức là những phương tiện có yêu cầu tải video.

Trong mô hình này, nhằm cải thiện QoS cho các ứng dụng truyền video trong VANET, bài toán tối ưu CAHO được xây dựng và giải để xác định:

- (i) quyết định lưu trữ video i tại các RSU thông qua bán kính lưu trữ h_i ;

- (ii) quyết định lưu trữ video i tại UAV dừng ở đoạn đường n , được biểu diễn bởi biến nhị phân $c_{n,i}$;
- (iii) vị trí dừng của UAV được xác định bởi chỉ số dừng b_n ở đoạn đường n với độ cao h_n .



Hình 3.1: Mô hình CAHO cho truyền video trong UAV-VANET

Mục tiêu của bài toán là cực tiểu hóa thời gian phục vụ trung bình, đồng thời sử dụng hiệu quả tài nguyên lưu trữ tại cả RSU và UAV. Thời gian phục vụ được định nghĩa là khoảng thời gian trung bình để các RV nhận được video yêu cầu từ RSU và UAV trong quá trình di chuyển.

Sau khi triển khai chiến lược CAHO, khi một RV yêu cầu video i và đi vào một đoạn đường, phương tiện đó sẽ được phục vụ bởi RSU hoặc UAV đã lưu trữ video tương ứng, nhằm đảm bảo phiên truyền video liên tục. Cần lưu ý rằng, trong trường hợp video i có kích thước lớn và không thể truyền hết khi RV đi qua toàn bộ N RSU, quá trình truyền sẽ tiếp tục thông qua một tập RSU khác được kết nối với một MBS khác. Kịch bản này vượt ra ngoài phạm vi của nghiên cứu trong chương và được dành cho các nghiên cứu trong tương lai. Các mô hình toán học chi tiết cho thiết kế và tối ưu CAHO sẽ được trình bày trong các mục tiếp theo.

3.3. TÍNH TOÁN THÔNG SỐ HỆ THỐNG

3.3.1. Mô hình T1D-PPP cho RV

Xét mô hình T1D-PPP mô tả phân bố không gian của các VU trên các tuyến đường một chiều [89], [90]. Gọi p_i là độ phổ biến của video i , khi đó phân bố của các RV đối với video i cũng tuân theo một quá trình Poisson một chiều độc lập, ký hiệu là Φ_i^{RV} với mật độ λ_i^{RV} , được tính như sau:

$$\lambda_i^{RV} = p_i \lambda, \quad (3.1)$$

Trong đó, độ phổ biến của video i được mô hình hóa bằng phân bố Zipf-like [46], như công thức (1.1).

3.3.2. Lưu trữ tại RSU và UAV

1) Lưu trữ tại RSU

Đối với lưu trữ tại RSU, bán kính lưu trữ h_i của video i được xác định rằng video này sẽ được lưu trữ sau mỗi h_i RSU. Khi đó, các RSU lưu trữ video i được xác định theo chỉ số:

$$n_{i,j} = h_i(j - 1) + 1, j = 1, 2, \dots, J_i,$$

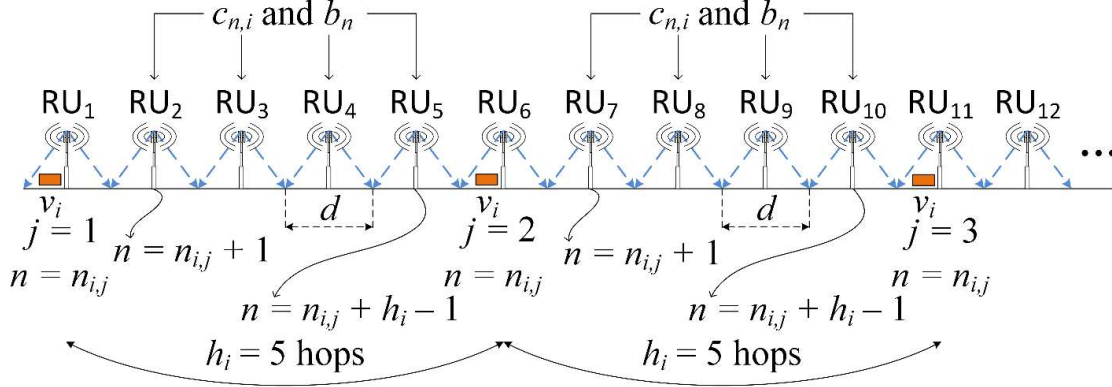
trong đó, J_i là số RSU lưu trữ video i , được tính theo công thức (1.2):

$$J_i = \left\lfloor \frac{N - 1}{h_i} \right\rfloor + 1.$$

Toán tử $\lfloor \cdot \rfloor$ trả về số nguyên lớn nhất không vượt quá giá trị thực cho trước.

2) Lưu trữ tại UAV

Đối với một video i , số RSU không lưu trữ video này là $N - J_i$, tạo thành các khoảng trống lưu trữ. Do số lượng UAV là hữu hạn, việc triển khai UAV cho từng khoảng trống là không khả thi. Vì vậy, với M UAV khả dụng, mỗi UAV cần lựa chọn lưu trữ một tập con các video và dừng tại các RSU phù hợp với độ cao nhất định nhằm đảm bảo phiên truyền video liên tục (hình 3.2). Cụ thể, một UAV dừng phía trên RSU n (với $n \neq n_{i,j}$) sẽ truyền video i được lưu trữ đến các RV nếu $c_{n,i} = 1$, ngược lại, nếu $c_{n,i} = 0$ thì UAV không lưu trữ video i tại RSU n .



Hình 3.2: Ví dụ về lưu trữ và dùng với $h_i = 5$

3.3.3. Dung lượng truyền từ RSU đến phương tiện (R2V)

Để xác định dung lượng truyền R2V, nghiên cứu áp dụng mô hình kênh R2V được đề xuất trong [34]. Theo mô hình này, đoạn đường do một RSU bao phủ được chia thành $K = 7$ vùng, với các tham số chi tiết được liệt kê trong bảng 1.1.

Vùng	1	2	3	4	5	6	7
d_k (m)	25	30	40	60	40	30	25
r_k (m)	125	100	70	30	70	100	125
R_k (Mbps)	1	2	5.5	11	5.5	2	1

Dung lượng truyền cho mỗi RV yêu cầu video i tại vùng k , do RSU n phục vụ, được xác định như sau:

$$C_{n,k,i}^{R2V} = \begin{cases} \frac{p_{k,i}^{RV} R_k}{V_{k,i}}, & \text{if } n = n_{i,j}, \\ 0, & \text{if } n \neq n_{i,j}, \end{cases} \quad (3.2)$$

Trong đó:

- $p_{k,i}^{RV}$ là xác suất tồn tại ít nhất một RV yêu cầu video i trong vùng k ;
- $V_{k,i}$ là số lượng RV yêu cầu video i trong vùng k .

Hai giá trị này được xác định lần lượt như sau:

$$p_{k,i}^{RV} = 1 - e^{-d_k \lambda_i^{RV}}, \quad (3.3)$$

và

$$V_{k,i} = d_k \lambda_i^{RV}, \quad (3.4)$$

với d_k là chiều dài của vùng k .

3.3.4. Dung lượng truyền từ UAV đến phương tiện (U2V)

Tại các RSU không lưu trữ video i ($n \neq n_{i,j}$), UAV dừng phía trên trục giữa của đoạn đường ở độ cao h_n và truyền video i đến các RV thông qua mô hình kênh không-đất [103], [107], [108], [109] như trong hình 3.3, trong đó $h_n = \frac{d}{2} \tan(\theta_n)$ với giá trị θ_n tùy thuộc vào môi trường trong bảng 3.2. Mô hình này bao gồm hai loại liên kết: đường truyền nhìn thẳng (LoS) và không nhìn thẳng (NLoS - Non-Line of Sight). Suy hao đường truyền cho liên kết LoS hoặc NLoS giữa UAV và RV điển hình trong vùng k thuộc đoạn đường n được xác định bởi:

$$\psi_{n,k}^A = \left(\frac{4\pi f_c d_{n,k}}{c} \right)^2 \eta_n^A, A \in \{\text{LoS}, \text{NLoS}\}, \quad (3.5)$$

Trong đó:

- f_c là tần số sóng mang,
- c là vận tốc ánh sáng,
- η_n^A là hệ số suy hao đối với liên kết LoS hoặc NLoS,
- $d_{n,k}$ là khoảng cách lớn nhất từ UAV đến RV trong vùng k , được xác định như sau: $d_{n,k} = \sqrt{h_n^2 + r_k^2}$

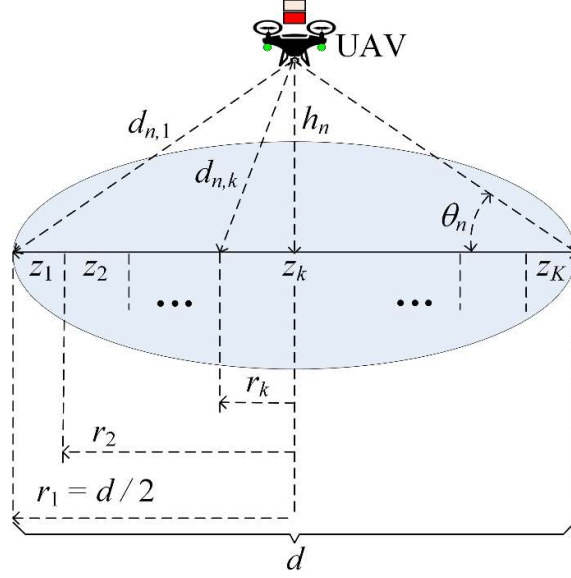
Kết hợp xác suất của đường truyền LoS ($P_{n,k}^{\text{LoS}}$) và xác suất của đường truyền NLoS ($P_{n,k}^{\text{NLoS}} = 1 - P_{n,k}^{\text{LoS}}$), suy hao đường truyền trung bình là:

$$\begin{aligned} \bar{\psi}_{n,k} &= \psi_{n,k}^{\text{LoS}} P_{n,k}^{\text{LoS}} + \psi_{n,k}^{\text{NLoS}} P_{n,k}^{\text{NLoS}} \\ &= \left(\frac{4\pi f_c d_{n,k}}{c} \right)^2 (\eta_n^{\text{LoS}} P_{n,k}^{\text{LoS}} + \eta_n^{\text{NLoS}} P_{n,k}^{\text{NLoS}}), \end{aligned} \quad (3.6)$$

trong đó, xác suất xảy ra đường truyền LoS được tính bởi:

$$P_{n,k}^{\text{LoS}} = \frac{1}{1 + \kappa_n \exp[-\nu_n(\xi_{n,k} - \kappa_n)]}, \quad (3.7)$$

với κ_n và ν_n là các tham số phụ thuộc môi trường truyền của đoạn đường n , và $\xi_{n,k} = \frac{180}{\pi} \arctan\left(\frac{h_n}{r_k}\right)$ [102], [110].



Hình 3.3: Mô hình vùng phủ của UAV trong đoạn đường n

Giả sử UAV được cấp phát tài nguyên phổ tần xuống (downlink) và không có can nhiễu, tỷ số tín hiệu trên nhiễu (SNR) của kênh truyền từ UAV đến RV trong vùng k được xác định bởi:

$$\gamma_{n,k,i} = \frac{b_n c_{n,i} (\bar{\psi}_{n,k})^{-1} P}{\sigma^2}, \quad (3.8)$$

trong đó, b_n là chỉ số dừng được sử dụng để xác định UAV có dừng tại RSU $n \neq n_{i,j}$ hay không, với $b_n = 1$ nếu UAV dừng và $b_n = 0$ nếu không dừng. Điều kiện $\sum_{n=1, n \neq n_{i,j}}^N b_n = M$ cho biết có tổng cộng M UAV được triển khai, trong đó $1 \leq M \leq N$. Ký hiệu P là công suất phát của UAV.

Từ biểu thức (3.8), ta thu được dung lượng truyền tương ứng, được xác định như sau:

$$C_{n,k,i}^{UV} = \begin{cases} p_{k,i}^{RV} W \log_2(1 + \gamma_{n,k,i}), & \text{if } n = n_{i,j}, \\ 0, & \text{if } n \neq n_{i,j}, \end{cases} \quad (3.9)$$

với W là băng thông hệ thống.

3.3.5. Thời gian phục vụ trung bình

Thời gian cần thiết để phục vụ các RV trong vùng k được xác định là thời gian phương tiện di chuyển qua vùng đó, được tính bởi:

$$t_{n,k} = \frac{d_k}{s_n}, \quad (3.10)$$

Dựa trên biểu thức này, thời gian phục vụ trung bình của hệ thống đối với I video, đồng thời là hàm mục tiêu của bài toán tối ưu, được xác định như sau:

$$\bar{T} = \sum_{i=1}^I p_i t_i, \quad (3.11)$$

trong đó t_i là tổng thời gian cần thiết để truyền video i thông qua sự phối hợp giữa RSU và UAV, được tính chi tiết trong thuật toán 3.1.

Thuật toán 3.1 mô tả quá trình phân phối nội dung video trong mô hình CAHO, trong đó, việc truyền video i được thực hiện thông qua sự phối hợp giữa các RSU và các UAV trong hệ thống. Thuật toán nhận đầu vào là kích thước của video S_i cần truyền tới phương tiện yêu cầu, đồng thời khởi tạo thời gian phục vụ ban đầu $t_i = 0$. Mục tiêu của thuật toán là xác định tổng thời gian cần thiết để phân phối toàn bộ nội dung video i tới phương tiện, dựa trên khả năng truyền dữ liệu từ các RSU và UAV trong vùng phục vụ.

Trước hết, thuật toán xét các RSU có khả năng cung cấp nội dung video. Vòng lặp ngoài cùng được thực hiện theo chỉ số j từ 1 đến J_i , trong đó J_i biểu diễn số RSU có lưu trữ video i . Vị trí các RSU này được xác định thông qua biểu thức $n_{i,j} = h_i(j - 1) + 1$. Sau khi xác định vị trí RSU, thuật toán lần lượt xét các phương tiện trong vùng phủ của RSU thông qua vòng lặp theo chỉ số k . Trong mỗi lần lặp theo k , nếu kích thước dữ liệu còn lại của video S_i vẫn lớn hơn 0, hệ thống sẽ thực hiện truyền một phần dữ liệu từ RSU đến phương tiện thông qua liên kết R2V. Lượng dữ liệu truyền được xác định dựa trên dung lượng kênh $C_{n_{i,j},k,i}^{R2V}$ và thời gian truyền tương ứng $t_{n_{i,j},k}$. Sau mỗi lần truyền, kích thước dữ liệu còn lại S_i được cập nhật bằng cách trừ đi phần dữ liệu vừa truyền, đồng thời thời gian phục vụ t_i được cộng thêm khoảng thời gian truyền tương ứng. Quá trình này tiếp tục cho đến khi dữ liệu video được truyền hết hoặc khi RSU hiện tại không thể cung cấp thêm dữ liệu.

