

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



BÙI MINH PHỤNG

**TỐI ƯU LƯU TRỮ
VÀ TRUYỀN VIDEO CỘNG TÁC
TRONG MẠNG 5G SIÊU DÀY ĐẶC**

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

Đà Nẵng – 2021

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



NCS. BÙI MINH PHỤNG

**TỐI ƯU LƯU TRỮ VÀ TRUYỀN VIDEO CỘNG TÁC
TRONG MẠNG 5G SIÊU DÀY ĐẶC**

CHUYÊN NGÀNH: Khoa học Máy tính

MÃ SỐ: 9.48.01.01

LUẬN ÁN TIẾN SĨ

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. VÕ NGUYỄN SƠN

Đà Nẵng – 2021

LỜI CAM ĐOAN

Tác giả xin cam đoan đây là công trình nghiên cứu do chính tác giả thực hiện. 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.

Tác giả luận án

LỜI CẢM ƠN

Để hoàn thành được luận án tiến sĩ này, tôi đã nhận được rất nhiều sự hỗ trợ của Nhà trường, của người hướng dẫn khoa học, của đồng nghiệp và của gia đình.

Tôi xin gửi lời cảm ơn chân thành đến TS. Võ Nguyên Sơn, là người thầy và cũng là đồng nghiệp đã nhiệt tình hướng dẫn, định hướng và cùng tôi nghiên cứu trong suốt quá trình tôi thực hiện luận án tiến sĩ tại viện Nghiên cứu khoa học Cơ bản và Ứng dụng – Trường Đại học Duy Tân, TP. HCM.

Tôi xin cảm ơn đến PGS. TS. Nguyễn Gia Như, Trưởng khoa Sau đại học – Trường Đại học Duy Tân, là người đã động viên, khuyến khích tôi trong suốt quá trình thực hiện luận án.

Tôi xin cảm ơn các đồng nghiệp cùng làm nghiên cứu ở viện Nghiên cứu khoa học Cơ bản và Ứng dụng – Trường Đại học Duy Tân, TP. HCM, các đồng nghiệp ở Khoa CNTT – trường Đại học Văn Lang đã hỗ trợ tôi thực hiện luận án.

Tôi cũng xin cảm ơn Trường Đại học Duy Tân, Trường Đại học Văn Lang đã tạo điều kiện cho tôi có được môi trường tốt để hoàn thành luận án.

Cuối cùng, tôi xin cảm ơn gia đình đã hỗ trợ và đồng hành cùng tôi trong suốt thời gian tôi thực hiện luận án này.

Mặc dù đã cố gắng và nỗ lực trong quá trình nghiên cứu để luận án được hoàn chỉnh nhất nhưng luận án có thể vẫn còn nhiều thiếu sót. Tác giả rất mong nhận được sự góp ý quý giá để hoàn thiện luận án cũng như tiếp tục cho các nghiên cứu sau này.

Tác giả

Bùi Minh Phụng

MỤC LỤC

LỜI CAM ĐOAN	i
MỤC LỤC	iii
DANH MỤC CÁC CHỮ VIẾT TẮT VÀ KÝ HIỆU	vi
DANH MỤC CÁC BẢNG	viii
DANH MỤC CÁC HÌNH VÀ BIỂU ĐỒ	ix
MỞ ĐẦU	1
1. Lý do chọn đề tài.....	1
2. Mục tiêu, đối tượng, phạm vi và phương pháp nghiên cứu	4
2.1. Mục tiêu nghiên cứu.....	4
2.2. Đối tượng nghiên cứu	4
2.3. Phạm vi nghiên cứu	4
2.4. Phương pháp nghiên cứu.....	5
3. Nhiệm vụ nghiên cứu và kết quả đạt được.....	5
3.1. Nhiệm vụ nghiên cứu	5
3.2. Kết quả đạt được.....	6
4. Bố cục luận án.....	6
CHƯƠNG 1. TỔNG QUAN VỀ LƯU TRỮ VÀ TRUYỀN VIDEO TRONG 5G UDN	9
1.1. Giới thiệu về 5G UDN	9
1.1.1. Mô hình 5G UDN.....	9
1.1.2. Các thách thức của 5G UDN.....	12
1.2. Mô hình lưu trữ và truyền video trong 5G UDN	13
1.2.1. Lưu trữ và truyền video đơn tầng	14
1.2.2. Lưu trữ và truyền video cộng tác đa tầng.....	15
1.2.3. Phương pháp đánh giá hiệu năng mô hình.....	15
1.3. Hiện trạng các nghiên cứu về cơ chế lưu trữ và chia sẻ tài nguyên.....	16
1.3.1. Các nghiên cứu trong nước.....	16
1.3.2. Các nghiên cứu ngoài nước	18
1.4. Các thuật giải ứng dụng trong các bài toán tối ưu lưu trữ và truyền video....	23
1.4.1. Thuật giải vét cạn.....	24

1.4.2. Thuật giải đàn dơi.....	24
1.4.3. Thuật giải di truyền.....	29
1.5. Tổng kết chương 1	37
CHƯƠNG 2. LƯU TRỮ VÀ CHIA SẺ TÀI NGUYÊN TỐI ƯU THEO MỐI QUAN HỆ XÃ HỘI NGƯỜI DÙNG NÂNG CAO DUNG LƯỢNG TRUYỀN VIDEO TRONG 5G UDN.....	39
2.1. Giới thiệu cơ chế SCS	39
2.2. Mô hình truyền video trong 5G UDN với cơ chế SCS	43
2.2.1. Các ký hiệu và giả thiết.....	43
2.2.2. Mô hình hệ thống với cơ chế SCS.....	45
2.3. Tính toán các thông số hệ thống với cơ chế SCS.....	47
2.3.1. Mối quan hệ xã hội giữa các cặp TX-RX.....	47
2.3.2. Mô hình kênh truyền không dây.....	49
2.3.3. Dung lượng phân phối hệ thống	51
2.4. Bài toán tối ưu SCS và thuật giải vét cạn.....	55
2.5. Đánh giá hiệu suất cơ chế SCS	59
2.5.1. Thiết lập thông số hệ thống.....	59
2.5.2. Đánh giá hiệu suất cơ chế SCS.....	60
2.6. Kết luận chương 2	70
CHƯƠNG 3. LƯU TRỮ ĐA PHÂN GIẢI VÀ CHIA SẺ TÀI NGUYÊN TỐI ƯU THEO NHU CẦU NGƯỜI DÙNG NÂNG CAO CHẤT LƯỢNG TRUYỀN VIDEO TRONG 5G UDN.....	71
3.1. Giới thiệu cơ chế CRS.....	71
3.2. Mô hình truyền video trong 5G UDN với cơ chế CRS.....	74
3.3. Tính toán các thông số hệ thống với cơ chế CRS	79
3.3.1. Xác suất đạt dung lượng tại SU.....	79
3.3.2. Xác suất đạt dung lượng tại CU	82
3.3.3. Xác suất đạt dung lượng tại NU	83
3.3.4. Chất lượng trung bình của video nhận được	85
3.4. Bài toán tối ưu CRS và thuật giải di truyền	86
3.4.1. Bài toán tối ưu CRS	86
3.4.2. Thuật giải di truyền cho bài toán CRS.....	89
3.5. Đánh giá hiệu suất cơ chế CRS và thuật giải di truyền.....	92
3.5.1. Thiết lập thông số hệ thống và cấu hình máy tính.....	92

3.5.2. <i>Đánh giá hiệu quả của thuật giải di truyền</i>	94
3.5.3. <i>Đánh giá hiệu suất cơ chế CRS</i>	98
3.6. Kết luận chương 3	103
KẾT LUẬN	104
1. Kết quả đạt được	104
2. Định hướng nghiên cứu.....	105
CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ CỦA NGHIÊN CỨU SINH	106
1. Công trình đã công bố của luận án.....	106
2. Công trình đã công bố khác của nghiên cứu sinh	106
TÀI LIỆU THAM KHẢO	108