Thuật toán 3.1: Phân phối video CAHO

Input: S_i
 $t_i = 0$
Output: t_i

```

1: for  $j = 1 : J_i$  do
2:    $n_{i,j} = h_i(j - 1) + 1$ 
3:   for  $k = 1 : K$  do
4:     if  $S_i > 0$  then
5:        $S_i = S_i - t_{n_{i,j},k} C_{n_{i,j},k,i}^{R2V}$ 
6:        $t_i = t_i + t_{n_{i,j},k}$ 
7:     else
8:       break
9:     end if
10:  end for
11:  for  $n = n_{i,j} + 1 : n_{i,j} + h_i - 1$  do
12:    if  $n \leq N$  then
13:      for  $k = 1 : K$  do
14:        if  $S_i > 0$  then
15:           $S_i = S_i - t_{n,k} C_{n,k,i}^{U2V}$ 
16:           $t_i = t_i + t_{n,k}$ 
17:        else
18:          break
19:        end if
20:      end for
21:    end if
22:  end for
23: end for

```

Sau khi khai thác khả năng truyền dữ liệu từ RSU, thuật toán tiếp tục xét đến các UAV trong hệ thống nhằm hỗ trợ phân phối phần dữ liệu còn lại. Cụ thể, vòng lặp tiếp theo được thực hiện với chỉ số n từ $n_{i,j} + 1$ đến $n_{i,j} + h_i - 1$, biểu diễn các UAV nằm trong vùng phục vụ của cụm RSU không lưu trữ video i . Tuy nhiên, quá trình này chỉ được thực hiện khi chỉ số n không vượt quá tổng số RSU trong hệ thống. Trong mỗi UAV, thuật toán tiếp tục duyệt qua các phương tiện trong vùng phủ thông qua chỉ số k . Tương tự như trường hợp RSU, nếu dữ liệu video vẫn còn chưa được truyền hết, UAV sẽ tiếp tục truyền dữ liệu đến phương tiện thông qua liên kết U2V. Lượng dữ liệu truyền trong mỗi bước được xác định dựa trên dung lượng kênh $C_{n,k,i}^{U2V}$ và thời gian truyền $t_{n,k}$.

Sau mỗi lần truyền, kích thước dữ liệu còn lại S_i và thời gian phục vụ t_i được cập nhật tương ứng. Nếu dữ liệu video đã được truyền hết, vòng lặp hiện tại sẽ được dừng lại để tránh các phép tính không cần thiết.

Như vậy, thuật toán thực hiện quá trình phân phối nội dung theo hai giai đoạn. Trước hết, hệ thống ưu tiên khai thác khả năng cung cấp dữ liệu từ các RSU. Trong trường hợp dữ liệu video vẫn chưa được truyền hết, các UAV sẽ tiếp tục tham gia hỗ trợ phân phối nội dung cho phương tiện. Cách tổ chức này phản ánh đúng đặc điểm của mô hình CAHO, trong đó, hạ tầng mặt đất đóng vai trò nguồn cung cấp nội dung chính, còn UAV được sử dụng như các nút hỗ trợ nhằm mở rộng vùng phủ và cải thiện khả năng phân phối nội dung. Kết quả cuối cùng của thuật toán là giá trị t_i , biểu diễn tổng thời gian phục vụ cần thiết để truyền video i đến phương tiện yêu cầu trong cấu hình hệ thống đã cho.

3.4. THIẾT KẾ TỐI ƯU CAHO

3.4.1. Bài toán tối ưu CAHO

Dựa trên hàm mục tiêu được xác định trong (3.11), đồng thời xét đến các ràng buộc về tài nguyên lưu trữ tại RSU và UAV, bài toán tối ưu CAHO được phát biểu như sau:

$$\min \bar{T} \quad (3.12a)$$

$$h_i, b_n, c_{n,i}$$

Với các ràng buộc:

$$1 \leq h_i \leq N - 1, \forall i, \quad (3.12b)$$

$$S^{RSU} + S^{UAV} \leq \delta N \sum_{i=1}^I S_i, \quad (3.12c)$$

$$\sum_{n=1}^N b_n \leq M, \quad (3.12d)$$

Với h_i : bán kính lưu trữ, b_n : quyết định UAV có dừng tại RSU n hay không, $c_{n,i}$: UAV tại RSU n có lưu trữ video i hay không, và trong đó ràng buộc (3.12c) giới hạn tổng dung lượng lưu trữ được sử dụng cho việc lưu trữ nội dung tại RSU và UAV. Hai đại lượng S^{RSU} và S^{UAV} lần lượt biểu thị tài nguyên lưu trữ tiêu thụ tại RSU và tại UAV, được xác định như sau:

$$S^{RSU} = \sum_{i=1}^I J_i S_i, \quad (3.13)$$

$$S^{UAV} = \sum_{n=1}^N \sum_{i=1}^I c_{n,i} b_n S_i. \quad (3.14)$$

Bài toán CAHO trong (3.12) nhằm cực tiểu thời gian phục vụ video trong mạng UAV-VANET bằng cách nên triển khai UAV ở những vị trí nào, mỗi UAV nên lưu trữ những video gì và đồng thời cấu hình lưu trữ tại các RSU ra sao với số lượng UAV và ngân sách lưu trữ chung cho cả hạ tầng mặt đất lẫn trên không giới hạn. Khác với DPC chỉ quyết định phân bổ nội dung giữa hai tầng lưu trữ đã tồn tại sẵn, CAHO phải quyết định đồng thời phân bổ nội dung và phân bổ không gian, tức là cả "lưu gì", "lưu ở đâu" và "vị trí UAV". Đây là sự khác biệt mang tính cấu trúc tạo nên đặc trưng của bài toán CAHO. Như vậy, ba biến tối ưu của CAHO có vai trò gắn kết chặt chẽ với nhau qua hàm mục tiêu thời gian phục vụ gồm:

- Biến nguyên $h_i \in \{1, \dots, N-1\}$ là bán kính lưu trữ video i tại các RSU, kế thừa ý nghĩa từ mô hình DPC nhưng được sử dụng trong bối cảnh có thêm UAV.
- Biến nhị phân $c_{n,i} \in \{0,1\}$ quyết định lưu trữ tại UAV: $c_{n,i} = 1$ nếu UAV đặt tại đoạn đường n lưu video i , ngược lại $c_{n,i} = 0$.
- Biến nhị phân $b_n \in \{0,1\}$ quyết định vị trí dừng của UAV: $b_n = 1$ nếu có UAV dừng tại đoạn đường n , ngược lại $b_n = 0$.

Ngoài ra, trong bài toán CAHO, ràng buộc (3.12b) tương tự như (2.14b) nhằm đảm bảo mỗi video phải được lưu trữ tại ít nhất một RSU và bán kính lưu trữ không vượt quá quy mô hệ thống. Trong bối cảnh CAHO, ràng buộc này còn có vai trò gián tiếp xác định các khoảng trống lưu trữ giữa các RSU, chính những khoảng trống này là cơ hội để UAV phát huy vai trò bù đắp thông qua các quyết định $c_{n,i}$ và b_n cùng với độ cao h_n phù hợp với môi trường tại đoạn đường n được cho trước. Ràng buộc (3.12c) là giới hạn tổng dung lượng lưu trữ, áp dụng cho cả RSU và UAV. Ràng buộc này làm cho hệ thống phải thực hiện đánh đổi dung lượng lưu trữ giữa RSU và UAV: mỗi byte lưu thêm tại UAV phải đổi lấy một byte tiết kiệm tại RSU và ngược lại. Điều này tạo ra

sự cạnh tranh tài nguyên giữa hạ tầng mặt đất và trên không, buộc nghiệm tối ưu phải xác định tỷ lệ phân bổ hợp lý giữa hai tầng dựa trên đặc điểm của hệ thống (mật độ phương tiện, độ phổ biến nội dung, chất lượng kênh U2V và R2V). Với ràng buộc (3.12d), đây là ràng buộc về số lượng UAV khả dụng. Ràng buộc này phản ánh thực tế triển khai khi số lượng UAV trong một hệ thống VANET luôn bị giới hạn bởi chi phí đầu tư, năng lực quản lý vận hành, quy định bay và các yếu tố kỹ thuật khác. Tham số M là số UAV tối đa được phép triển khai, do đó bài toán phải lựa chọn không quá M vị trí dừng tối ưu trong tổng số N đoạn đường xem xét.

3.4.2. Giải bài toán bằng GA

GA đã được chứng minh là một công cụ hiệu quả để giải các bài toán tối ưu phức tạp trong nhiều lĩnh vực thực tiễn khác nhau, từ hệ thống truyền thông đến các ứng dụng truyền video [38], [39], [40], [41], [95], [96], [97], [98]. Ưu điểm của GA xuất phát từ các nguyên lý tiến hóa tự nhiên, cho phép tìm kiếm hiệu quả nghiệm tối ưu hoặc xấp xỉ tối ưu toàn cục trong các không gian tìm kiếm có tính đa đỉnh và phi tuyến, đồng thời tránh được hiện tượng hội tụ cục bộ.

Trong nghiên cứu này, GA [99] được áp dụng để giải bài toán tối ưu CAHO trong (3.12). Tuy nhiên, GA chỉ hỗ trợ trực tiếp các ràng buộc dạng cận trên và cận dưới đơn giản, như ràng buộc (3.12b), trong khi các ràng buộc phức tạp hơn, chẳng hạn (3.12c) và (3.12d), cần được xử lý gián tiếp thông qua phương pháp hàm phạt (penalty method) [39]. Theo đó, bài toán tối ưu có ràng buộc được chuyển đổi thành một bài toán không ràng buộc bằng cách biểu diễn các ràng buộc (3.12c) và (3.12d) dưới dạng như sau:

$$\begin{cases} \Delta S = \delta N \sum_{i=1}^I S_i - S^{RSU} - S^{UAV} \geq 0, \\ \Delta M = M - \sum_{n=1}^N b_n \geq 0 \end{cases} \quad (3.15)$$

Từ đó, hàm phạt được xây dựng dưới dạng:

$$F = \rho_1(\min\{0, \Delta S\})^2 + \rho_2(\min\{0, \Delta M\})^2, \quad (3.16)$$

trong đó ρ_1 và ρ_2 là các hệ số điều chỉnh mức độ phạt khi nghiệm vi phạm các ràng buộc tương ứng.

Các hệ số phạt ρ_1 và ρ_2 được điều chỉnh tăng dần theo từng bước nhằm đảm bảo các nghiệm tìm được không vi phạm các ràng buộc của bài toán. Ban đầu, các hệ số này được đặt ở mức nhỏ, cụ thể là 10^{-3} , sau đó giá trị của hàm phạt trong (3.16) được kiểm tra. Nếu tại thời điểm đó hàm phạt vẫn còn lớn và các hạng tử ràng buộc tương ứng chưa triệt tiêu, các hệ số ρ_1 và ρ_2 sẽ được tăng lên theo hệ số 10. Trong quá trình điều chỉnh, mỗi hệ số chỉ tiếp tục được tăng khi ràng buộc tương ứng của nó vẫn còn bị vi phạm. Ngược lại, nếu một ràng buộc đã được thỏa mãn và hạng tử tương ứng trong hàm phạt đã bằng 0, hệ số phạt liên quan sẽ được giữ cố định để tránh làm thay đổi cân bằng của các thành phần khác trong hàm mục tiêu.

Quá trình điều chỉnh này được lặp lại cho đến khi các hạng tử vi phạm trong hàm phạt giảm dần và tiến về 0. Ở các bước cuối cùng của quá trình tối ưu, hệ số tăng của ρ_1 và ρ_2 được giảm xuống các giá trị nhỏ hơn, chẳng hạn như 5 hoặc 2 thay vì 10. Việc điều chỉnh theo các bước nhỏ hơn giúp tránh những biến động lớn của hàm phạt, đồng thời cho phép quá trình tìm kiếm nghiệm hội tụ ổn định hơn, đảm bảo các ràng buộc được thỏa mãn mà không làm ảnh hưởng đến giá trị của hàm mục tiêu.

Cuối cùng, bài toán tối ưu CAHO được chuyển thành bài toán tối ưu không ràng buộc như sau:

$$\min_{h_i, b_n, c_{n,i}} T_F = \bar{T} + F \quad (3.17)$$

Chi tiết GA áp dụng để giải bài toán (3.17) được trình bày trong thuật toán 3.2. Do bài toán CAHO bao gồm đồng thời biến tối ưu nguyên (h_i) và biến tối ưu nhị phân ($c_{n,i}, b_n$), các biến này không thể được mã hóa trong cùng một nhiễm sắc thể cũng như không thể sử dụng chung các toán tử lai ghép và đột biến. Vì vậy, mỗi cá thể trong GA được biểu diễn bằng hai chuỗi con:

- Chuỗi $X_1^{(z)}$ mã hóa các biến nguyên h_i ;
- Chuỗi $X_2^{(z)}$ mã hóa các biến nhị phân $c_{n,i}$ và b_n .

GA được thực hiện thông qua các toán tử chính bao gồm chọn lọc, lai ghép và đột biến, và được lặp lại cho đến khi thỏa mãn một trong hai điều kiện dừng sau:

(i) giá trị hàm phạt trung bình nhỏ hơn 10^{-3} trong 10 thế hệ liên tiếp;

hoặc (ii) số thế hệ đạt đến giá trị cực đại $N_G = 100$.

Vì GA là thuật toán tìm kiếm heuristic thích nghi, độ phức tạp của thuật toán 3.2 chịu ảnh hưởng trực tiếp bởi kích thước không gian tìm kiếm N_P và số lần lặp N_G . Giá trị của N_P và N_G có thể thay đổi tùy thuộc vào nhiều yếu tố khác nhau, bao gồm:

- (i) quy mô hệ thống, đặc trưng bởi số UAV M , số RSU N và số video I ;
- (ii) gánh nặng tính toán phát sinh từ hàm mục tiêu, các ràng buộc và các biến tối ưu trong bài toán (3.12); và
- (iii) yêu cầu về độ chính xác của các ứng dụng và dịch vụ truyền video.

Ngoài ra, độ phức tạp tổng thể của thuật toán còn chịu ảnh hưởng bởi việc triển khai các toán tử di truyền, chẳng hạn như chọn lọc, lai ghép và đột biến. Do đó, độ phức tạp của thuật toán 3.2 có bậc xấp xỉ $\mathcal{O}(N_P N_G)$ [22], [26].