DANH MỤC CÁC CHỮ VIẾT TẮT VÀ KÝ HIỆU

Viết tắt	Tiếng Anh	Tiếng Việt
AWGN	Additive White Gaussian Noise	Nhiều trắng Gaussian
BA	Bat Algorithm	Thuật giải đàn dơi
BBA	Binary Bat Algorithm	Thuật giải đàn dơi nhị phân
CP	Content Provider	Nhà cung cấp nội dung
CRS	Multi-Resolution Caching and Resource Sharing Optimization	Lưu trữ đa phân giải và chia sẻ tài nguyên tối ưu
CS	Caching Strategy	Chiến lược lưu trữ
CU	Caching User	Người dùng có lưu trữ nội dung
D2D	Device-to-Device	Truyền thông từ thiết bị đến thiết bị
EA	Exhaustive Algorithm	Thuật giải vét cạn
FBS	Femto Base Station	Trạm cơ sở nhỏ femto
FRS	Full Rate Caching and Resource Sharing	Lưu trữ video với mức phân giải cao nhất và chia sẻ tài nguyên
GA	Genetic Algorithm	Thuật giải di truyền
MBS	Macro Base Station	Trạm cơ sở lớn
MU	Mobile User	Người dùng di động
NU	Normal User	Người dùng bình thường
PSNR	Peak Signal-to-Noise Ratio	Tỉ số tín hiệu cực đại trên nhiễu
QoS	Quality of Service	Chất lượng dịch vụ
RD	Rate-Distortion	Mối quan hệ giữa méo và tốc độ mã hóa

RS	Resource Sharing Strategy	Chiến lược chia sẻ tài nguyên
RX	Receiver	Thiết bị nhận
SBS	Smallcell Base Station	Trạm cơ sở nhỏ
SCS	Social-aware Caching and Resource Sharing	Lưu trữ và chia sẻ tài nguyên tối ưu theo mối quan hệ xã hội người dùng
SINR	Signal to Interference plus Noise Ratio	Tỉ số tín hiệu trên can nhiễu và nhiễu trắng
SNR	Signal to Noise-Ratio	Tỉ số tín hiệu trên nhiễu trắng
SU	Spectrum Owning User	Người dùng sở hữu kênh truyền sẵn sàng chia sẻ
TX	Transmitter	Thiết bị truyền
UDN	Ultra-dense Network	Mạng siêu dày đặc
VAS	Video Streaming Applications and Services	Ứng dụng và dịch vụ truyền video

DANH MỤC CÁC BẢNG

Bảng 2-1. Các ký hiệu sử dụng cho bài toán SCS	44
Bảng 2-2. Các tham số đầu vào cho bài toán SCS	59
Bảng 3-1. Các ký hiệu sử dụng cho bài toán CRS	75
Bảng 3-2. Các tham số đầu vào cho bài toán CRS	93
Bảng 3-3. Cấu hình máy tính	94
Bảng 3-4. So sánh độ chính xác giữa GA, BBA và EA.....	95
Bảng 3-5. Thực thi GA theo số N_{PG} với kết quả xấu nhất	96

DANH MỤC CÁC HÌNH VÀ BIỂU ĐỒ

Hình 1-1. Mô tả kết nối giữa các thiết bị, các công nghệ và ứng dụng trong 5G UDN [2]	10
Hình 1-2. Cơ chế lưu trữ và chia sẻ tài nguyên cho các dịch vụ và ứng dụng video trong 5G UDN.....	14
Hình 1-3. Lưu đồ thuật giải dànơi nhị phân.....	28
Hình 1-4. Lai ghép đơn điểm	31
Hình 1-5. Lai ghép đa điểm (m=5)	32
Hình 1-6. Đột biến bù nhị phân.....	33
Hình 1-7. Lưu đồ thuật giải di truyền	34
Hình 2-1. Mô hình hệ thống của SCS	45
Hình 2-2. Lưu đồ hoạt động cơ chế SCS	46
Hình 2-3. Dung lượng hệ thống theo khả năng lưu trữ của FBS	61
Hình 2-4. Dung lượng hệ thống theo hệ số δ	62
Hình 2-5. Dung lượng hệ thống theo độ nổi tiếng của video.....	63
Hình 2-6. Dung lượng hệ thống theo số cặp D2D	64
Hình 2-7. Dung lượng hệ thống theo số lượng FBS	65
Hình 2-8. Dung lượng hệ thống theo Gamma.....	66
Hình 2-9. Dung lượng hệ thống theo khoảng cách giữa MBS và các MU	67
Hình 2-10. Dung lượng hệ thống theo khoảng cách giữa MBS và các MU với C^* thấp hơn.....	69
Hình 3-1. Mô hình hệ thống của CRS.....	75
Hình 3-2. Lưu đồ hoạt động cơ chế CRS	77
Hình 3-3. Đánh giá độ ổn định của GA và BBA	95
Hình 3-4. Mức độ hội tụ của thuật giải di truyền.....	97
Hình 3-5. Chất lượng hệ thống (HT) của CRS, OCC, ORS, NCS theo số lượng FBS	99
Hình 3-6. Chất lượng HT của CRS, OCC, ORS, NCS theo số lượng SU	100

Hình 3-7. Chất lượng HT của CRS, OCC, ORS, NCS theo α	101
Hình 3-8. Chất lượng HT của CRS với các tham số HT khác nhau theo α	102
Hình 3-9. So sánh kết quả của CRS và FRS	103

MỞ ĐẦU

1. Lý do chọn đề tài

Thế giới đang bước vào kỷ nguyên của công nghiệp 4.0, song song đó, theo dự đoán sẽ có khoảng 50 tỉ thiết bị được kết nối để truyền thông với nhau và truy cập các dịch vụ/tiện ích đa phương tiện vào năm 2021 [1]. Trong bối cảnh này, mạng thế hệ thứ 5 (5G – Fifth Generation) sẽ đối diện với thách thức của một lượng lớn người dùng di động (MU – Mobile User) yêu cầu các dịch vụ và ứng dụng có dung lượng dữ liệu và tốc độ cao. Điều này sẽ khiến cho mạng 5G trở nên suy yếu bởi vấn đề tắc nghẽn do xung đột lưu lượng xảy ra tại các tuyến trục (backhaul link) của các trạm cơ sở lớn (MBS – Macro Base Station) cũng như của các trạm cơ sở nhỏ (SBS – Small cell Base Station¹). Trong khi việc đầu tư phát triển các tuyến trục tốc độ cao tốn rất nhiều chi phí và thậm chí phải thay đổi kiến trúc mạng thì những giải pháp và kỹ thuật mới có thể thay thế một cách hiệu quả hơn.

Mạng siêu dày đặc (UDN – Ultra-Dense Network) được xem như là một kiến trúc hứa hẹn có khả năng đáp ứng được các yêu cầu của mạng 5G như tăng dung lượng hệ thống gấp 1000 lần và độ trễ truy xuất dữ liệu nhỏ hơn 1ms [2]. Tuy nhiên, việc phát triển UDN còn phải yêu cầu nhiều hơn nữa các công nghệ, kỹ thuật và các thiết kế tối ưu đột phá nhằm cung cấp cho lượng lớn MU những dịch vụ và ứng dụng có dung lượng dữ liệu và tốc độ cao, ví dụ như dịch vụ và ứng dụng truyền video (VAS – Video Applications and Services), với chất lượng dịch vụ (QoS – Quality of Service) và hiệu quả sử dụng tài nguyên tốt nhất. Để đạt được điều này, nhiều công nghệ, kỹ thuật và thiết kế tối ưu cho UDN đã được nghiên cứu tập trung vào việc làm thế nào để tận dụng các nguồn tài nguyên về không gian, thời gian, mã, phổ tần, băng thông, năng lượng và dung lượng lưu trữ, cũng như làm thế nào để đưa các dịch vụ tiên tiến đến gần với các MU hơn.

¹ SBS là từ chung cho các trạm cơ sở nhỏ, trong đó bao gồm cả Femtocell, Picocell, Metrocell và Microcell theo phạm vi phủ sóng từ nhỏ tới lớn.

Một số ví dụ điển hình về công nghệ, kỹ thuật và thiết kế kết hợp với UDN có thể kể đến như: công nghệ truyền thông sóng milimet (mmWave) và MIMO cỡ rất lớn (Massive Multiple Input Multiple Output) được nghiên cứu để cải tiến hiệu quả phổ và mở rộng băng thông truyền [3-5]; công nghệ mmWave có quan tâm tới giao thoa và điều khiển công suất trong mạng truyền thông từ thiết bị đến thiết bị (D2D – Device-to-Device) để gia tăng hiệu quả sử dụng năng lượng và phổ tần [6-8]; kỹ thuật truyền thông D2D [9-12]; và đặc biệt là các cơ chế lưu trữ [13-48] nhằm đưa các dịch vụ đến gần người dùng hơn có thể được áp dụng vào trong UDN để cải tiến dung lượng hệ thống, giảm xung đột tại các tuyến trục của MBS và SBS; đã đạt được một số tiêu chí của QoS và hiệu quả sử dụng tài nguyên cao.