Trong thực tế, thuật toán 3.2 phù hợp hơn khi được triển khai tại một trạm gốc (MBS) duy nhất, nơi chiến lược CAHO được thực hiện tập trung để điều phối toàn bộ các RSU, UAV và các video phổ biến trong một đoạn đường giới hạn. Tuy nhiên, khi quy mô hệ thống mở rộng với tuyến đường dài hơn, tương ứng với các giá trị M , N và I tăng lên, việc triển khai thực tế CAHO phải đối mặt với hai thách thức chính:

- Thứ nhất, việc thu thập thông tin hệ thống, như thông tin trạng thái kênh, để xây dựng bài toán tối ưu CAHO trở nên ngày càng khó khăn.
- Thứ hai, độ phức tạp tính toán của thuật toán 3.2 gia tăng đáng kể.

Để khắc phục những thách thức này, số lượng RSU, UAV và các video phổ biến được gán cho mỗi MBS cần được lựa chọn cẩn trọng, phù hợp với quy mô hệ thống và năng lực xử lý của MBS. Hơn nữa, trong các kịch bản liên quan đến video có kích thước lớn, nhiều MBS có thể phối hợp với nhau để cung cấp các dịch vụ truyền video liên tục cho các RV [30].

Thuật toán 3.2: GA cho CAHO**Đầu vào:** Thông số trong bảng 3.3, $Gen = 1$ **Đầu ra:** $T_F^*(h_i^*, c_{n,i}^*, b_n^*)$

1. Khởi tạo ngẫu nhiên

- N_p chuỗi con $\{X_1^{(z)}\}$ với $z = 1, 2, \dots, N_p$, mỗi chuỗi bao gồm I biến nguyên tối ưu $\{h_i\}$.
- N_p chuỗi con $\{X_2^{(z)}\}$, mỗi chuỗi gồm $(N \times I + N)$ bit để biểu diễn một tập gồm $N \times I$ biến tối ưu nhị phân cho $\{c_{n,i}\}$ và N biến tối ưu nhị phân cho $\{b_n\}$.

2. Tính giá trị hàm thích nghi (fitness) theo công thức (3.17) được tập $\{T_F^{(z)}\}$ tức là $\{T_F^{(z)}(X_1^{(z)}, X_2^{(z)})\}$.

3. Trong khi điều kiện dừng TC chưa thỏa mãn, thực hiện:

4. Đặt $\{X^{(z)}\} = \{[X_1^{(z)}][X_2^{(z)}]\}$ cùng giá trị hàm phạt $T_F^{(z)}$ tương ứng vào tập xếp hạng để lai tạo.5. Lựa chọn $N_{PG} = N_p \times N_G$ cá thể tốt nhất có giá trị hàm thích nghi nhỏ nhất bằng toán tử lấy mẫu phổ quát ngẫu nhiên [18] để lai ghép, được tập:

$$\{X^{(t)}\} = \{[X_1^{(t)}][X_2^{(t)}]\}, t=1, 2, \dots, N_{PG}$$

6. Chọn một cặp cá thể bố mẹ để tạo cá thể con bằng phép lai ghép một điểm với xác suất P_{ci} và P_{cb} lần lượt áp dụng cho $X_1^{(t)}$ và $X_2^{(t)}$.7. Thực hiện đột biến các cá thể con $\{X_1^{(t)}\}$ và $\{X_2^{(t)}\}$ tương ứng với các xác suất P_{mi} và P_{mb} tương ứng, bằng phép đảo bit nhằm khôi phục các đặc trưng di truyền tích cực có thể đã bị mất trong các bước trước, để thu được: $\{X^{(t),*}\} = \{[X_1^{(t),*}][X_2^{(t),*}]\}$ 8. Lặp lại bước 2 để có $T_F^{(t),*}(X_1^{(t),*}, X_2^{(t),*})$ 9. Chèn $\{X^{(t),*}\}$ và $\{T_F^{(t),*}\}$ vào quần thể hiện tại để thu được tập mới $\{X^{(z)}\}$ và $\{T_F^{(z)}\}$.10. Cập nhật $Gen = Gen + 1$.

11. Kết thúc vòng lặp.

12. Tìm giá trị hàm thích nghi tốt nhất $T_F^*(h_i^*, c_{n,i}^*, b_n^*) \in \{T_F^{(z)}(X_1^{(z)}, X_2^{(z)})\}$ **3.4.3. Triển khai giải pháp CAHO**

Để hoàn thiện quá trình thiết kế giải pháp CAHO, các bước thực hiện triển khai

với một số thay đổi so với DPC, cụ thể như sau:

- Bước 1: Một trạm gốc chủ được bầu chọn từ các trạm gốc liên quan đến tuyến đường đang xét.
- Bước 2: Trạm gốc chủ yêu cầu các trạm gốc thành viên thu thập thông tin của các RSU, UAV và phương tiện trên các đoạn đường được bao phủ liên quan.
- Bước 3: Dựa vào thông tin thu thập, trạm gốc chủ xây dựng và giải bài toán tối ưu để tìm ra h_i , $c_{n,i}$ và b_n .
- Bước 4: Triển khai kết quả tối ưu vào tuyến đường theo h_i , $c_{n,i}$ và b_n như được mô tả trong hình 3.2.

Các UAV vẫn tiếp tục dừng tại vị trí được chỉ định cho đến khi sắp hết năng lượng hoặc nhận nhiệm vụ mới (khi quá trình trên được thực hiện lại). Trong trường hợp sắp hết năng lượng, trạm nền chủ sẽ chỉ định UAV khác đến thay thế để UAV hiện tại bay đến trạm sạc gần nhất.

3.5. ĐÁNH GIÁ HIỆU SUẤT

3.5.1. Thiết lập thông số

Trong phần đánh giá hiệu năng, các thông số hệ thống cùng với các tham số của GA được thiết lập như trình bày trong bảng 3.1. Bên cạnh đó, các thông số đặc trưng cho môi trường truyền dẫn tại các đoạn đường khác nhau được lựa chọn ngẫu nhiên và được tổng hợp trong bảng 3.2, nhằm phản ánh tính không đồng nhất của điều kiện truyền trong các kịch bản VANET thực tế.

Các thông số mô phỏng trong mô hình CAHO được lựa chọn cẩn thận nhằm phản ánh các đặc trưng cơ bản của môi trường UAV-VANET, bao gồm mật độ phương tiện, phạm vi phủ sóng của RSU/UAV, kích thước nội dung video, dung lượng lưu trữ và điều kiện kênh truyền R2V/U2V. Các thông số này cho phép tạo ra các kịch bản đại diện để đánh giá xu hướng và hiệu năng của mô hình đề xuất dưới nhiều điều kiện khác nhau. Đặc biệt, các tham số như độ lệch phân bố phổ biến nội dung α , mật độ phương tiện λ và số lượng UAV M được thay đổi trong quá trình đánh giá nhằm kiểm tra phản ứng của hệ thống và tính ổn định của phương pháp.

Bảng 3.1: Thông số hệ thống CAHO

Ký hiệu	Giá trị	Đơn vị
M	5	UAV
N	25	RSU
I	10	video
K	7	vùng/RSU
S_i	Ngẫu nhiên trong khoảng 0.25 đến 2.5	Gbit
s_n	Ngẫu nhiên trong khoảng 30 đến 60	km/h
d	250	m
λ	0.025	VU/m
α	1	
δ	0.5	
W	10	MHz
P	5	W
σ^2	10^{-12}	W
c	3×10^8	m/s
f_c	2	GHz
ρ_1, ρ_2	$10^{-6}, 1$	
N_p	3000	cá thể
N_G	100	thế hệ
P_G	0.9	
P_{ci}, P_{cb}	0.9, 0.6	
P_{mi}, P_{mb}	$10^{-12}, 10^{-12}$	

Để đánh giá toàn diện hiệu quả của phương pháp đề xuất CAHO, nghiên cứu tiến hành so sánh với bốn phương pháp đối sánh bao gồm:

- (i) RAHO (RAnDOM HOVering), trong đó, vị trí dừng của UAV được lựa chọn ngẫu nhiên;
- (ii) RACR (RAnDOM Caching at RSUs), với chiến lược lưu trữ tại các RSU được xác định ngẫu nhiên;
- (iii) RACU (RAnDOM Caching at UAVs), trong đó, nội dung lưu trữ tại UAV được lựa chọn ngẫu nhiên; và

- (iv) NUAV (None-UAV deployment), tương ứng với trường hợp hệ thống không triển khai UAV.

Các biến tối ưu được xem xét trong từng phương pháp đối sánh được tóm tắt trong bảng 3.3. Cụ thể, đối với phương pháp RAHO, các biến lưu trữ h_i tại RSU và $c_{n,i}$ tại UAV được tối ưu, trong khi biến chỉ định vị trí dừng b_n được sinh ngẫu nhiên. Ngược lại, trong RACR, các biến $c_{n,i}$ và b_n được tối ưu, chiến lược lưu trữ tại RSU h_i được lựa chọn ngẫu nhiên. Đối với RACU, các biến h_i và b_n được tối ưu, trong khi $c_{n,i}$ được sinh ngẫu nhiên. Riêng với NUAV, UAV không được triển khai, tức là $M = 0$.

Bảng 3.2: Thông số môi trường [102], [110]

Thông số Môi trường	κ_n	ν_n	η_n^{LoS}	η_n^{NLoS}	θ_n
Ngoại ô	4.88	0.43	0.1	21	20.24°
Đô thị	9.61	0.16	1	20	42.44°
Đô thị dày đặc	12.08	0.11	1.6	23	54.62°
Đô thị cao tầng	27.23	0.08	2.3	34	75.52°

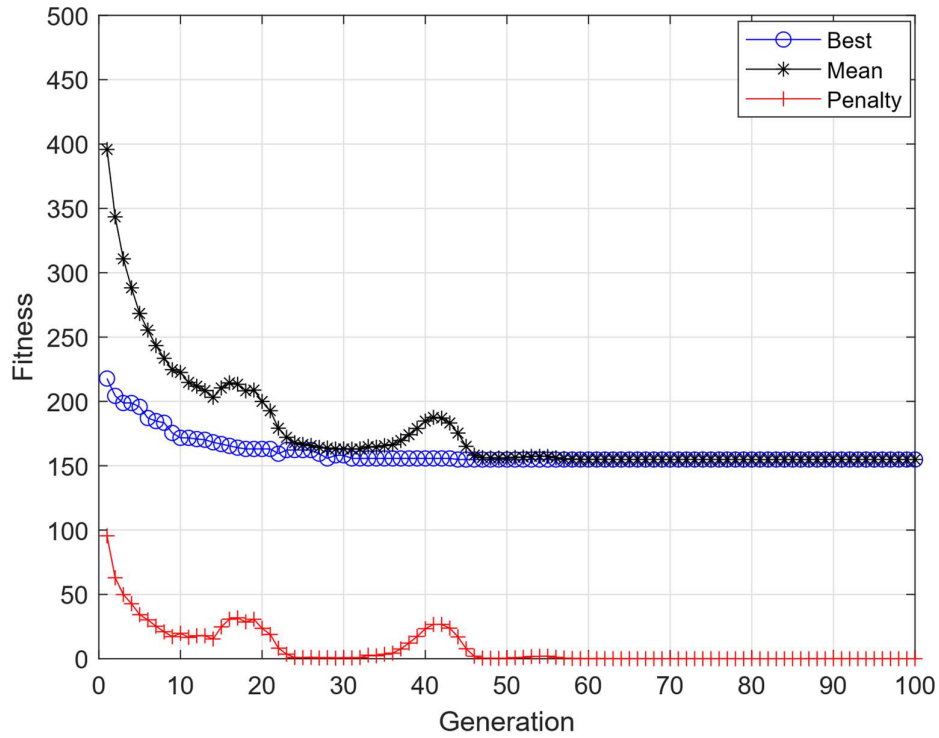
Để đảm bảo tính công bằng trong so sánh, tất cả các phương pháp đối sánh đều được ràng buộc sử dụng dung lượng lưu trữ bằng hoặc xấp xỉ (trong một số trường hợp có thể nhỏ hơn) so với phương pháp CAHO. Cách thiết lập này nhằm đảm bảo rằng sự khác biệt về hiệu năng thu được chủ yếu đến từ chiến lược tối ưu đề xuất, thay vì do lợi thế về tài nguyên lưu trữ.

Bảng 3.3: Giá trị tối ưu của các phương pháp so sánh

Giá trị tối ưu Phương pháp	h_i	$c_{n,i}$	b_n
RAHO	Y	Y	N
RACR	N	Y	Y
RACU	Y	N	Y
NUAV	Y	N	N
CAHO	Y	Y	Y

3.5.2. Đánh giá sự hội tụ của GA

Hiệu quả GA được đánh giá thông qua quá trình hội tụ của hàm mục tiêu và tính ổn định của nghiệm thu được trong các kịch bản hệ thống khác nhau. Cụ thể, GA được sử dụng để giải bài toán tối ưu CAHO nhằm cực tiểu hóa thời gian phục vụ trung bình, trong đó các biến tối ưu bao gồm chiến lược lưu trữ tại RSU, lưu trữ tại UAV và vị trí dừng của UAV. Hình 3.4 thể hiện quá trình hội tụ của GA thông qua ba giá trị: giá trị fitness tốt nhất (Best), giá trị fitness trung bình (Mean) được xác định trong (3.17), và giá trị hàm phạt trung bình (Penalty) được xác định trong (3.16), theo số thế hệ.