Tuy nhiên, các giải pháp trên đều chưa thể áp dụng hiệu quả vào VAS được ước tính là sẽ chiếm đến 78% tổng dung lượng dữ liệu vào năm 2021 [49]. Nguyên nhân đầu tiên đó là các tiêu chí đánh giá chất lượng dịch vụ truyền video rất nhạy cảm với các yếu tố của môi trường truyền không dây; hành vi và mối quan hệ của MU; tài nguyên của mạng và của MU; các đặc tính của video và các kỹ thuật mã hóa – đóng gói video. Nguyên nhân thứ hai đó là tài nguyên mạng không được tận dụng một cách triệt để nhằm 1) lưu trữ các nội dung video (sử dụng dung lượng lưu trữ) tại các thiết bị di động, SBS, và MBS gần với MU hơn và 2) cộng tác phân phối nội dung video (bằng cách tận dụng thêm các nguồn tài nguyên khác như băng thông, năng lượng và phổ) từ điểm lưu trữ đến MU một cách hiệu quả.

Gần đây, kỹ thuật lưu trữ (caching) đã thu hút sự chú ý của các nhà nghiên cứu trong cả lĩnh vực học thuật và công nghiệp để mang lại lợi ích cho các nhà cung cấp dịch vụ Internet (ISP – Internet Service Provider) và nhà cung cấp nội dung (CP – Content Provider) cũng như đáp ứng nhu cầu cao của người dùng cuối. Trong lĩnh vực này, các nhà nghiên cứu Việt Nam đã có cơ hội hợp tác với các chuyên gia hàng đầu thế giới nhằm phát triển các dự án và công bố các công trình nghiên cứu ở các tạp chí uy tín [14, 19, 21-23, 29-31, 35, 38, 44, 50]. Tuy vậy, các nhóm nghiên cứu này vẫn chưa triển khai một cách tổng thể mô hình, phân tích và thiết kế tối ưu lưu trữ và truyền video hợp tác trong 5G UDN. Các công trình trong nước chủ yếu

tập trung nghiên cứu tối ưu lưu trữ hoặc tối ưu chi phí cho nhà cung cấp dịch vụ hoặc nhà phát triển nội dung thực hiện trên đơn tầng ở trạm cơ sở chính (MBS) hoặc tại trạm cơ sở nhỏ (SBS), nhưng chưa có các nghiên cứu định hướng đến lợi ích người dùng (MU) và đa tầng trong mạng 5G UDN. Và như vậy, vẫn còn nhiều vấn đề thách thức chưa được giải quyết nhằm nâng cao năng lực cho 5G UDN trong việc cung cấp cho MU các dịch vụ tiên tiến có QoS và hiệu suất sử dụng tài nguyên (dung lượng, băng thông, năng lượng, phổ tần) cao.

Ở các nước phát triển, các công trình nghiên cứu trong lĩnh vực này nhiều hơn về số lượng và tốt hơn về chất lượng, cụ thể như: lưu trữ tại các thiết bị (D2D caching) [15, 20, 33, 42, 45], lưu trữ tại trạm cơ sở nhỏ femto (femtocaching) [26, 27, 34], lưu trữ tại các trạm cơ sở nhỏ (small-cell caching) [17, 32, 37, 41, 43, 46], lưu trữ tại trạm cơ sở lớn (MBS caching) [39, 40], và lưu trữ đa tầng (multi-tier caching) [16, 18, 24, 25, 28, 36, 47, 48].

Có thể nhận thấy rằng, mặc dù các công trình nghiên cứu được thực hiện tại các quốc gia phát triển nhiều hơn về số lượng, đã thực hiện trên đa tầng (MBS, FBS, MU), và có những đóng góp quan trọng hơn so với các công trình được thực hiện bởi các nhà nghiên cứu trong nước. Tuy nhiên, các công trình nghiên cứu ngoài nước vẫn chưa nghiên cứu đồng thời tối ưu lưu trữ đa tầng kết hợp chia sẻ tài nguyên có quan tâm đến mối quan hệ xã hội của người dùng để nâng cao hiệu quả của hệ thống. Và như vậy, vẫn còn nhiều vấn đề chưa được giải quyết, cần phải có những kỹ thuật, mô hình, phân tích, thiết kế tối ưu và các tiêu chuẩn đột phá nhằm đáp ứng các yêu cầu cao và phức tạp của lưu trữ và truyền/phân phối video trong 5G UDN.

Từ những vấn đề còn tồn đọng trong các nghiên cứu trong và ngoài nước như trên, trong luận án này, NCS sẽ đề xuất cơ chế lưu trữ đa tầng kết hợp với chia sẻ tài nguyên tối ưu theo định hướng người dùng nhằm nâng cao hiệu quả truyền video trong 5G UDN.

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

2.1. Mục tiêu nghiên cứu

Mục tiêu nghiên cứu của luận án là đề xuất các cơ chế lưu trữ và truyền video hợp tác đa tầng trong 5G UDN nhằm cung cấp cho người dùng di động các dịch vụ tiên tiến có chất lượng dịch vụ và hiệu suất sử dụng tài nguyên cao.

2.2. Đối tượng nghiên cứu

- 5G UDN: tập trung vào nghiên cứu các mô hình, các đặc điểm của 5G UDN có nhiều tầng thiết bị thu phát gồm MBS, SBS (như microcell, picocell, femtocell) và truyền thông D2D.
- Video: các loại video chuẩn được dùng phổ biến, mô hình mối quan hệ giữa méo và tốc độ mã hóa của video, các cơ chế lưu trữ và truyền video trong 5G UDN.
- Mô hình: mối quan hệ xã hội, QoS và tài nguyên trong 5G UDN.
- Thuật giải: các thuật giải nhằm giải bài toán tối ưu quá trình lưu trữ, chia sẻ tài nguyên và truyền video từ MBS và SBS đến các MU và giữa các MU với nhau.

2.3. Phạm vi nghiên cứu

Trong luận án này, NCS giới hạn các phạm vi nghiên cứu liên quan đến kỹ thuật, người dùng, mạng và dịch vụ cũng như toán và các công cụ hỗ trợ, cụ thể như sau:

- Về kỹ thuật: luận án tập trung vào kỹ thuật lưu trữ và chia sẻ tài nguyên trong 5G UDN.
- Về người dùng, mạng và dịch vụ: luận án tập trung vào dịch vụ và ứng dụng truyền video (VAS) trong 5G UDN, các tiêu chí của QoS, hiệu suất sử dụng tài nguyên, hành vi và mối quan hệ xã hội của người dùng dựa trên mô hình Buffet Ấn Độ (IBM – Indian Buffet Model).

- Về toán và các công cụ hỗ trợ: luận án tập trung vào các thuật giải tìm kiếm, một số công cụ phân tích, công cụ mã hóa video và mô hình đường cong đặc tính méo – tốc độ (Rate – Distortion) của video.

2.4. Phương pháp nghiên cứu

Để đạt kết quả phù hợp với các mục tiêu nghiên cứu, dựa trên các đối tượng cần nghiên cứu, 02 phương pháp nghiên cứu được áp dụng gồm 1) Phương pháp phân tích và tổng hợp và 2) Phương pháp nghiên cứu định lượng, cụ thể như sau

- Phương pháp phân tích và tổng hợp: Phân tích và đánh giá các mô hình và kết quả nghiên cứu được công bố liên quan, từ đó, xác định được các vấn đề còn tồn đọng, những câu hỏi và ý tưởng, đặt ra các giả thuyết nghiên cứu cũng như khẳng định sự cần thiết phải có mô hình và giải pháp mới tốt hơn. Sau đó, các kết quả phân tích và đánh giá sẽ được tổng hợp, liên kết, kết hợp và tổ chức lại một cách có hệ thống nhằm đề xuất các mô hình và giải pháp tốt hơn dựa trên giả thuyết nghiên cứu được đặt ra.
- Phương pháp nghiên cứu định lượng: Trong phương pháp định lượng, các mô hình đề xuất và các yếu tố liên quan được lượng hóa bởi các biểu thức tính toán. Các biểu thức tính toán mô tả bản chất của mô hình hệ thống được kiểm định tính đúng đắn thông qua việc thực hiện mô phỏng và quan sát đáp ứng của hệ thống. Tính ưu việt của giải pháp đề xuất được kiểm chứng thông qua việc đối sánh với các giải pháp của các nghiên cứu khác liên quan.

3. Nhiệm vụ nghiên cứu và kết quả đạt được

3.1. Nhiệm vụ nghiên cứu

- Nhiệm vụ nghiên cứu 1: Đề xuất cơ chế lưu trữ và chia sẻ tài nguyên tối ưu theo mối quan hệ xã hội người dùng nâng cao dung lượng truyền video trong 5G UDN, mô hình toán các thông số của mô hình đề xuất, xây dựng và giải

bài toán tối ưu bằng các giải thuật tìm kiếm, mô phỏng tính toán và đánh giá hiệu quả của mô hình.

- Nhiệm vụ nghiên cứu 2: Đề xuất cơ chế lưu trữ đa phân giải và chia sẻ tài nguyên tối ưu theo nhu cầu người dùng nâng cao chất lượng truyền video trong 5G UDN, mô hình toán các thông số của mô hình đề xuất, xây dựng và giải bài toán tối ưu bằng các giải thuật tìm kiếm, mô phỏng tính toán và đánh giá hiệu quả của mô hình.

3.2. Kết quả đạt được

Trong quá trình nghiên cứu và hoàn thành luận án, NCS đã công bố 2 bài báo quốc tế và 1 bài hội nghị quốc tế. Kết quả nghiên cứu của luận án góp phần vào việc tối ưu lưu trữ và phân phối tài nguyên là video nhằm cung cấp cho người dùng dịch vụ và chất lượng video tốt hơn.

- Kết quả 1: Đề xuất cơ chế lưu trữ và truyền video cộng tác đa tầng có quan tâm đến hành vi và mối quan hệ xã hội của người dùng (SCS) thông qua việc tận dụng tài nguyên lưu trữ và tài nguyên phổ tần sẵn có của các phần tử trong 5G UDN.
- Kết quả 2: Đề xuất cơ chế lưu trữ đa phân giải và truyền video cộng tác có quan tâm đến nhu cầu người dùng (CRS) thông qua việc tận dụng hiệu quả tài nguyên lưu trữ và tài nguyên phổ tần sẵn có của các phần tử trong 5G UDN.

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:

Mở đầu

Trình bày về lý do chọn đề tài, mục tiêu nghiên cứu, các đối tượng nghiên cứu, phạm vi nghiên cứu và phương pháp nghiên cứu.

Chương 1: Tổng quan về lưu trữ và truyền video trong 5G UDN

Chương 1 tập trung vào tìm hiểu các cơ chế lưu trữ và truyền video cộng tác trong 5G UDN, đồng thời, đánh giá các mô hình để hiểu rõ được ưu và nhược điểm của các mô hình, là cơ sở để đề xuất các cơ chế mới cho truyền video trong 5G UDN.

Chương 2: Lưu trữ và chia sẻ tài nguyên tối ưu theo mối quan hệ xã hội người dùng nâng cao dung lượng truyền video trong 5G UDN

Chương 2 đề xuất cơ chế lưu trữ và truyền video cộng tác đa tầng có quan tâm đến hành vi và mối quan hệ xã hội của MU. Trong cơ chế này, các thông số được xem xét bao gồm: tài nguyên về dung lượng lưu trữ để lưu các video, tài nguyên phổ tần sẵn có để chia sẻ cho truyền thông D2D, và yếu tố hành vi và mối quan hệ xã hội của MU trong 5G UDN. Các thông số của hệ thống được tính toán để xây dựng bài toán tối ưu nhằm cực đại dung lượng phân phối video của hệ thống đến các MU. Bài toán tối ưu được giải bằng thuật giải vét cạn, các mô phỏng tính toán được thực thi trên Matlab để đánh giá cơ chế được đề xuất và so sánh với các cơ chế khác.

Chương 3: Lưu trữ đa phân giải và chia sẻ tài nguyên tối ưu theo nhu cầu người dùng nâng cao chất lượng truyền video trong 5G UDN

Chương 3 tập trung vào việc cải tiến cơ chế lưu trữ và truyền video đã đề xuất ở Chương 2. Cụ thể là thay đổi thông số đánh giá QoS từ dung lượng phân phối thành chất lượng video thu được; tận dụng tài nguyên lưu trữ hiệu quả hơn bằng cách chọn lựa lưu trữ video theo các mức phân giải khác nhau và tận dụng tài nguyên phổ tần hiệu quả hơn bằng chọn lựa theo bộ 3 thiết bị (SU – Spectrum owning User, CU – Caching User và NU – Normal User) thay

vì chỉ bộ 2 (SU và một cặp D2D đã định sẵn); phục vụ MU linh động và hiệu quả hơn với nhiều sự lựa chọn video với các mức phân giải khác nhau (khác nhau về chất lượng) phù hợp với năng lực xử lý của thiết bị người dùng (nhu cầu người dùng). Các thông số của hệ thống được tính toán để xây dựng bài toán tối ưu nhằm cực đại chất lượng video thu được tại các MU. Bài toán tối ưu được giải bằng thuật giải di truyền, các mô phỏng tính toán được thực thi trên Matlab để đánh giá cơ chế được đề xuất và so sánh với các cơ chế khác.

Kết luận

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

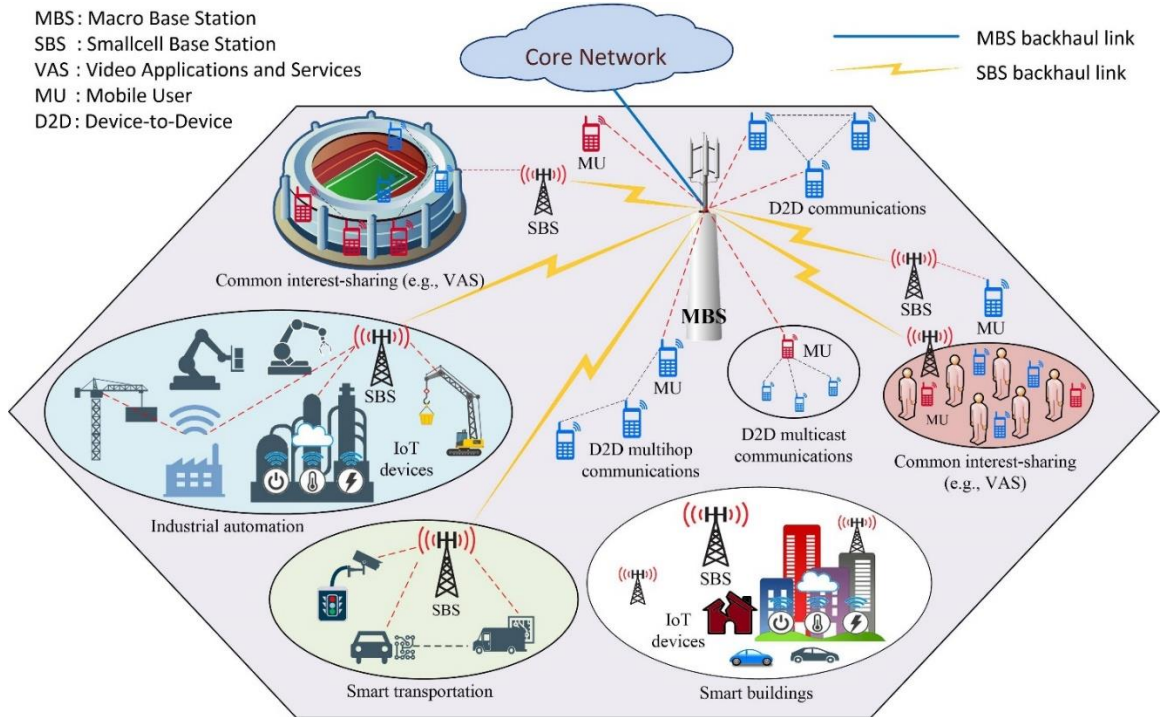
CHƯƠNG 1. TỔNG QUAN VỀ LƯU TRỮ VÀ TRUYỀN VIDEO TRONG 5G UDN

Tóm tắt: Chương 1 sẽ giới thiệu về các kỹ thuật lưu trữ và truyền video trong mạng 5G siêu dày đặc (UDN – Ultra-dense Network). Trong đó, các ưu điểm và những vấn đề cần được nghiên cứu của 5G UDN được trình bày, đặc biệt, kỹ thuật lưu trữ đa tầng và truyền video cộng tác tận dụng hiệu quả tài nguyên (dung lượng lưu trữ và phổ tần) được nhấn mạnh như là một giải pháp hiệu quả. Ngoài ra, các khái niệm, kết quả nghiên cứu liên quan, những công cụ và thuật giải cần thiết cũng được đề cập, phân tích và đánh giá một cách chi tiết. Từ đó, làm tiền đề cho những cơ chế mới được đề xuất trong các chương tiếp theo cũng như làm nổi bật ý nghĩa và đóng góp của luận án.