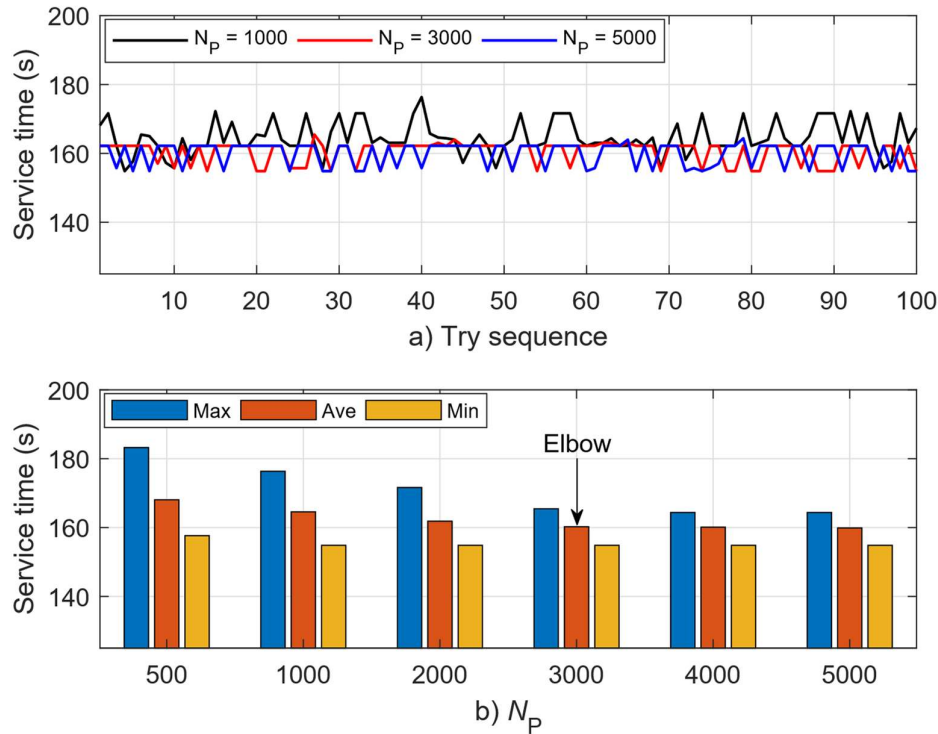
Hình 3.4: Tốc độ hội tụ của GA trong phương pháp CAHO

Trong các thế hệ đầu, khoảng cách giữa Best và Mean phản ánh độ đa dạng của quần thể khởi tạo và cho phép GA khám phá không gian nghiệm. Khi số thế hệ tăng, Best và Mean tiến gần nhau, thể hiện xu hướng hội tụ của thuật toán.

Mặc dù Best và Mean đã trở nên khá gần nhau kể từ khoảng thế hệ thứ 30, GA vẫn tiếp tục cải thiện giá trị fitness đồng thời đưa giá trị hàm phạt về gần 0. Điều này cho thấy trong giai đoạn sau, thuật toán không chỉ tinh chỉnh nghiệm theo hướng tối ưu hơn mà còn đảm bảo tất cả các ràng buộc của bài toán tối ưu CAHO đều được thỏa mãn.

Sau khoảng 45 thế hệ, giá trị Mean hội tụ hoàn toàn về Best và duy trì ổn định cho đến thế hệ cuối, cho thấy GA đã đạt được nghiệm tối ưu hoặc xấp xỉ tối ưu. Bên cạnh đó, sự suy giảm nhanh của giá trị Penalty về 0 xác nhận rằng nghiệm thu được là nghiệm khả thi đối với các ràng buộc của hệ thống.

Hình 3.5 trình bày kết quả của 100 lần chạy độc lập cho mỗi kích thước quần thể $N_p = \{500, 1000, 2000, 3000, 4000, 5000\}$, như trong hình 3.5. Trong hình 3.5(a), khi kích thước quần thể nhỏ ($N_p = 1000$), thời gian phục vụ dao động mạnh hơn và thường có các giá trị cao hơn so với các trường hợp có kích thước quần thể lớn hơn. Ngược lại, khi N_p tăng lên 3000 và 5000, phương sai giảm rõ rệt và GA đạt nghiệm tối ưu ổn định hơn.



Hình 3.5: Độ ổn định và chính xác của GA trong phương pháp CAHO

Để phân tích sâu hơn về độ chính xác, giá trị lớn nhất (Max), giá trị trung bình (Ave) và giá trị nhỏ nhất (Min) của thời gian phục vụ trong 100 lần thực hiện độc lập được tổng hợp và thể hiện trong hình 3.5(b). Với $N_p = 500$, số lượng cá thể quá ít khiến GA không thể tìm được các nghiệm cực tiểu đáng tin cậy. Khi kích thước quần thể tăng

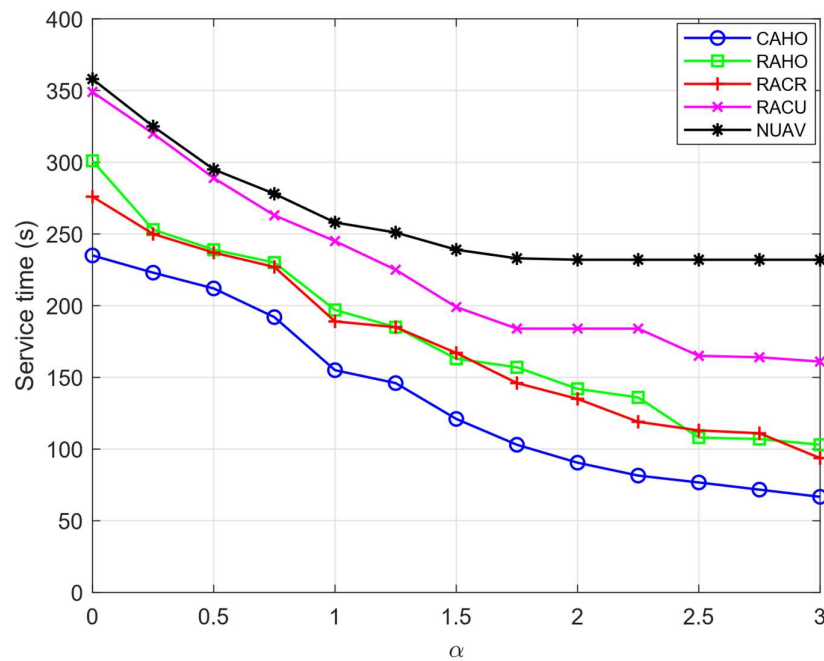
lên, xác suất xuất hiện các nghiệm gần tối ưu tăng theo, từ đó thu hẹp khoảng cách giữa Ave và Min.

Các kết quả cho thấy rằng lợi ích thu được không đáng kể khi $N_p > 3000$. Việc tiếp tục tăng kích thước quần thể vượt quá giá trị này không mang lại cải thiện rõ rệt về thời gian phục vụ, trong khi chi phí tính toán tăng nhanh. Do đó, điểm khuỷu (elbow point) tại $N_p = 3000$ được xem là lựa chọn hợp lý, tại đó GA đạt được độ chính xác cao với chi phí tính toán chấp nhận được.

Tổng hợp các kết quả trên cho thấy GA có tính khả thi và hiệu quả cao khi áp dụng cho bài toán tối ưu CAHO, đồng thời đạt được sự cân bằng tốt giữa độ chính xác và độ phức tạp tính toán, phù hợp cho các kịch bản triển khai thực tế.

3.5.3. Đánh giá hiệu suất CAHO

1) Hiệu năng theo tham số α

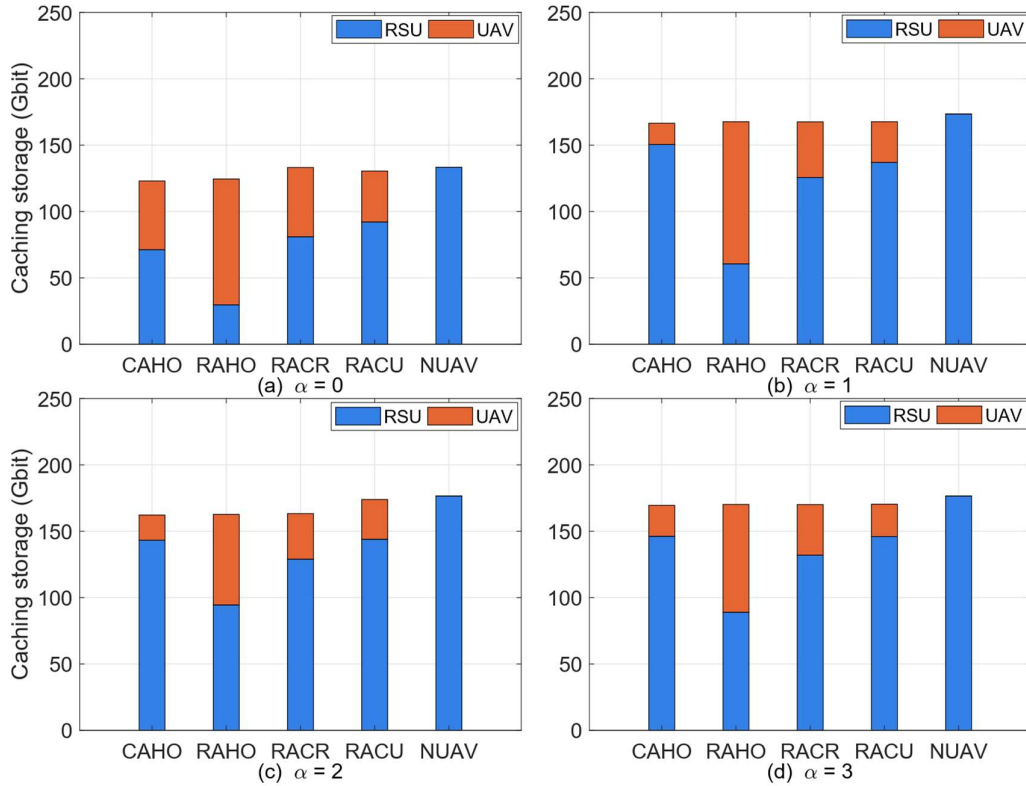


Hình 3.6: Thời gian phục vụ theo α

Hình 3.6 trình bày thời gian phục vụ của các phương pháp khác nhau theo tham số α , đại diện cho mức độ lệch của phân bố độ phổ biến video. Có thể quan sát rằng khi α tăng, thời gian phục vụ của tất cả các phương pháp đều giảm, cho thấy lợi ích của việc khai thác tài nguyên lưu trữ cho các video phổ biến nhất. Trong tất cả các phương pháp

được so sánh, CAHO đạt hiệu năng tốt nhất, với thời gian phục vụ thấp nhất, nhờ khả năng tối ưu đồng thời cả ba biến quan trọng, bao gồm: bán kính lưu trữ tại RSU (h_i), chiến lược lưu trữ tại UAV ($c_{n,i}$) và vị trí dừng của UAV (b_n). Sau CAHO, hai phương pháp RAHO và RACR cho hiệu năng tương đối tương đương, cho thấy rằng việc tối ưu vị trí dừng của UAV và chiến lược lưu trữ tại RSU có vai trò gần như nhau trong việc giảm thời gian phục vụ.

Xét về hiệu năng $\text{RAHO} \approx \text{RACR} > \text{RACU}$ trong hình 3.6 cho biết vai trò của từng biến tối ưu. Thứ nhất, vị trí dừng UAV (b_n) và chiến lược lưu trữ tại RSU (h_i) có mức đóng góp tương đương vào việc giảm thời gian phục vụ. Thứ hai, chiến lược lưu trữ tại UAV ($c_{n,i}$) có ảnh hưởng lớn nhất, do RACU, phương pháp duy nhất sinh $c_{n,i}$ ngẫu nhiên, có hiệu năng kém hơn cả hai phương pháp còn lại. Quan sát này lý giải vì sao CAHO, khi tối ưu cả ba biến đồng thời, đạt hiệu năng tốt nhất.



Hình 3.7: Dung lượng lưu trữ theo α

Xét về dung lượng lưu trữ sử dụng, hình 3.7 thể hiện mối quan hệ giữa tài nguyên lưu trữ của các phương pháp theo α . Mặc dù tổng dung lượng lưu trữ của các phương

pháp là gần tương đương (đảm bảo tính công bằng trong so sánh), cách phân bổ lưu trữ giữa RSU và UAV lại có sự khác biệt rõ rệt. Khi α tăng, UAV có xu hướng tập trung lưu trữ vào một số ít video phổ biến, dẫn đến việc dịch chuyển gánh nặng lưu trữ về phía RSU. Trong các phương pháp, CAHO sử dụng tài nguyên lưu trữ hiệu quả nhất, trong khi RAHO có hiệu quả thấp nhất do vị trí dừng của UAV được lựa chọn ngẫu nhiên, buộc UAV phải lưu trữ nhiều nội dung hơn để bù đắp cho sự kém hiệu quả này.

Hình 3.7 cho thấy tổng dung lượng lưu trữ của các phương pháp tương đương nhau (đảm bảo ràng buộc (3.12c)), nhưng cách phân bổ giữa RSU và UAV khác biệt rõ rệt. Khi α tăng, CAHO chuyển lưu trữ từ UAV sang RSU vì các video phổ biến nhất có thể được phục vụ hiệu quả qua R2V với h_i nhỏ. Ngược lại, RAHO buộc phải duy trì S^{UAV} cao trên toàn dải α do vị trí dừng ngẫu nhiên không trùng với khoảng trống lưu trữ tại RSU, dẫn đến lưu trữ dư thừa. Khoảng cách S^{UAV} giữa CAHO và RAHO phản ánh chi phí lưu trữ mà việc không tối ưu b_n phải trả.

2) Hiệu năng theo số lượng UAV M

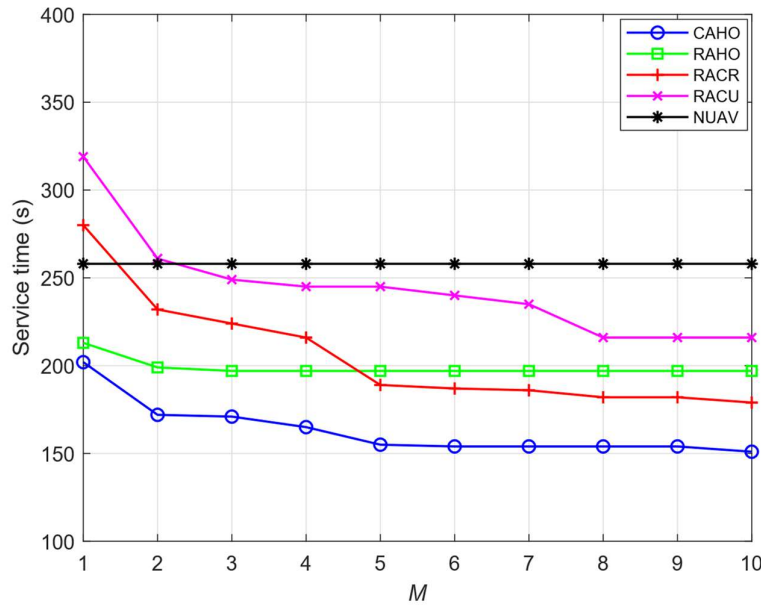
Hình 3.8 thể hiện sự thay đổi của thời gian phục vụ theo số lượng UAV M . Kết quả cho thấy, đối với tất cả các phương pháp có triển khai UAV, thời gian phục vụ giảm khi M tăng. Tuy nhiên, khi số lượng UAV đạt đến một ngưỡng nhất định, việc tăng thêm UAV không còn mang lại cải thiện đáng kể, do bị giới hạn bởi tài nguyên lưu trữ và các khoảng trống lưu trữ còn lại.

Hiệu quả khai thác UAV của các phương pháp khác nhau rõ rệt theo M . RAHO bão hòa sớm do nhiều UAV dừng tại các vị trí không có khoảng trống lưu trữ cần bù đắp. RACR khai thác UAV tốt hơn nhờ tối ưu b_n . CAHO duy trì hiệu quả biên cao nhất cho đến $M = 5$, đây là ngưỡng tại đó số khoảng trống lưu trữ tối đa ($N - J_i$) ứng với cấu hình h_i tối ưu đã được phủ kín. Tăng M vượt ngưỡng này không cải thiện thời gian phục vụ vì các UAV bổ sung không có vị trí có ý nghĩa để dừng.

Với phương pháp RAHO, hiện tượng bão hòa xảy ra sớm hơn, bởi vị trí dừng của UAV không được tối ưu, làm hạn chế khả năng khai thác hiệu quả số lượng UAV lớn. Ngược lại, RACR tận dụng UAV tốt hơn nhờ tối ưu vị trí dừng. Phương pháp NUAV, do không triển khai UAV, nên hiệu năng không thay đổi theo M . So với các

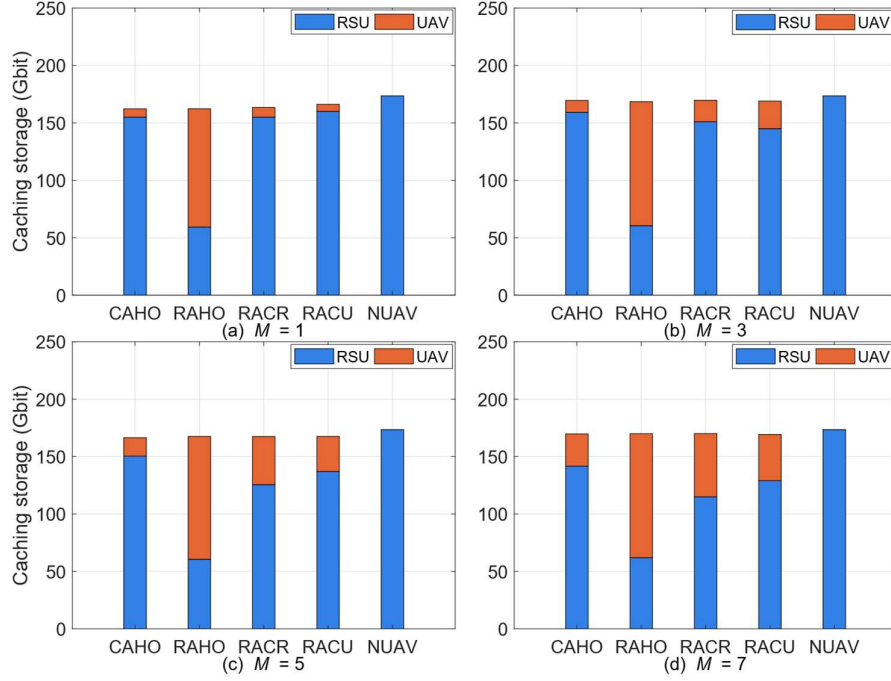
phương pháp còn lại, CAHO khai thác số lượng UAV hiệu quả nhất, đạt thời gian phục vụ thấp nhất trên toàn bộ dải giá trị M . Quan trọng là, số lượng UAV cần được lựa chọn cẩn thận cho từng kịch bản VANET, lý tưởng nhất là tại điểm “khuỷu bão hòa” (ví dụ, $M = 5$). Mức tiêu thụ bộ nhớ lưu trữ của tất cả các phương pháp theo M được khảo sát trong hình 3.9, trong đó xu hướng tương tự như trong hình 3.7. Ngoài ra, có thể thấy rõ rằng khi M tăng, các UAV tận dụng nhiều cơ hội hơn để đạt được dung lượng truyền U2V tốt hơn nhằm lấp đầy các khoảng trống lưu trữ, dẫn đến sự gia tăng dung lượng lưu trữ được sử dụng tại các UAV.

Hình 3.9 cho thấy CAHO tăng dần S^{UAV} theo M cho đến khi đạt ngưỡng bão hòa tại $M = 5$, sau đó S^{UAV} giữ ổn định và S^{RSU} giảm tương ứng. Hành vi này khớp với điểm khuỷu trong hình 3.8: khi số khoảng trống lưu trữ đã được phủ kín, GA không phân bổ thêm tài nguyên cho UAV bổ sung. Ngược lại, RAHO duy trì S^{UAV} cao trên toàn dải M do mỗi UAV ngẫu nhiên đều được cấp lưu trữ bất kể vị trí có hiệu quả hay không.



Hình 3.8: Thời gian phục vụ theo M

Khi M tăng đến ngưỡng $M = 5$, CAHO đạt điểm khuỷu bão hòa, ngưỡng tại đó số khoảng trống lưu trữ tối đa ($N - J_i$) ứng với cấu hình h_i tối ưu đã được phủ kín. Tăng M vượt ngưỡng này không cải thiện thời gian phục vụ vì các UAV bổ sung không có vị trí có ý nghĩa để dừng.

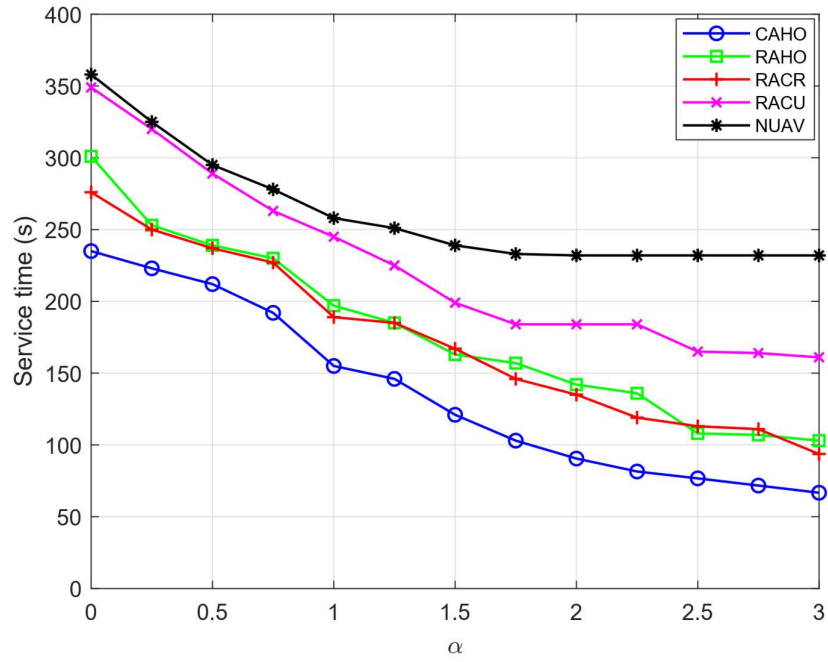


Hình 3.9: Dung lượng lưu trữ theo M

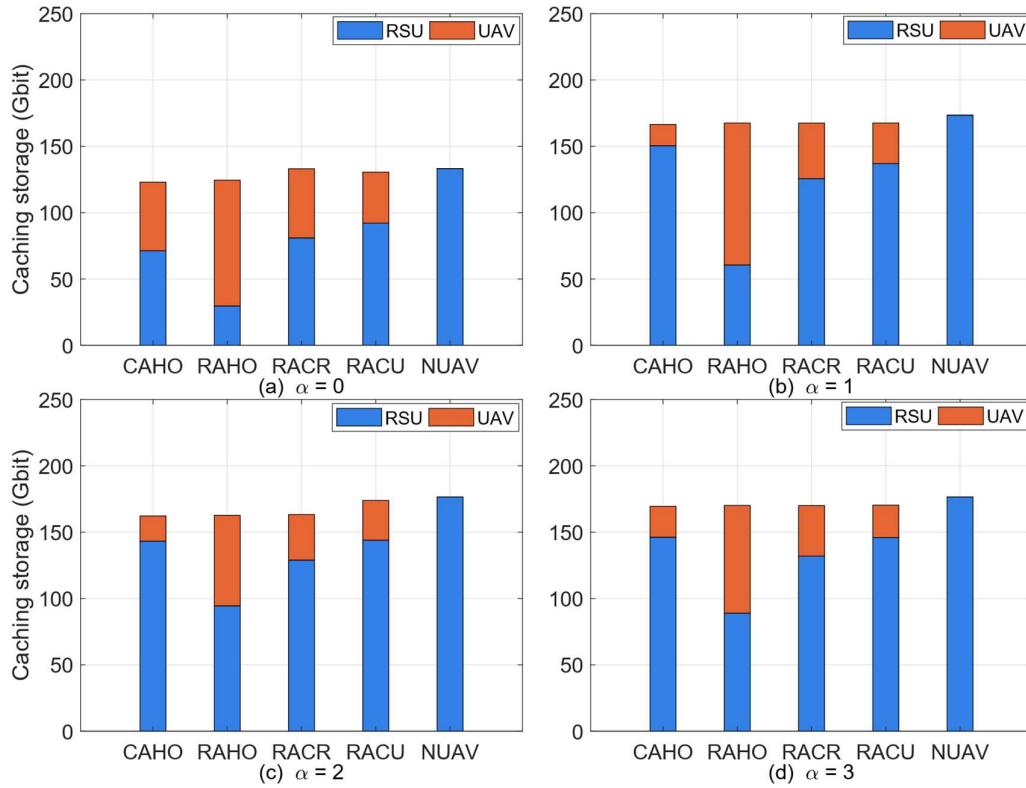
3) Hiệu năng theo mật độ phương tiện λ

Hình 3.10 minh họa thời gian phục vụ của các phương pháp theo mật độ phương tiện λ . Đối với các phương pháp có hỗ trợ UAV, thời gian phục vụ giảm khi λ tăng, do xác suất tồn tại các phương tiện yêu cầu cao hơn, giúp khai thác hiệu quả các liên kết U2V. Ngược lại, phương pháp NUAV có thời gian phục vụ tăng, do dung lượng R2V trong biểu thức (3.2) bị chia cho số lượng phương tiện yêu cầu lớn hơn. Khi mật độ phương tiện tăng, lợi thế của đường truyền có UAV hỗ trợ trở nên rõ rệt hơn, cho phép UAV bù đắp các hạn chế của truyền dẫn R2V trong phương pháp NUAV. Hệ quả là dung lượng lưu trữ tại UAV tăng theo λ , như thể hiện trong hình 3.11.

Khi λ tăng, thời gian phục vụ của CAHO giảm rõ rệt trong khi NUAV tăng (hình 3.10). Sự đối lập này phản ánh đặc tính kênh truyền của hai phương pháp: NUAV chỉ phục vụ qua R2V nên dung lượng trên mỗi RV giảm khi số RV chia sẻ băng thông tăng theo λ ; CAHO bổ sung kênh U2V có xác suất LoS cao và không bị cạnh tranh tài nguyên với R2V, do đó càng nhiều phương tiện thì việc khai thác U2V càng hiệu quả. Khoảng cách hiệu năng giữa CAHO và NUAV mở rộng theo λ là minh chứng định lượng cho vai trò bù đắp của UAV ở vùng mật độ cao.



Hình 3.10: Thời gian phục vụ theo λ



Hình 3.11: Dung lượng lưu trữ theo λ

Khi mật độ phương tiện λ tăng, dung lượng lưu trữ tại UAV trong mô hình CAHO tăng theo (hình 3.11), phản ánh việc hệ thống điều chuyển tài nguyên lưu trữ từ RSU sang UAV để khai thác liên kết U2V có xác suất LoS cao. Cơ chế này giải thích vì sao thời gian phục vụ của CAHO trong hình 3.10 giảm liên tục theo λ , trong khi NUAV thể hiện xu hướng ngược lại do dung lượng R2V bị chia sẻ giữa nhiều RV.

3.5.4. Độ phức tạp của mô hình CAHO

So sánh giải thuật 3.1 và 2.1 có thể thấy rằng, mô hình CAHO có độ phức tạp hay chi phí tính toán hàm mục tiêu C_0 của bài toán tương tự như mô hình DPC. Tuy nhiên, việc đưa tầng UAV vào hệ thống làm tăng số lượng biến quyết định cũng như kích thước không gian tìm kiếm. Kích thước không gian tìm kiếm tăng sẽ yêu cầu số cá thể N_P và số thế hệ N_G lớn hơn để đảm bảo GA hội tụ ổn định và chính xác. Đáng chú ý trong bài toán CAHO, số cá thể và số thế hệ được thiết lập là $N_P = 3000$ và $N_G = 100$ giống như bài toán DPC, đủ để đảm bảo tính ổn định và chính xác của GA. Và do đó, tương tự như DPC, độ phức tạp tính toán tổng thể của GA trong mô hình CAHO được xấp xỉ theo bậc $O(N_P \times N_G \times C_0)$. Mô hình CAHO vẫn có khả năng áp dụng cho các hệ thống UAV-VANET với quy mô lớn, đặc biệt khi kết hợp với các chiến lược giảm độ phức tạp hoặc các thuật toán tối ưu lai trong các nghiên cứu tiếp theo.

Đối với mô hình CAHO, độ phức tạp tăng cao hơn DPC do mô hình bổ sung tầng UAV vào bài toán tối ưu. Mỗi cá thể trong CAHO bao gồm I biến nguyên h_i , $N \times I$ biến nhị phân $c_{n,i}$ cho quyết định lưu trữ video tại UAV và N biến nhị phân b_n cho quyết định vị trí dừng UAV. Do đó, chiều dài nhiễm sắc thể có thể biểu diễn xấp xỉ là $(I + NI + N)$. Khi số lượng video I , số RSU N hoặc số UAV M tăng, không gian nghiệm của CAHO mở rộng mạnh theo dạng tổ hợp. Số UAV M ảnh hưởng đến số cách lựa chọn vị trí dừng UAV trong N vị trí ứng viên, có thể xấp xỉ bởi tổ hợp $C(N, M)$. Vì vậy, so với DPC, CAHO có chi phí tìm kiếm cao hơn và có thể yêu cầu số cá thể hoặc số thế hệ lớn hơn để đạt nghiệm ổn định trong các kịch bản quy mô lớn.

Do đặc tính hội tụ chậm và chi phí tính toán cao của GA, mô hình CAHO cũng phù hợp hơn với cơ chế tối ưu ngoại tuyến hoặc tái tối ưu định kỳ tại MBS/MEC. Trong các hệ thống UAV-VANET quy mô lớn, khả năng mở rộng có thể được cải thiện bằng

các hướng như giới hạn vùng ứng viên cho UAV, phân cụm RSU, tối ưu theo từng khu vực, khởi tạo nghiệm bằng heuristic, song song hóa đánh giá fitness hoặc kết hợp GA với các thuật toán tối ưu lai.