1.1. Giới thiệu về 5G UDN

1.1.1. Mô hình 5G UDN

Mạng thông tin di động là một hình thức mạng truyền thông hiện đại cho phép con người có thể liên lạc, kết nối và chia sẻ thông tin với nhau ở bất kỳ đâu. Với sự phát triển nhanh chóng của khoa học kỹ thuật và công nghệ, cuộc cách mạng công nghiệp 4.0 đã và đang diễn ra ở nhiều nơi, nhiều nước trên thế giới, cùng hàng tỉ thiết bị di động được kết nối không dây. Theo ước tính, sẽ có khoảng 50 tỷ thiết bị được kết nối vào mạng không dây để truyền thông tin và truy cập các dịch vụ tiên tiến [51]. Trong bối cảnh này, lượng lớn người dùng yêu cầu các dịch vụ tiên tiến sẽ khiến cho mạng không dây, đặc biệt là mạng di động thế hệ thứ 5 (5G – Fifth Generation), trở nên suy yếu bởi vấn đề tắc nghẽn lưu lượng tại các tuyến trục (backhaul link) của các trạm cơ sở lớn (MBS – Macro Base Station) và các trạm cơ sở nhỏ (SBS – Small cell Base Station). Gần đây, mạng 5G UDN [2] được xem là mô hình khả thi nhờ vào khả năng cung cấp lượng lớn kết nối các dịch vụ dữ liệu lớn và tốc độ cao cũng như trì hoãn thấp so với các thế hệ mạng di động trước đó.



Hình 1-1. Mô tả kết nối giữa các thiết bị, các công nghệ và ứng dụng trong 5G UDN [2]

Mô hình của 5G UDN gồm nhiều tầng, cung cấp cho người dùng di động (MU – Mobile User) nhiều dịch vụ và ứng dụng khác nhau, được mô tả trong Hình 1-1. Trong mô hình này, tầng trên sử dụng dải phổ tần cao hơn, tín hiệu mạnh hơn với tốc độ truyền cao hơn, trong khi đó, không gian dịch vụ và diện tích phủ sóng sẽ được mở rộng hơn nhờ vào các tầng trung gian và tầng thấp [2, 52]. 5G UDN sẽ cho phép kết hợp linh hoạt các kỹ thuật truyền thông dựa vào việc tận dụng và chia sẻ tài nguyên giữa các phần tử/nhóm phần tử trong mạng nhằm đảm bảo kết nối, đáp ứng nhu cầu ngày càng cao của người dùng. Cụ thể, những tính năng và yêu cầu chính của 5G UDN [2] bao gồm:

- Tốc độ dữ liệu và độ trễ: cho phép truyền dữ liệu với tốc độ cao, lên đến vài Gbp/s và độ trễ chỉ khoảng 1 mili giây.
- Giao tiếp đa thiết bị/máy: cho phép kết nối nhiều thiết bị cùng lúc và có thể chia sẻ những dữ liệu dùng chung từ các thiết bị như là điện thoại thông minh, thiết bị gia dụng, tivi, xe ô tô, thiết bị cảm ứng....

- Phổ tần: sự gia tăng nhu cầu kết nối của các thiết bị đòi hỏi lưu lượng, tài nguyên phổ tần cho 5G UDN cũng tăng lên. Các băng tần cao thuộc bước sóng xăng-ti-mét (cmWave) và milimét (mmWave) là các băng tần tiềm năng vì chúng có khả năng cung cấp kênh băng thông lớn. Do đó, cung cấp được tốc độ dữ liệu cao, cho phép truyền tải với băng thông rộng hơn băng thông 20MHz của mạng 4G.
- Kết hợp nhiều công nghệ: 5G UDN không phải là sự thay đổi công nghệ mới, mà nó hỗ trợ và phát triển thêm nhiều công nghệ. 5G UDN kết hợp nhiều hệ thống như GSM, HSPA (phương thức kết nối gói tốc độ cao), LTE và các hệ thống hỗ trợ truy cập với hiệu suất cao hơn.
- 5G UDN hoạt động gồm nhiều tầng như được mô tả trong Hình 1-1, gồm MBS, SBS, và các kết nối từ thiết bị đến thiết bị (D2D – Device-to-Device). Việc chia thành nhiều tầng sẽ đảm bảo hiệu suất hệ thống (năng lượng, phổ tần và không gian) tốt hơn.
- Hỗ trợ tối đa khai thác dữ liệu: trong quá trình khai thác, mỗi thiết bị truy cập có các yêu cầu khác nhau về tốc độ, độ trễ, nội dung truy cập. Trong 5G UDN, vấn đề này sẽ được giải quyết linh hoạt và thông minh bởi vì các tầng có thể hỗ trợ nhau để chuyển tải đến người dùng nội dung tốt nhất. Dữ liệu có thể được truy xuất từ MBS, từ SBS, hoặc từ người dùng khác qua truyền thông D2D.
- Hiệu suất dụng tài nguyên cao: một trong những thách thức chính của 5G UDN là nâng cao hiệu quả sử dụng năng lượng của các thiết bị kết nối mạng, làm sao để nâng cao thời gian sử dụng của pin, cải thiện hiệu quả năng lượng với năng lượng tiêu thụ thấp kết hợp hấp thu năng lượng (energy harvesting) từ môi trường xung quanh (sóng/năng lượng vô tuyến) ở khoảng cách gần, cụ thể là từ các MBS [53].

1.1.2. Các thách thức của 5G UDN

Tuy nhiên, 5G UDN vẫn còn những thách thức về vật lý cũng như quá trình sử dụng và khai thác dữ liệu của MU như:

- Mật độ kết nối dày đặc từ những phạm vi truy cập và lưu lượng truy cập khác nhau tại các vị trí khác nhau. Điều đó có thể gây tắc nghẽn cục bộ, làm mất cân bằng hiệu suất và quyền hạn truy cập của các thiết bị.
- Những giới hạn truy cập riêng hoặc chung ở các tầng khác nhau sẽ tạo ra nhiều mức can nhiễu khác nhau. Ví dụ: các phiên truyền thông D2D khác nhau có thể ảnh hưởng can nhiễu cho nhau và thậm chí làm gián đoạn quá trình truy cập.
- Vấn đề ưu tiên truy cập các kênh ở các tần số khác nhau và ưu tiên trong các chiến lược phân bổ tài nguyên.

Để khắc phục các thách thức đó, 5G UDN phải thiết kế và xây dựng đảm bảo các yếu tố sau:

- Tối ưu đa tầng: đồng bộ và tối ưu theo mô hình đa tầng để quản lý và cộng tác nhằm truyền tải dữ liệu đến người dùng tốt nhất với phân bổ tài nguyên hợp lý nhất. Quản lý và cộng tác hiệu quả liên tầng (inter-tier) và trong tầng (intra-tier) sẽ đảm bảo sự linh hoạt trong quá trình truyền dữ liệu, làm giảm thiểu sự can thiệp nhau trong vận hành. Một thiết bị có thể đồng thời tải dữ liệu ở các tầng khác nhau hoặc từ các thiết bị khác gần đó. Tuy nhiên, điều này cần phải tích hợp chặt chẽ giữa các nút mạng thành một hệ thống sử dụng và chia sẻ tin cậy và nhanh chóng.
- Thiết kế các phương pháp hiệu quả để hỗ trợ đồng thời nhiều thiết bị kết nối: 5G UDN đa tầng sẽ tăng thông lượng hệ thống và giảm tỷ lệ hao hụt, tăng hiệu quả tài nguyên cho những người sử dụng gần nhau. Vì vậy, cần tăng cường sử dụng các thiết bị mới sử dụng công nghệ hiện đại, đảm bảo tiêu thụ ít năng lượng và cho phép kết nối một lúc nhiều thiết bị với nhau.

Các thiết bị mới phải tối ưu trong mỗi điểm mạng để đảm bảo quá trình khai thác và sử dụng không gián đoạn.