3.6. KẾT LUẬN CHƯƠNG 3

Trong chương này, NCS đã đề xuất một mô hình tối ưu kết hợp lưu trữ và vị trí dừng của UAV, được gọi là CAHO, nhằm hỗ trợ dịch vụ truyền video hợp tác trong mạng VANET có hỗ trợ UAV. Mô hình CAHO cho phép tối ưu đồng thời chiến lược lưu trữ tại các RSU, lưu trữ tại UAV và vị trí dừng của UAV, với mục tiêu giảm thiểu thời gian phục vụ trung bình của hệ thống trong điều kiện tài nguyên lưu trữ và số lượng UAV bị giới hạn. Trên cơ sở mô hình hóa phân bố các phương tiện yêu cầu bằng quá trình Poisson một chiều, kết hợp với các mô hình truyền R2V và U2V, bài toán tối ưu CAHO đã được xây dựng và giải bằng GA. Các điều chỉnh được thêm vào GA để giải quyết tính phức tạp của các biến tối ưu dạng số nguyên và số nhị phân. Kết quả đạt được chứng minh khả năng hội tụ, tính ổn định và tính chính xác của GA đã được điều chỉnh.

Các kết quả mô phỏng cho thấy CAHO vượt trội so với các phương pháp đối sánh trong nhiều kịch bản khác nhau, xét theo các tham số hệ thống như độ lệch phân bố độ phổ biến video, số lượng UAV và mật độ phương tiện. Cụ thể, phương pháp đề xuất đạt được thời gian phục vụ thấp hơn, đồng thời sử dụng tài nguyên lưu trữ hiệu quả hơn thông qua việc phân bổ hợp lý giữa RSU và UAV. Những kết quả này khẳng định vai trò quan trọng của UAV trong việc bù đắp các khoảng trống lưu trữ và nâng cao hiệu năng của hệ thống truyền video trong VANET. Trong các nghiên cứu tiếp theo, mô hình CAHO có thể được mở rộng theo nhiều hướng khác nhau, chẳng hạn như xem xét quỹ đạo bay động của UAV, tối ưu công suất truyền, hoặc phối hợp nhiều MBS để hỗ trợ các dịch vụ truyền video liên tục trong các kịch bản mạng quy mô lớn và phức tạp hơn.

KẾT LUẬN

1. Kết quả đạt được

Luận án giải quyết bài toán tối ưu lưu trữ và phân phối nội dung video trong mạng UAV-VANET, với tiêu chí trung tâm xuyên suốt là tối thiểu hóa thời gian phục vụ, là đại lượng phản ánh trực tiếp cả QoS và QoE của hệ thống và người dùng cuối. Khác với các nghiên cứu trước sử dụng các chỉ tiêu trung gian như cache hit ratio hay thông lượng, lựa chọn này cho phép luận án xây dựng một khung tối ưu thống nhất, trong đó, các quyết định về lưu trữ và cấu hình mạng đều quy về một mục tiêu chung có ý nghĩa thực tiễn.

Luận án đề xuất bài toán DPC tối ưu kết hợp lưu trữ xác định (DC) tại RSU và lưu trữ theo xác suất (PC) tại phương tiện trong một hàm mục tiêu thống nhất. Hai cơ chế DC và PC, vốn được xử lý riêng lẻ trong các nghiên cứu trước, được gắn kết qua biến tối ưu chung gồm bán kính lưu trữ h_i và xác suất lưu trữ ρ_i dưới ràng buộc tài nguyên hệ thống. Bài toán có cấu trúc tối ưu hỗn hợp nguyên–thực, được giải bằng GA với cơ chế mã hóa hai chuỗi con độc lập và toán tử di truyền tách biệt cho từng loại biến. Kết quả mô phỏng cho thấy DPC vượt trội so với các phương pháp đơn tầng OCV, OCR và đạt mức tiêu thụ tài nguyên lưu trữ hiệu quả hơn.

Mô hình CAHO mở rộng bài toán sang không gian ba chiều của UAV-VANET, trong đó UAV được định vị đúng vai trò là nút lưu trữ di động thay vì chỉ là trạm chuyển tiếp như nhiều nghiên cứu trước. CAHO lần đầu tiên tối ưu đồng thời ba thành phần gồm bán kính lưu trữ tại RSU (h_i), quyết định lưu trữ tại UAV ($c_{n,i}$) và vị trí dừng của UAV (b_n), trong cùng một khung tối ưu hướng tới cực tiểu thời gian phục vụ. Sự gắn kết chặt chẽ giữa quyết định lưu trữ và quyết định vị trí là điểm khác biệt so với các nghiên cứu trước, vốn tách rời hai khía cạnh này. Bài toán có cấu trúc tối ưu hỗn hợp nguyên–nhị phân với hai ràng buộc tổ hợp độc lập, được giải bằng GA điều chỉnh với hàm phạt hai thành phần.

DPC và CAHO không phải là hai mô hình độc lập mà là quá trình mở rộng logic, từ môi trường VANET hai chiều sang UAV-VANET ba chiều, với tiêu chí thời gian phục vụ và giải thuật tối ưu meta-heuristic là trục xuyên suốt. Cách xây dựng này tạo nên

một khung tiếp cận có khả năng mở rộng, trong đó mỗi bước kế thừa kết quả của bước trước và tiến gần hơn tới bối cảnh triển khai thực tế, theo từng kịch bản ứng dụng được trình bày trong bảng sau:

Mô hình	Kiến trúc	Mô hình truyền	Bài toán tối ưu	Kịch bản ứng dụng	Mục đích
DPC	2D	R2V và V2V	- Tại RSU: tìm bán kính lưu trữ video - Tại phương tiện: tìm xác suất lưu trữ video	Mật độ người dùng cao, RSU phân bố không đồng đều và hạn chế, có nhiều xe tham gia chia sẻ tài nguyên lưu trữ để tận dụng ưu thế truyền thông V2V bên cạnh R2V	Giảm thiểu thời gian phục vụ, đảm bảo dịch vụ truyền video liên tục và sử dụng hiệu quả tài nguyên lưu trữ của hệ thống
CAHO	3D	R2V và U2V	- Tại RSU: tìm bán kính lưu trữ video - Tại UAV: tìm vị trí lưu trữ video và vị trí dừng (hovering)	Mật độ người dùng cao, RSU phân bố không đồng đều và hạn chế, có ít xe tham gia chia sẻ tài nguyên lưu trữ nên phải cần kênh truyền thông U2V để hỗ trợ R2V	

Các đóng góp chính của luận án bao gồm:

- Xây dựng khung mô hình thời gian phục vụ video và giải thuật phân phối video cho hệ thống lưu trữ phân tầng trong UAV-VANET.
- Đề xuất hai bài toán tối ưu mới với cấu trúc biến hỗn hợp: DPC (nguyên–thực) và CAHO (nguyên–nhị phân).
- Phát triển hai phiên bản GA điều chỉnh phù hợp với đặc thù từng bài toán, có hàm phạt chặt chẽ và mã hóa tách chuỗi con phù hợp.
- Phân tích định lượng các phát hiện thiết kế quan trọng, bao gồm ngưỡng tối ưu của τ , hiện tượng bão hòa tại $M = 5$ và kịch bản sử dụng DPC hoặc CAHO.

Các kết quả của luận án có ý nghĩa cả về lý thuyết lẫn ứng dụng, mở ra tiềm năng triển khai trong các hệ thống MEC/VEC, mạng 5G/6G và hệ thống giao thông thông minh thế hệ tiếp theo.

2. Các hướng nghiên cứu tiếp theo

Các kết quả của luận án được xây dựng trên một số giả định nhằm tập trung vào mục tiêu và đóng góp chính. Những giả định này đồng thời xác định các hướng phát triển cụ thể trong tương lai, được tổ chức theo ba trục chính: độ tin cậy và đặc tính động của hệ thống, tối ưu đa mục tiêu, và phương pháp tối ưu.

Về độ tin cậy và đặc tính động của hệ thống: Mô hình hiện tại giả định RSU và UAV luôn sẵn sàng hoạt động, mật độ phương tiện và phân bố độ phổ biến nội dung là tĩnh (hoặc có đặc tính thống kê thay đổi theo giờ). Hướng mở rộng là xây dựng mô hình tối ưu bền vững, trong đó trạng thái hoạt động của các nút mạng được mô hình hóa qua chuỗi Markov hoặc biến ngẫu nhiên, và chiến lược lưu trữ thích nghi theo thời gian được cập nhật liên tục dựa trên trạng thái mạng quan sát được. Mô hình kênh truyền có thể được chi tiết hóa bằng cách tích hợp fading đa đường, shadowing và can nhiễu trong môi trường đô thị. Ngoài ra, việc tích hợp mô phỏng giao thông vi mô (SUMO, car-following) và các tập dữ liệu giao thông thực tế sẽ giúp kiểm chứng mô hình ở điều kiện vận hành sát thực tế. Việc xét sai số dự báo độ phổ biến và xây dựng cơ chế lưu trữ thích nghi theo thời gian thực là một hướng nghiên cứu tiếp theo của luận án.

Về tối ưu đa mục tiêu: Hàm mục tiêu hiện tại tập trung vào thời gian phục vụ, trong khi chi phí năng lượng và quỹ đạo bay của UAV chưa được xem xét. Hướng tiếp theo là phát triển bài toán tối ưu đa mục tiêu kết hợp đồng thời thời gian phục vụ, dung lượng lưu trữ, tiêu thụ năng lượng và quỹ đạo bay của UAV.

Về phương pháp tối ưu: GA cho thấy hiệu quả trong phạm vi luận án nhưng có chi phí tính toán tăng khi quy mô hệ thống mở rộng. Hai hướng cải tiến được xác định là phát triển thuật toán lai như Hybrid GA-PSO hoặc memetic algorithms để giảm độ phức tạp, và ứng dụng deep reinforcement learning để xây dựng cơ chế tối ưu thích nghi thời gian thực.

CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ CỦA NGHIÊN CỨU SINH

Công trình đã công bố của luận án

STT	Công trình đã công bố
[C1]	Nguyen-Son Vo, Quynh-Anh Nguyen, Thanh-Minh Phan, Thanh-Hieu Nguyen, and Thanh-Dung Tran, "Caching Placement Minimised Service Time for Video Streaming over Roadside Units in VANETs," <i>Journal of Aviation Science & Technology</i> , vol. 1, no. 2, pp. 1-6, Nov. 2022. Tạp chí Quốc gia – ISSN: 2815-5661.
[C2]	Quynh-Anh Nguyen, Nguyen-Son Vo, Thuong C. Lam, Tan Do-Duy, Haejoon Jung, and Trung Q. Duong, "Joint Deterministic and Probabilistic Edge Caching Minimized Service Time for Cooperative Video Transmission in VANETs," <i>IEEE Transactions on Network Science and Engineering</i> , vol. 13, pp. 1932-1943, Sep. 2025. Tạp chí Quốc tế - SCIE/Q1 – ISSN: 2327-4697.
[C3]	Quynh-Anh Nguyen, Nguyen-Son Vo, Van-Ca Phan, Dac-Binh Ha, Minh-Phung Bui, and Long D. Nguyen, "Joint Caching and Hovering Minimized Service Time for Cooperative Video Streaming in UAV-assisted VANETs," <i>REV Journal on Electronics and Communications</i> , vol. 16, no. 1, pp. 24-35, Jan.-Mar. 2026. Tạp chí Quốc gia - ISSN: 1859-378X.

Các công trình khác

STT	Công trình đã công bố
[C4]	Quang-Nhat Tran, Nguyen-Son Vo, Minh-Phung Bui, Thanh-Minh Phan, Quynh-Anh Nguyen and Trung Q. Duong, "Spectrum Sharing and Power Allocation Optimised Multihop Multipath D2D Video Delivery in Beyond 5G Networks," <i>IEEE Transactions on Cognitive Communications and Networking</i> , vol. 8, no. 2, pp. 919-930, June 2022.
[C5]	Quang-Nhat Tran, Nguyen-Son Vo, Quynh-Anh Nguyen , Minh-Phung Bui, Thanh-Minh Phan, Van-Viet Lam, Antonino Masaracchia, " D2D multi-hop multipath communications in B5G networks: A survey on models, techniques, and applications," <i>EAI Endorsed Transactions on Industrial Networks and Intelligent Systems</i> , Vol. 7, Issue 25, 2021.

TÀI LIỆU THAM KHẢO

- [1] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, “Vehicular Fog Computing: A Viewpoint of Vehicles as the Infrastructures,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3860–3873, Jun. 2016, doi: 10.1109/TVT.2016.2532863.
- [2] H. Hartenstein and K. P. Laberteaux, “A tutorial survey on vehicular ad hoc networks,” *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, Jun. 2008, doi: 10.1109/MCOM.2008.4539481.
- [3] D. E. F. Dressler and C. Sommer, “Vehicular Networking - Christoph Sommer, Falko Dressler - Google Books.”. [Online]. Available: https://books.google.com.vn/books/about/Vehicular_Networking.html?id=xzDCBQAAQBAJ&redir_esc=y
- [4] N. Cheng *et al.*, “Air-ground integrated mobile edge networks: Architecture, challenges, and opportunities,” *IEEE Communications Magazine*, vol. 56, no. 8, pp. 26–32, Aug. 2018, doi: 10.1109/MCOM.2018.1701092.
- [5] N. H. Hussein, C. T. Yaw, S. P. Koh, S. K. Tiong, and K. H. Chong, “A Comprehensive Survey on Vehicular Networking: Communications, Applications, Challenges, and Upcoming Research Directions,” *IEEE Access*, vol. 10, pp. 86127–86180, Aug. 2022, doi: 10.1109/ACCESS.2022.3198656.
- [6] A. Raza, S. H. R. Bukhari, F. Aadil, and Z. Iqbal, “An UAV-VANET architecture for intelligent transportation system in smart cities,” *International Journal of Distributed Sensor Networks*, vol. 17, no. 7, Jul. 2021, doi: 10.1177/15501477211031750.
- [7] M. Mozaffari, W. Saad, M. Bennis, Y. H. Nam, and M. Debbah, “A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems,” *IEEE Communications Surveys and Tutorials*, vol. 21, no. 3, pp. 2334–2360, Mar. 2019, doi: 10.1109/COMST.2019.2902862.
- [8] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya, “A survey on vehicular cloud computing,” *Journal of Network and Computer Applications*, vol. 40, no. 1, pp. 325–344, Apr. 2014, doi: 10.1016/j.jnca.2013.08.004.
- [9] X. Hou, Y. Li, D. Jin, D. O. Wu, and S. Chen, “Modeling the impact of mobility on the connectivity of vehicular networks in large-scale urban environments,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2753–2758, Apr. 2016, doi: 10.1109/TVT.2015.2418574.