- Quản lý can nhiễu trong 5G UDN đa tầng cho các người dùng khác nhau được tái sử dụng tài nguyên với sự sắp xếp can nhiễu hợp lý, đảm bảo được sự linh hoạt, không trùng lặp dữ liệu và quá tải tại một điểm/phần tử mạng nào đó. Các thiết kế phải vượt qua những hạn chế này để trở thành một mạng tối ưu nhất có thể.

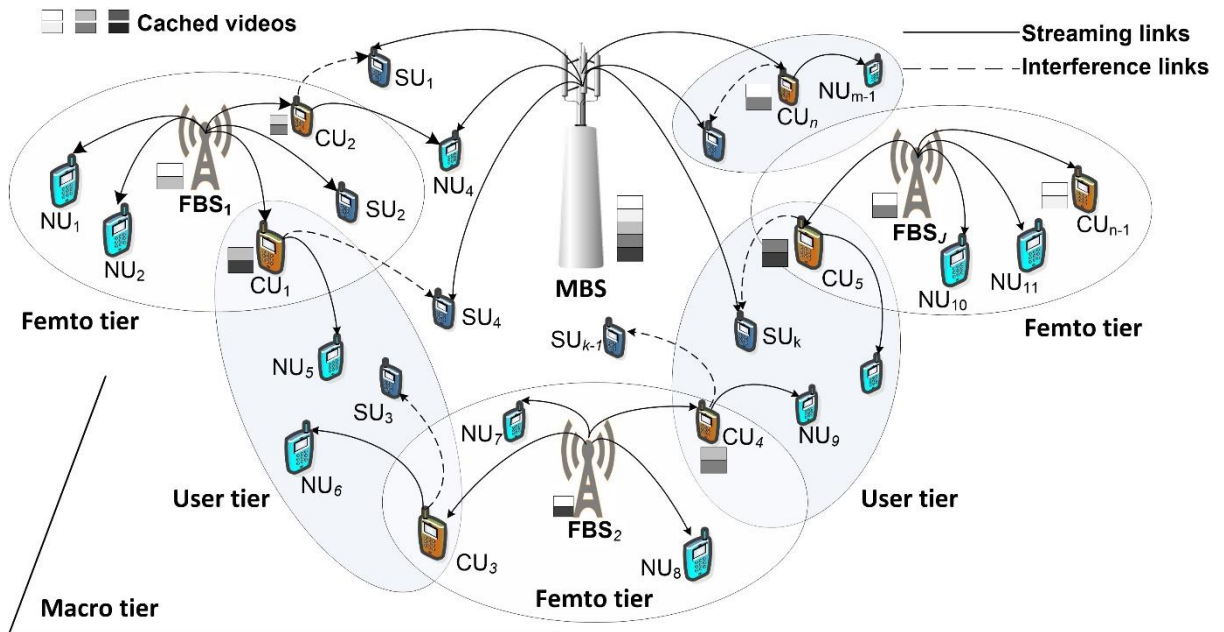
Như vậy, những nghiên cứu nhằm đưa ra các giải pháp khắc phục các vấn đề thách thức nêu trên trong 5G UDN là cấp thiết để sử dụng tối ưu nguồn tài nguyên, nâng cao chất lượng mạng di động và cải thiện sự hài lòng của người dùng.

1.2. Mô hình lưu trữ và truyền video trong 5G UDN

Một trong những giải pháp có thể nhìn thấy để giải quyết vấn đề tắc nghẽn do mật độ kết nối dày đặc đó là nâng cấp các tuyến trục (backhaul link) như được mô tả trong Hình 1-1. Tuy nhiên, đây là giải pháp tốn kém và cần phải thay đổi nhiều trong kiến trúc hệ thống. Thay vào đó, kỹ thuật lưu trữ để mang các dịch vụ/ứng dụng/nội dung, đặc biệt là các dịch vụ và ứng dụng có dung lượng lớn và tốc độ dữ liệu cao như truyền video (VAS – Video Applications and Service), đến gần với MU hơn cũng giải quyết được bài toán tắc nghẽn nhưng lại không làm thay đổi kiến trúc hệ thống. Kỹ thuật lưu trữ trong 5G UDN sẽ cho phép MU truy suất nội dung video ngay trong nội bộ mạng (khu vực phát của MBS hoặc của SBS) mà không cần phải truy cập đến máy chủ lưu trữ trong mạng lõi (core network). Thậm chí, MU không cần phải truy xuất nội dung video từ MBS và SBS, mà truy xuất từ những MU khác đã lưu trữ nội dung thông qua truyền thông D2D trong khoảng cách ngắn. Dĩ nhiên, trong kỹ thuật lưu trữ sẽ có những yêu cầu hỗ trợ của các kỹ thuật khác trong truyền thông, ví dụ như kỹ thuật chia sẻ tài nguyên phổ tần, để nâng cao hiệu quả lưu trữ.

Kỹ thuật lưu trữ và chia sẻ tài nguyên phổ tần đã và đang thu hút sự chú ý của các nhà nghiên cứu trong và ngoài nước để mang lại lợi ích cho các nhà cung cấp

dịch vụ Internet (ISP – Internet Service Provider) và nhà cung cấp nội dung (CP – Content Provider) cũng như đáp ứng nhu cầu cao của MU trong việc truy xuất các dịch vụ tiên tiến tốc độ cao như truyền video. Các mô hình lưu trữ kết hợp chia sẻ tài nguyên phổ biến cho các ứng dụng và dịch vụ truyền video trong 5G UDN được trình bày như sau.



Hình 1-2. Cơ chế lưu trữ và chia sẻ tài nguyên cho các dịch vụ và ứng dụng video trong 5G UDN

1.2.1. Lưu trữ và truyền video đơn tầng

Lưu trữ và truyền video đơn tầng là kỹ thuật cho phép hệ thống lưu trữ các video tại tầng MBS (tại macro tier) hoặc tại tầng SBS (femto tier) hoặc tại tầng người dùng di động (user tier) có khả năng lưu trữ (CU – Caching User) như được mô tả trong Hình 1-2. Lưu trữ tại MBS là một trong những kỹ thuật lưu trữ đơn giản nhưng hiệu quả để thay thế cho việc đầu tư tuyến trực tốc độ cao nhưng rất tốn kém. Ngoài ra, lưu trữ tại MBS cũng giúp cho việc phục vụ số lượng lớn các MU với độ trễ truy xuất thấp, tính liên tục cao và chất lượng hiển thị cao hơn so với truy xuất nội dung từ mạng lõi. Lưu trữ tại các SBS cho phép linh động mang nội dung video đến gần một nhóm MU có cùng mối quan tâm đến các nội dung video để phục vụ hiệu quả hơn. Tuy nhiên, việc lưu trữ tại SBS phải cần đến sự hỗ trợ của các cơ chế quản lý và

phân bổ tài nguyên phổ tần hợp lý để loại bỏ/hạn chế can nhiễu liên tầng (giữa MBS và SBS) và can nhiễu đồng tầng (giữa các SBS với nhau) để truyền video đến MU. Lưu trữ tại CU để truyền video trong khoảng cách gần qua truyền thông D2D được xem là kỹ thuật lưu trữ tại vùng biên của 5G UDN hiệu quả nhằm làm giảm lưu lượng chuyển tiếp tại các tuyến trục của cả MBS và SCS. Kỹ thuật lưu trữ này thông thường đòi hỏi sự chia sẻ tài nguyên phổ tần từ các thiết bị khác, hay còn gọi là thiết bị sở hữu tài nguyên phổ tần (SU – Spectrum Owning User) sẵn sàng chia sẻ. SU sẽ chia sẻ tài nguyên phổ tần của nó cho CU để CU dùng lại phổ tần này cho việc truyền video qua truyền thông D2D. Việc quản lý và chia sẻ tài nguyên phổ tần của SU đóng vai trò rất quan trọng trong kỹ thuật lưu trữ tại CU vì việc CU dùng lại phổ tần để truyền video sẽ gây can nhiễu lên SU.

1.2.2. Lưu trữ và truyền video cộng tác đa tầng

Lưu trữ và truyền video cộng tác đa tầng là kỹ thuật lưu trữ video được thực hiện kết hợp tại nhiều hơn 1 tầng trong 5G UDN. Trong lưu trữ đa tầng, mặc dù kỹ thuật sẽ phức tạp hơn so với đơn tầng, nhưng các MU được chọn lựa truy xuất video một cách linh động từ tầng nào được xem là tốt nhất. Các nghiên cứu liên quan đến lưu trữ đa tầng chủ yếu triển khai trên 2 tầng (sẽ được trình bày chi tiết trong Mục 1.3). Trong luận án này, lưu trữ đa tầng được hiểu là 3 tầng có kết hợp cơ chế chia sẻ tài nguyên phổ tần kênh truyền xuống của các SU cho truyền thông video D2D. Các tầng sẽ phối hợp với nhau trong quá trình lưu trữ và chia sẻ tài nguyên để phục vụ MU một cách tốt nhất với hiệu quả sử dụng tài nguyên cao nhất.

1.2.3. Phương pháp đánh giá hiệu năng mô hình

Để đánh giá được hiệu năng lưu trữ và truyền video trong 5G UDN, các đặc tính của kênh truyền, các thông số của video (độ nổi tiếng, kích thước, mối quan hệ giữa méo và tốc độ mã hóa (Rate-Distortion)), thông tin người dùng (mối quan hệ xã hội và nhu cầu), và tài nguyên hệ thống (dung lượng lưu trữ và phổ tần) cần phải được xem xét và mô hình toán dưới dạng bài toán tối ưu. Trong bài toán tối ưu, hàm mục tiêu sẽ đặc trưng cho một trong những thông số đánh giá hiệu năng của mô hình, cụ

thể như tỷ lệ truy cập video thành công (hit rate), dung lượng phân phối video, và chất lượng truyền video. Bài toán tối ưu còn xem xét đến các ảnh hưởng của ràng buộc nhằm tận dụng tốt tài nguyên lưu trữ và phổ tần được chia sẻ trong việc nâng cao hiệu năng mô hình.

Ngoài ra, tùy theo từng mục tiêu của việc đánh giá mô hình, các bài toán tối ưu sẽ được giải bởi những thuật giải phù hợp khác nhau với độ chính xác và thời gian thực thi hợp lý. Cụ thể, thuật giải vét cạn sẽ được áp dụng cho bài toán tối ưu có kích thước hệ thống nhỏ nhằm kiểm chứng tính đúng đắn cũng như cho phép có những đánh giá ban đầu về hiệu năng hệ thống. Thuật giải vét cạn cũng là cơ sở tham chiếu để thiết lập các thông số thực thi cho các thuật giải khác như thuật toán metabeuristic và thuật giải di truyền để đảm bảo độ chính xác. Sau đó, thuật toán metabeuristic và thuật giải di truyền được triển khai cho các bài toán tối ưu phức tạp hơn và kích thước hệ thống lớn hơn đảm bảo yêu cầu về độ chính xác và thời gian thực thi.

1.3. Hiện trạng các nghiên cứu về cơ chế lưu trữ và chia sẻ tài nguyên

1.3.1. Các nghiên cứu trong nước

Nhóm nghiên cứu của Nguyễn Trần Minh Thư và Đồng Thị Bích Thủy (Đại học Khoa học tự nhiên, ĐHQG TP.HCM) đã đề xuất các chiến lược quản lý lưu trữ cho các hệ thống di động không dây. Mục tiêu chính là cải thiện tỷ lệ truy xuất thành công [22, 23] và giảm mức tiêu thụ năng lượng của số lượng MU đang ngày càng tăng, trong khi sử dụng băng thông hệ thống hiệu quả [13, 21]. Tuy nhiên, việc các hệ thống được xem xét đơn giản và thiếu các chỉ số đánh giá hiệu năng hệ thống đã làm cho [13, 21-23] không đủ tốt để áp dụng vào trong 5G UDN. Trong [14], Phạm Tuấn Minh từ Đại học Sư phạm Hà Nội đã áp dụng mô hình lý thuyết trò chơi để kiểm soát lượng nội dung được lưu trữ nhằm đạt được tỷ lệ hiệu quả chi phí cao tại các ISP và CP. Vấn đề của [14] là hiệu suất lưu trữ không được nghiên cứu sâu, mà chỉ có đáp ứng của các CP được phân tích.

Đóng góp quan trọng hơn trong lĩnh vực này phải được đề cập đến đó là nhóm nghiên cứu của Võ Lưu Phương và các cộng sự từ Đại học Quốc tế - ĐHQG TP.HCM,

Học viện Công nghệ Bưu chính Viễn thông, Đại học Thủ Dầu Một, Đại học Tôn Đức Thắng, Đại học Lạc Hồng, Đại học Bách khoa TP.HCM, và các Đại học khác đến từ Hàn Quốc, Mỹ, Phần Lan, Singapore [19, 29, 35, 38, 44]. Trong [35], các tác giả đã xây dựng bài toán tối ưu vị trí lưu trữ nội dung và sau đó giải bài toán để tìm ra số lượng video tối ưu được lưu trữ tương ứng với từng cấp độ/tốc độ mã hóa bit nhất định. Kết quả là số lượng người dùng hài lòng trong hệ thống truyền hình theo yêu cầu được cực đại theo ràng buộc về dung lượng lưu trữ của máy chủ biên (edge server). Tiếp tục phát triển từ [35], số lượng video tối ưu được lưu trữ tại cổng dịch vụ (service gateway) được tính toán nhằm cực đại QoE (được đặt trung bởi hàm Lôgarít của tốc độ bit) [44]. Giải pháp đề xuất cũng cung cấp tỷ lệ truy xuất thành công cao, lưu lượng chuyển tiếp tại các tuyến trực thấp.

Tương tự với [14], các tác giả trong [29] sử dụng mô hình đấu giá ngược (reverse auction) để thiết kế mạng lưu trữ cho mạng hướng nội dung nhằm mang lại lợi ích cho cả ISP và khách hàng bằng cách giảm lưu lượng chuyển tiếp và trì hoãn. Các tác giả cũng phân tích hiệu suất (về tỷ lệ truy xuất thành công và trì hoãn) của các chính sách tìm vị trí lưu trữ nội dung khác nhau trong mạng hướng nội dung để chỉ ra những ưu nhược điểm của từng chính sách [38]. Trong [19], các tác giả đề xuất cơ chế trả phí cho việc lưu trữ dựa trên lý thuyết hợp đồng (contract theory). Mục tiêu là để khuyến khích các ISP sử dụng nhiều dung lượng cho việc lưu trữ hơn nhằm đạt được lợi ích cao nhất. Các công trình của Võ Lưu Phương và các cộng sự [19, 29, 35, 38, 44] tập trung vào việc cung cấp: 1) cho người dùng QoE cao đặc trưng bởi tỷ lệ truy xuất thành công, trì hoãn, và tốc độ bit mã hóa và 2) cho ISP và CP cơ chế tính giá tốt và đem lại lợi ích cao. Tuy nhiên, các tác giả không xem xét vấn đề phân phối nội dung qua môi trường truyền thông vô tuyến, mà ở đó, tài nguyên (năng lượng, phổ và băng thông), đặc tính của kênh truyền vô tuyến, và thậm chí các mối quan hệ của người dùng, có ảnh hưởng quan trọng đến hiệu suất lưu trữ. Ngoài ra, các cơ chế lưu trữ đề xuất được triển khai cơ bản chỉ tại trạm cơ sở, máy chủ biên, hoặc cổng dịch vụ, sẽ không thể áp dụng vào trong kiến trúc không đồng nhất và phân cấp của 5G UDN.

Gần đây nhất, trong [30], Trần Hoàng Nguyên và các cộng sự từ Đại học Sydney, Úc, đã thiết kế một cơ chế trả phí kết hợp giữa lưu trữ nội dung có trả phí và chính sách thay thế nội dung dựa trên giá trong mạng dữ liệu được định danh (named data networking). Mục đích của công trình này gần giống như của [19, 29, 35, 38, 44], nhưng các tác giả trong [30] tập trung vào phân tích và đánh giá sâu hơn. Trần Duy Thịnh và Lê Bảo Long từ Đại học Université du Québec, Canada, đã nghiên cứu sự kết hợp giữa lưu trữ nội dung và phân bổ tài nguyên trong các mạng vô tuyến hướng nội dung được ảo hóa (virtualized content-centric wireless networks) của mạng 5G [31]. Các tác giả đã tận dụng tài nguyên lưu trữ và tài nguyên vô tuyến để cực tiểu tỷ lệ từ chối yêu cầu nội dung cực đại của các nhà mạng di động ảo (mobile virtual network operator) khác nhau trong các tế bào khác nhau. Tuy nhiên, các tác giả không nghiên cứu vấn đề làm thế nào để phân phối nội dung video từ vị trí lưu trữ đến người dùng một cách hiệu quả. Hơn nữa, những khía cạnh khác của QoS như chất lượng hiển thị và cũng như mối quan hệ xã hội của người dùng cũng không được nghiên cứu để cực đại hiệu quả lưu trữ.