- [10] A. Khanpour *et al.*, “UAV-Based Intelligent Traffic Surveillance System: Real-Time Vehicle Detection, Classification, Tracking, and Behavioral Analysis,” *arxiv.org*, pp. 1-15, Sep. 2025, doi: 10.48550/arXiv.2509.04624
- [11] M. M. Ashraf, S. Boudjit, S. Zeadally, N. E. H. Bahloul, and N. Bashir, “Integrating Unmanned Aerial Vehicles (UAVs) with Vehicular Ad-hoc NETWORKS (VANETs): Architectures, applications, opportunities,” *Computer Networks*, vol. 255, pp. 110873, Dec. 2024, doi: 10.1016/j.comnet.2024.110873.
- [12] C. Zhao, Q. Zeng, Y. Tang, and B. Yang, “Multi-Path Routing Protocol for the Video Service in UAV-VANETs,” in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, Norman, OK, USA, Sep. 2021, pp. 1-5, doi: 10.1109/VTC2021-Fall52928.2021.9625278.
- [13] C. J. Huang, L. C. Chen, Y. Sen Cheng, K. W. Hu, and M. E. Jian, “Enhancing Infotainment Services in Integrated Aerial–Ground Mobility Networks,” *Sensors* 2025, vol. 25, no. 13, pp. 3891, Jun. 2025, doi: 10.3390/s25133891.
- [14] F. Zeng, R. Zhang, X. Cheng, and L. Yang, “UAV- data dissemination scheduling in VANETs,” in *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, May. 2018, pp. 1-6, doi: 10.1109/ICC.2018.8422219.
- [15] M. Alshehri *et al.*, “UAV-based intelligent traffic surveillance using recurrent neural networks and Swin transformer for dynamic environments,” *Frontiers in Neurorobotics*, vol. 19, pp. 1681341, Oct. 2025, doi: 10.3389/fnbot.2025.1681341.
- [16] Z. Fan *et al.*, “A Novel UAV-VANET Routing Protocol for Post-Disaster Emergency Communications,” *IEEE Transactions on Network Science and Engineering*, vol. 13, pp. 4863–4882, Dec. 2025, doi: 10.1109/TNSE.2025.3644432.
- [17] O. S. Oubbati, N. Chaib, A. Lakas, P. Lorenz, and A. Rachedi, “UAV- Supporting Services Connectivity in Urban VANETs,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3944–3951, Apr. 2019, doi: 10.1109/TVT.2019.2898477.
- [18] A. D. Sree and K. Sharma, “Machine Learning Based Secure Routing Protocol with UAV- for Autonomous Vehicles,” *International Journal of Sensors, Wireless Communications and Control*, vol. 14, no. 1, pp. 55–65, Feb. 2024, doi: 10.2174/0122103279273609231213075003.
- [19] M. M. V. B. Murali Krishna, “A Survey UAV-VANET Routing Protocol,” *International Journal of Computer Science Trends and Technology*, vol. 8, pp. 1-

- 7, Aug. 2020, Available: <https://www.ijcstjournal.org/volume-8/issue-4/IJCST-V8I4P12.pdf>.
- [20] A. D. Sree and K. Sharma, "Machine Learning Based Secure Routing Protocol with UAV- for Autonomous Vehicles," *International Journal of Sensors, Wireless Communications and Control*, vol. 14, no. 1, pp. 55–65, Feb. 2024, doi: 10.2174/0122103279273609231213075003.
 - [21] N. Hu, X. Qin, N. Ma, Y. Liu, Y. Yao, and P. Zhang, "Energy-efficient Caching and Task offloading for Timely Status Updates in UAV-VANETs," *2022 IEEE/CIC International Conference on Communications in China, ICCIC*, Aug. 2022, pp. 1032–1037, doi: 10.1109/ICCC55456.2022.9880683.
 - [22] Y. Sun, X. Zhong, F. Wu, X. Chen, S. Zhang, and N. Dong, "Multi-UAV Content Caching Strategy and Cooperative, Complementary Content Transmission Based on Coalition Formation Game," *Sensors* 2022, vol. 22, no. 9, pp. 3123, Apr. 2022, doi: 10.3390/s22093123.
 - [23] W. Wang, Y. Liu, Y. Dai, and Y. He, "Joint cooperative caching and power control for UAV- internet of vehicles," *Scientific Reports* 2024, vol. 14, no. 1, pp. 9341, Apr. 2024, doi: 10.1038/s41598-024-59823-9.
 - [24] O. B. Piramuthu and M. Caesar, "UAV-VANET authentication for real-time highway surveillance," *Proceedings of the ACM Symposium on Applied Computing*, pp. 1925–1931, Apr. 2022, doi: 10.1145/3477314.3507021.
 - [25] Cisco, "Cisco Annual Internet Report - Cisco Annual Internet Report (2018–2023) White Paper - Cisco," Online. Accessed: Feb. 04, 2026. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
 - [26] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update," 2019.
 - [27] S. Berri, J. Zhang, B. Bensaou, and H. Labiod, "Joint Content-Prefetching, Transmission Scheduling, and Rate Adaptation in Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4348–4358, Apr. 2022, doi: 10.1109/TVT.2022.3147979.
 - [28] C. Tang, C. Zhu, X. Wei, Q. Li, and J. J. P. C. Rodrigues, "Task caching in vehicular edge computing," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Vancouver, BC, Canada*, May. 2021, pp. 1–6, doi: 10.1109/INFOCOMWKSHPS51825.2021.9484498.

- [29] Q. Shen, B. J. Hu, and E. Xia, "Dependency-Aware Task Offloading and Service Caching in Vehicular Edge Computing," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 12, pp. 13182–13197, Dec. 2022, doi: 10.1109/TVT.2022.3196544.
- [30] T. Q. Duong, K. J. Kim, Z. Kaleem, M. P. Bui, and N. S. Vo, "UAV caching in 6G networks: A Survey on models, techniques, and applications," *Physical Communication*, vol. 51, pp. 101532, Apr. 2022, doi: 10.1016/j.phycom.2021.101532.
- [31] Q. A. Nguyen, N. S. Vo, T. C. Lam, T. Do-Duy, H. Jung, and T. Q. Duong, "Joint Deterministic and Probabilistic Edge Caching Minimized Service Time for Cooperative Video Transmission in VANETs," *IEEE Transactions on Network Science and Engineering*, vol. 13, pp. 1932–1943, Sep. 2025, doi: 10.1109/TNSE.2025.3605822.
- [32] M. Asefi, J. W. Mark, and X. Shen, "A mobility-aware and quality-driven retransmission limit adaptation scheme for video streaming over VANETs," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011, Houston, TX, USA*, Dec. 2011, pp. 1-5, doi: 10.1109/GLOCOM.2011.6134469.
- [33] B. Feng, C. Feng, G. Min, and T. Q. S. Quek, "Two-Timescale Adaptive Live Video Streaming Transmission Mechanism for Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 4, pp. 6823–6828, Apr. 2025, doi: 10.1109/TVT.2024.3507816.
- [34] J. Chen, H. Wu, P. Yang, F. Lyu, and X. Shen, "Cooperative Edge Caching with Location-Based and Popular Contents for Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10291–10305, Sep. 2020, doi: 10.1109/TVT.2020.3004720.
- [35] Y. Guo, Q. Yang, F. R. Yu, and V. C. M. Leung, "Cache-Enabled Adaptive Video Streaming over Vehicular Networks: A Dynamic Approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5445–5459, Jun. 2018, doi: 10.1109/TVT.2018.2817210.
- [36] T. Gong and Q. Zhu, "Geolocation-based UAV- Caching Strategies in Vehicular Networks," *Journal of Internet Technology*, vol. 24, no. 7, pp. 1457–1468, Dec. 2023, doi: 10.53106/160792642023122407007.
- [37] Q.-A. Nguyen, N.-S. Vo, V.-C. Phan, D.-B. Ha, M.-P. Bui, and D.-L. Nguyen, "Joint Caching and Hovering Minimized Service Time for Cooperative Video Streaming in UAV-VANETs," *REV Journal on Electronics and Communications*, vol. 16, no. 1, pp. 1-12, Feb. 2026, doi: 10.21553/REV-JEC.430.

- [38] M. I. Mushtaq, O. Chughtai, Y. Zhang, A. H. Alenezi, and N. Alqahtani, "Framework for Optimized Resource Allocation in Multi-User, Multi-Service, Multi-Device Aerial Networks," *IEEE Access*, vol. 12, pp. 54866–54878, Apr. 2024, doi: 10.1109/ACCESS.2024.3384562.
- [39] T. Fang and L. P. Chau, "GOP-based channel rate allocation using genetic algorithm for scalable video streaming over error-prone networks," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1323–1330, Jun. 2006, doi: 10.1109/TIP.2005.864159.
- [40] N. S. Vo, T. Q. Duong, H. D. Tuan, and A. Kortun, "Optimal Video Streaming in Dense 5G Networks With D2D Communications," *IEEE Access*, vol. 6, pp. 209–223, Oct. 2017, doi: 10.1109/ACCESS.2017.2761978.
- [41] H. Zhang, X. Cao, J. K. L. Ho, and T. W. S. Chow, "Object-Level Video Advertising: An Optimization Framework," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 520–531, Apr. 2017, doi: 10.1109/TII.2016.2605629.
- [42] J. Mattingley and S. Boyd, "Automatic Code Generation for Real-Time Convex Optimization", Chapter 1 of *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, Eds., Cambridge University Press, 2010, pages 1–41.
- [43] L. D. Nguyen, A. Kortun, and T. Q. Duong, "An Introduction of Real-time Embedded Optimisation Programming for UAV Systems under Disaster Communication," *European Alliance for Innovation Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 5, no. 17, pp. 1–8, Dec. 2018, doi: 10.4108/eai.19-12-2018.156080.
- [44] B. T. Tinh, L. D. Nguyen, H. H. Kha, and T. Q. Duong, "Practical Optimization and Game Theory for 6G Ultra-Dense Networks: Overview and Research Challenges," *IEEE Access*, vol. 10, pp. 13311–13328, Jan. 2022, doi: 10.1109/ACCESS.2022.3146335.
- [45] E. K. P. Chong and S. H. Zak, "An Introduction to Optimization: Third Edition," *An Introduction to Optimization: Third Edition*, pp. 1–586, Oct. 2011, doi: 10.1002/9781118033340.
- [46] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," *Proceedings - IEEE INFOCOM*, vol. 1, pp. 126–134, Mar. 1999, doi: 10.1109/INFCOM.1999.749260.
- [47] X. Chen *et al.*, "Vehicle as a Service (VaaS): Leverage Vehicles to Build Service Networks and Capabilities for Smart Cities," *IEEE Communications Surveys and*

- Tutorials*, vol. 26, no. 3, pp. 2048–2081, Feb. 2024, doi: 10.1109/COMST.2024.3370169.
- [48] P. Laclau, S. Bonnet, B. Ducourthial, X. Li, and T. Lin, “Enhancing Automotive User Experience With Dynamic Service Orchestration for Software Defined Vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 1, pp. 824–834, Nov. 2024, doi: 10.1109/TITS.2024.3485487.
 - [49] F. Zeng, R. Zhang, X. Cheng, and L. Yang, “Channel Prediction Based Scheduling for Data Dissemination in VANETs,” *IEEE Communications Letters*, vol. 21, no. 6, pp. 1409–1412, Jun. 2017, doi: 10.1109/LCOMM.2017.2676766.
 - [50] H. Tian *et al.*, “CoPace: Edge Computation Offloading and Caching for Self-Driving with Deep Reinforcement Learning,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13281–13293, Dec. 2021, doi: 10.1109/TVT.2021.3121096.
 - [51] W. Yang and Z. Liu, “Efficient Vehicular Edge Computing: A Novel Approach with Asynchronous Federated and Deep Reinforcement Learning for Content Caching in VEC,” *IEEE Access*, vol. 12, pp. 13196–13212, Jan. 2024, doi: 10.1109/ACCESS.2024.3355462.
 - [52] S. A. Elsayed, S. Abdelhamid, and H. S. Hassanein, “Predictive Proactive Caching in VANETs for Social Networking,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5298–5313, May. 2022, doi: 10.1109/TVT.2022.3148720.
 - [53] T. C. Lam, N. S. Vo, M. P. Bui, C. D. T. Thai, H. Jung, and V. C. Phan, “Service Time-Aware Caching, Power Allocation, and 3D Trajectory Optimised Multimedia Content Delivery in UAV- IoT Networks,” *IEEE Transactions on Vehicular Technology*, vol. 74, no. 4, pp. 6419–6432, Apr. 2025, doi: 10.1109/TVT.2024.3510621.
 - [54] L. Zhao, H. Li, N. Lin, M. Lin, C. Fan, and J. Shi, “Intelligent Content Caching Strategy in Autonomous Driving Toward 6G,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9786–9796, Jul. 2022, doi: 10.1109/TITS.2021.3114199.
 - [55] Z. Lin, Y. Fang, P. Chen, F. Chen, and G. Zhang, “Modeling and Analysis of Edge Caching for 6G mmWave Vehicular Networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7422–7434, Jul. 2023, doi: 10.1109/TITS.2022.3147696.

- [56] X. Zhou *et al.*, “Edge Computation Offloading With Content Caching in 6G-Enabled IoV,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 3, pp. 2733–2747, Mar. 2024, doi: 10.1109/TITS.2023.3239599.
- [57] Y. Wang, H. Dai, X. Han, P. Wang, Y. Zhang, and C. Xu, “Cost-Driven Data Caching in Edge-Based Content Delivery Networks,” *IEEE Transactions on Mobile Computing*, vol. 22, no. 3, pp. 1384–1400, Mar. 2023, doi: 10.1109/TMC.2021.3108150.
- [58] T. C. Lam, N. S. Vo, V. V. Lam, T. Hoang, M. P. Bui, and T. Q. Duong, “Multi-rate selection and power allocation assisted probabilistic edge caching for cooperative video transmission in dense D2D networks,” *Alexandria Engineering Journal*, vol. 126, no. 20, pp. 555–564, Jul. 2025, doi: 10.1016/j.aej.2025.04.057.
- [59] S. Zhang, T. Cai, D. Wu, D. Schupke, N. Ansari, and C. Cavdar, “IoRT Data Collection With LEO Satellite-Assisted and Cache-Enabled UAV: A Deep Reinforcement Learning Approach,” *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 5872–5884, Apr. 2024, doi: 10.1109/TVT.2023.3336262.
- [60] M. H. T. Nguyen *et al.*, “Real-Time Optimized Clustering and Caching for 6G Satellite-UAV-Terrestrial Networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 3, pp. 3009–3019, Mar. 2024, doi: 10.1109/TITS.2023.3287279.
- [61] Y. Zhao *et al.*, “Joint Content Caching, Service Placement, and Task Offloading in UAV-Enabled Mobile Edge Computing Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 1, pp. 51–63, 2025, doi: 10.1109/JSAC.2024.3460049.
- [62] H. Wang *et al.*, “CAAF: An NDN-Based Cache-Aware Adaptive Forwarding Strategy for Reliable Content Delivery in VANETs,” *IEEE Transactions on Mobile Computing*, vol. 24, no. 7, pp. 6361–6375, Jul. 2025, doi: 10.1109/TMC.2025.3543458.
- [63] T. Cao, Z. Zhang, X. Wang, H. Xiao, and C. Xu, “PTCC: A Privacy-Preserving and Trajectory Clustering-Based Approach for Cooperative Caching Optimization in Vehicular Networks,” *IEEE Transactions on Sustainable Computing*, vol. 9, no. 4, pp. 615–630, Aug. 2024, doi: 10.1109/TSUSC.2024.3350386.
- [64] Y. T. Wang *et al.*, “Markov Clustering-Based Content Placement in Roadside-Unit Caching With Deadline Constraint,” *IEEE Transactions on Intelligent*

- Transportation Systems*, vol. 25, no. 9, pp. 11881–11892, Sep. 2024, doi: 10.1109/TITS.2024.3368413.
- [65] Z. Jin, T. Song, W. Jiang, and J. Hu, “A Centralized Edge Cooperative Caching Strategy for VANETs,” *2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United Arab Emirates*, Apr. 2024, pp. 1-6, doi: 10.1109/WCNC57260.2024.10570522.
 - [66] Z. Su, Y. Hui, Q. Xu, T. Yang, J. Liu, and Y. Jia, “An edge caching scheme to distribute content in vehicular networks,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5346–5356, Jun. 2018, doi: 10.1109/TVT.2018.2824345.
 - [67] Y. Zhang, C. Li, T. H. Luan, Y. Fu, W. Shi, and L. Zhu, “A Mobility-Aware Vehicular Caching Scheme in Content Centric Networks: Model and Optimization,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3100–3112, Apr. 2019, doi: 10.1109/TVT.2019.2899923.
 - [68] Z. Zhao, L. Guardalben, M. Karimzadeh, J. Silva, T. Braun, and S. Sargento, “Mobility prediction-assisted over-the-top edge prefetching for hierarchical VANETs,” *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 8, pp. 1786–1801, Aug. 2018, doi: 10.1109/JSAC.2018.2844681.
 - [69] B. Feng, C. Feng, D. Feng, Y. Wu, and X. G. Xia, “Proactive Content Caching Scheme in Urban Vehicular Networks,” *IEEE Transactions on Communications*, vol. 71, no. 7, pp. 4165–4180, Jul. 2023, doi: 10.1109/TCOMM.2023.3277530.
 - [70] Y. Alnagar, R. H. Gohary, S. Hosny, and A. A. El-Sherif, “Mobility-Aware Edge Caching for Minimizing Latency in Vehicular Networks,” *IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 68–84, Feb. 2022, doi: 10.1109/OJVT.2022.3150241.
 - [71] G. Deng, L. Wang, F. Li, and R. Li, “Distributed Probabilistic Caching strategy in VANETs through Named Data Networking,” in *2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), San Francisco, CA*, Sep. 2016, pp. 314–319, doi: 10.1109/INFCOMW.2016.7562093.
 - [72] W. Zhao, Y. Qin, D. Gao, C. H. Foh, and H. C. Chao, “An Efficient Cache Strategy in Information Centric Networking Vehicle-to-Vehicle Scenario,” *IEEE Access*, vol. 5, pp. 12657–12667, Jun. 2017, doi: 10.1109/ACCESS.2017.2714191.
 - [73] S. A. Elsayed, S. Abdelhamid, and H. S. Hassanein, “Probabilistic Cooperative Caching in VANETs for Social Networking,” in *2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates*, Dec. 2018, pp. 1-7, doi: 10.1109/GLOCOM.2018.8647302.

- [74] G. Manzo *et al.*, “DeepNDN: Opportunistic Data Replication and Caching in Support of Vehicular Named Data,” in *2020 IEEE 21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Cork, Ireland*, Aug. 2020, pp. 234–243, doi: 10.1109/WoWMoM49955.2020.00051.
- [75] L. Yao, Y. Wang, X. Wang, and G. Wu, “Cooperative Caching in Vehicular Content Centric Network Based on Social Attributes and Mobility,” *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 391–402, Feb. 2021, doi: 10.1109/TMC.2019.2944829.
- [76] H. Wu, Y. Fan, J. Jin, H. Ma, and L. Xing, “Social-Aware Decentralized Cooperative Caching for Internet of Vehicles,” *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14834–14845, Aug. 2023, doi: 10.1109/JIOT.2022.3229009.
- [77] R. Wang, Z. Kan, Y. Cui, D. Wu, and Y. Zhen, “Cooperative Caching Strategy with Content Request Prediction in Internet of Vehicles,” *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8964–8975, Jun. 2021, doi: 10.1109/JIOT.2021.3056084.
- [78] Z. Xue, Y. Liu, G. Han, F. Ayaz, Z. Sheng, and Y. Wang, “Two-Layer Distributed Content Caching for Infotainment Applications in VANETs,” *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 1696–1711, Feb. 2022, doi: 10.1109/JIOT.2021.3089280.
- [79] C. Wang, C. Chen, Q. Pei, N. Lv, and H. Song, “Popularity Incentive Caching for Vehicular Named Data Networking,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3640–3653, Apr. 2022, doi: 10.1109/TITS.2020.3038924.
- [80] C. Li, Y. Zhang, and Y. Luo, “A Federated Learning-Based Edge Caching Approach for Mobile Edge Computing-Enabled Intelligent Connected Vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3360–3369, Mar. 2023, doi: 10.1109/TITS.2022.3224395.
- [81] H. Wu, B. Wang, H. Ma, X. Zhang, and L. Xing, “Multiagent Federated Deep-Reinforcement-Learning-Based Collaborative Caching Strategy for Vehicular Edge Networks,” *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 25198–25212, Jul. 2024, doi: 10.1109/JIOT.2024.3392329.
- [82] J. Huang, D. Fang, Y. Qian, and R. Q. Hu, “Recent advances and challenges in security and privacy for V2X communications,” *IEEE Open Journal of Vehicular Technology*, vol. 1, pp. 244–266, Jun. 2020, doi: 10.1109/OJVT.2020.2999885.
- [83] B. Bandyopadhyay, P. Kuila, and M. Chandra Govil, “Popularity-Conscious Service Caching and Offloading in Digital Twin and NOMA-Aided Connected

- Autonomous Vehicular Systems,” *IEEE Transactions on Network and Service Management*, vol. 21, no. 6, pp. 6451–6464, Dec. 2024, doi: 10.1109/TNSM.2024.3462481.
- [84] R. R. Neetu, G. Ghatak, V. A. Bohara, and A. Srivastava, “Performance Analysis of Cache-Enabled Handover Management for Vehicular Networks,” *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 1, pp. 1151–1164, Jan. 2024, doi: 10.1109/TNSE.2023.3321296.
- [85] K. Zhang, J. Cao, S. Maharjan, and Y. Zhang, “Digital Twin Empowered Content Caching in Social-Aware Vehicular Edge Networks,” *IEEE Transactions on Computational Social Systems.*, vol. 9, no. 1, pp. 239–251, Feb. 2022, doi: 10.1109/TCSS.2021.3068369.
- [86] H. Yan, X. Xu, M. Bilal, X. Xia, W. Dou, and H. Wang, “Customer Centric Service Caching for Intelligent Cyber-Physical Transportation Systems with Cloud-Edge Computing Leveraging Digital Twins,” *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 1787–1797, Feb. 2024, doi: 10.1109/TCE.2023.3326969.
- [87] J. Zheng, T. H. Luan, G. Li, Z. Yin, Y. Wu, and M. Dong, “ACDV: Adaptive Content Delivery for Vehicular Digital Twin Networks,” *IEEE Transactions on Vehicular Technology*, vol. 74, no. 5, pp. 7084–7098, May. 2025, doi: 10.1109/TVT.2024.3517657.
- [88] B. Mao, Y. Liu, J. Liu, and N. Kato, “AI-Assisted Edge Caching for Metaverse of Connected and Automated Vehicles: Proposal, Challenges, and Future Perspectives,” *IEEE Vehicular Technology Magazine*, vol. 18, no. 4, pp. 66–74, Dec. 2023, doi: 10.1109/MVT.2023.3327514.
- [89] J. P. Jeyaraj, M. Haenggi, A. H. Sakr, and H. Lu, “The Transdimensional Poisson Process for Vehicular Network Analysis,” *IEEE Transactions on Wireless Communications.*, vol. 20, no. 12, pp. 8023–8038, Dec. 2021, doi: 10.1109/TWC.2021.3089553.
- [90] S. Zhao, Y. Zhang, Y. Zhu, Z. Zhao, and Y. Liu, “Optimal Probabilistic Collaborative Caching in UAV- Vehicular Networks,” *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China*, Oct. 2022, pp. 377–382, doi: 10.1109/ITSC55140.2022.9921794.
- [91] N. S. Vo, T. Q. Duong, L. Shu, X. Du, H. J. Zepernick, and W. Cheng, “Cross-layer design for video replication strategy over multihop wireless networks,” *2011 IEEE International Conference on Communications (ICC), Kyoto, Japan*, Jul. 2011, pp. 1–6, doi: 10.1109/icc.2011.5962536.

- [92] S. Zhu *et al.*, “Probability distribution of rician K-factor in urban, suburban and rural areas using real-world captured data,” *IEEE Transactions Antennas Propagation*, vol. 62, no. 7, pp. 3835–3839, Jul. 2014, doi: 10.1109/TAP.2014.2318072.
- [93] B. Hu, L. Fang, X. Cheng, and L. Yang, “Vehicle-to-vehicle distributed storage in vehicular networks,” *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, May. 2018, pp. 1-6, doi: 10.1109/ICC.2018.8422220.
- [94] B. Hu, L. Fang, X. Cheng, and L. Yang, “In-vehicle caching (IV-Cache) via dynamic distributed storage relay (D2SR) in vehicular networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 843–855, Jan. 2019, doi: 10.1109/TVT.2018.2880969.
- [95] S. Katoch, S. S. Chauhan, and V. Kumar, “A review on genetic algorithm: past, present, and future,” *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, Feb. 2021, doi: 10.1007/s11042-020-10139-6.
- [96] X. Qi, S. Khattak, A. Zaib, and I. Khan, “Energy Efficient Resource Allocation for 5G Heterogeneous Networks Using Genetic Algorithm,” *IEEE Access*, vol. 9, pp. 160510–160520, Nov. 2021, doi: 10.1109/ACCESS.2021.3131823.
- [97] H. Li, S. Wu, J. Jiao, X. H. Lin, N. Zhang, and Q. Zhang, “Energy-Efficient Task Offloading of Edge-Aided Maritime UAV Systems,” *IEEE Transactions on Vehicular Technology*, vol. 72, no. 1, pp. 1116–1126, Jan. 2023, doi: 10.1109/TVT.2022.3205127.
- [98] G. Lai, F. F. Leymarie, and W. Latham, “On Mixed-Initiative Content Creation for Video Games,” *IEEE Transaction on Games*, vol. 14, no. 4, pp. 543–557, Dec. 2022, doi: 10.1109/TG.2022.3176215.
- [99] A. Chipperfield, P. Fleming, H. Pohlheim and C. Fonseca, “Genetic Algorithms Toolbox,” University of Sheffield. Accessed: Feb. 05, 2026. [Online]. Available: <https://uos-codem.github.io/GA-Toolbox/>
- [100] H. Feng, S. Guo, L. Yang, and Y. Yang, “Collaborative Data Caching and Computation Offloading for Multi-Service Mobile Edge Computing,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9408–9422, Sep. 2021, doi: 10.1109/TVT.2021.3099303.
- [101] Z. Ning *et al.*, “Intelligent Edge Computing in Internet of Vehicles: A Joint Computation Offloading and Caching Solution,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2212–2225, Apr. 2021, doi: 10.1109/TITS.2020.2997832.

- [102] R. I. Bor-Yaliniz, A. El-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," *2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia*, May. 2016, pp. 1-5, doi: 10.1109/ICC.2016.7510820.
- [103] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *2014 IEEE Global Communications Conference, Austin, TX, USA*, Dec. 2014, pp. 2898–2904, doi: 10.1109/GLOCOM.2014.7037248.
- [104] K. Xiao, K. Feng, A. Dong, and Z. Mei, "Efficient Data Dissemination Strategy for UAV in UAV-VANETs," *IEEE Access*, vol. 11, pp. 40809–40819, Apr. 2023, doi: 10.1109/ACCESS.2023.3270279.
- [105] A. Al-Hilo, M. Samir, C. Assi, S. Sharafeddine, and D. Ebrahimi, "UAV- Content Delivery in Intelligent Transportation Systems-Joint Trajectory Planning and Cache Management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 5155–5167, Aug. 2021, doi: 10.1109/TITS.2020.3020220.
- [106] G. Chen, Y. Han, F. Shen, Q. Zeng, and Y. D. Zhang, "Edge Collaborative Caching Based on Incentive-Driven D3QN Combined With User Preferences in UAV-Vehicular Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 12, pp. 22945–22961, Dec. 2025, doi: 10.1109/TITS.2025.3604600.
- [107] W. Saad, M. Bennis, M. Mozaffari, and X. Lin, "Resource Management for UAV Networks," *Wireless Communications and Networking for Unmanned Aerial Vehicles*, pp. 145–180, Apr. 2020, doi: 10.1017/9781108691017.007.
- [108] ITU, "Propagation data and prediction methods required for the design of terrestrial broadband millimetric radio access systems operating in a frequency range of about 20-50 GHz Annex 1," pp. 1–34, Apr. 2012.
- [109] M. H. T. Nguyen, E. Garcia-Palacios, T. Do-Duy, L. D. Nguyen, S. T. Mai, and T. Q. Duong, "Spectrum-Sharing UAV- Mission-Critical Communication: Learning-Aided Real-Time Optimisation," *IEEE Access*, vol. 9, pp. 11622–11632, Jan. 2021, doi: 10.1109/ACCESS.2021.3050522.
- [110] M. Alzenad, A. El-Keyi, and H. Yanikomeroglu, "3-D Placement of an Unmanned Aerial Vehicle Base Station for Maximum Coverage of Users with Different QoS Requirements," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 38–41, Feb. 2018, doi: 10.1109/LWC.2017.2752161.