1.3.2. Các nghiên cứu ngoài nước

Các công trình nghiên cứu được thực hiện ngoài nước nhiều và đa dạng hơn cũng như có những đóng góp quan trọng hơn so với các công trình được thực hiện trong nước. Để thuận tiện cho việc đánh giá, các công trình này được phân loại theo vị trí lưu trữ khác nhau gồm: lưu trữ tại các thiết bị (device-to-device caching) có chia sẻ tài nguyên phổ tần [15, 20, 33, 42, 45], lưu trữ tại trạm cơ sở nhỏ femto (femtocaching) [26, 27, 34], lưu trữ tại các trạm cơ sở nhỏ (small-cell caching) [17, 32, 37, 41, 43, 46], lưu trữ tại trạm cơ sở lớn (MBS caching) [39, 40], và lưu trữ đa tầng (multi-tier caching) [16, 18, 24, 25, 28, 36, 47, 48], được trình bày chi tiết như sau.

1) *Lưu trữ tại các thiết bị (D2D caching)*: Lưu trữ tại các thiết bị (Device-to-Device – D2D) được xem là cơ chế lưu trữ tại vùng biên của mạng 5G một cách hiệu quả nhằm làm giảm lưu lượng chuyển tiếp tại tuyến trục của các MBS và SBS [15,

20, 33, 42, 45]. Trong [20], C. Xu và các cộng sự tại trường North China Electric Power University, Trung Quốc, đã tận dụng ưu điểm của truyền thông D2D có kết hợp lưu trữ để thiết lập mạng đa chặng D2D. Cụ thể, các tác giả đã áp dụng kỹ thuật thiết kế tối ưu lớp chéo (cross-layer optimization) được thực hiện tại lớp mạng (network layer) bởi thuật giải định tuyến nội dung và tại lớp điều khiển truy cập đa phương tiện (media access control layer) bởi cơ chế phân bổ tài nguyên phổ tần. Kết quả là trì hoãn trung bình được cực tiểu nhằm cung cấp dịch vụ phân phối nội dung nhanh hơn trong mạng đa chặng D2D mà không cần chuyển tiếp lưu lượng qua các tuyến trực của các MBS. Các tác giả từ các trường Universit'e Paris-Sud, Pháp và Northeastern University, Trung Quốc, đã xây dựng các bài toán tối ưu vị trí lưu trữ theo xác suất trong mạng truyền thông D2D ngẫu nhiên [15, 42]. Giải các bài toán này sẽ cực đại xác suất phân phối nội dung. Các kết quả tối ưu đạt được cho thấy rằng cách tiếp cận cực đại hóa dựa trên thông lượng có hỗ trợ lưu trữ sẽ tốt hơn dựa trên xác suất truy xuất thành công [42] và giải pháp khả thi để đạt được xác suất phân phối nội dung cao nhất là cơ chế lưu trữ tập trung vào các nội dung có độ phổ biến cao [15].

Bằng cách xem xét đến sự di chuyển của các MU trong thực tế, R. Wang và các cộng sự từ trường Hong Kong University of Science and Technology, Hồng Kông, đã xây dựng và giải bài toán tối ưu vị trí lưu trữ [33]. Kết quả khẳng định rằng giải pháp đề xuất đạt được tỷ lệ tải dữ liệu không thông qua tuyến trực (offloading) cao hơn so với giải pháp lưu trữ theo độ phổ biến của nội dung và lưu trữ ngẫu nhiên. Trong [45], các tác giả từ trường Pohang University of Science and Technology, Hàn Quốc, đã thiết kế hệ thống thông minh tải dữ liệu một phần không thông qua tuyến trực hỗ trợ MBS cung cấp các dịch vụ truyền video không gián đoạn. Hệ thống này có khả năng quản lý các nội dung video, các MU đã lưu trữ các video, và các MU đang yêu cầu nội dung video, để tính toán cho quá trình thực hiện tải video không thông qua tuyến trực. Bằng cách này, hệ thống cho phép đạt được mục tiêu về lưu lượng nhằm cung cấp cho MU các dịch vụ video không gián đoạn. Tuy nhiên, các giải pháp lưu trữ trên không tận dụng được hết tài nguyên dung lượng lưu trữ và phổ

của các thiết bị di động, SBSs, và MBS để cực đại dung lượng hệ thống và từ đó hợp tác phân phối các video đến MU đạt được QoS cao.

2) *Lưu trữ tại trạm cơ sở nhỏ femto (FBS caching)*: Nhờ vào dung lượng lớn, lưu trữ tại các trạm cơ sở nhỏ femto (FBS – Femto cell Base Station) là một trong những giải pháp hữu hiệu nhằm giảm thiểu xung đột tải tại MBS [26, 27, 34]. Cụ thể, Negin Golrezaei và các cộng sự từ trường University of Southern California, Los Angeles, CA, đã thiết kế một kiến trúc mạng vô tuyến mới để phân phối video [26]. Mục đích của công trình này nhằm tăng thông lượng của hệ thống bằng cách kết hợp giữa kỹ thuật truyền thông D2D và lưu trữ tại FBS. Nhằm cực tiểu thời gian trung bình để tải nội dung đa phương tiện và cực đại số lượng MU được phục vụ bởi các FBS, các tác giả từ trường University of Texas at Austin, Austin, TX và trường King Fahd University of Petroleum & Minerals, Ả Rập Xê Út, đã giải bài toán vị trí lưu trữ trong các FBS theo độ phổ biến của các nội dung đa phương tiện [27, 34]. Những cơ chế lưu trữ tại FBS cũng có các vấn đề còn tồn đọng tương tự như trong lưu trữ D2D.

3) *Lưu trữ tại các trạm cơ sở nhỏ (SBS Caching)*: Hiệu quả hơn so với lưu trữ tại FBS, lưu trữ tại các trạm cơ sở nhỏ (SBS – Small-cell Base Station) tận dụng không những tài nguyên dung lượng lưu trữ tại các FBS mà còn tại các trạm nên nhỏ khác như pico (PCSs) và micro [17, 32, 37, 41, 43, 46]. Trong [41, 46], các tác giả từ trường CentraleSupélec, Pháp và trường Beijing University of Posts and Telecommunications, Trung Quốc, đã đề xuất các cơ chế lưu trữ thụ động (proactive caching) triển khai tại các SBS. Dựa trên việc lưu trữ các nội dung có độ phổ biến cao có xem xét đến mối quan hệ xã hội của MU, các cơ chế lưu trữ có thể nâng cao thông lượng, giảm trì hoãn, cải thiện tỷ lệ thỏa mãn, và làm giảm tải chuyển tiếp tại các tuyến trực của MBS. Y. Chen và các cộng sự tại trường University of Sydney, Australia, đã ứng dụng lý thuyết hình học ngẫu nhiên vào trong lưu trữ tại SBS [43]. Các tác giả đã tính toán được xác suất tải lý thuyết, từ đó có được xác suất lưu trữ tối ưu cho các nhóm lưu trữ khác nhau. Trong [17, 37], các tác giả từ trường University College London, UK, đã giải bài toán tìm vị trí lưu trữ tối ưu tại các SBS để trực tiếp cực tiểu lưu lượng chuyển tiếp qua tuyến trực của các MBS. Đặc biệt, các cơ chế phân

phối cũng được nghiên cứu theo phương thức truyền thông điểm – đa điểm (multicast) tận dụng sự độc lập giữa các gói được mã hóa phân tách có khoảng cách cực đại (MDS – maximum distance separable) để làm giảm hơn nữa tải lưu lượng tại các tuyến trục này. Được thực hiện bởi Emre Ozfatura và Deniz Gündüz từ trường Imperial College London, Vương Quốc Anh, theo cách tiếp cận thực tế hơn về chính sách lưu trữ phân phối tối ưu, các yếu tố về sự di chuyển của các MU, giới hạn trì hoãn, và độ phổ biến của nội dung được xem xét [32]. Kết quả là tải lưu lượng tại các tuyến trục của MBS được cực tiểu một cách hiệu quả. Tuy nhiên, các công trình [17, 32, 37, 41, 43, 46] không xem xét đến khả năng lưu trữ tại các MBS để cực đại dung lượng hệ thống cũng như không xem xét hết các khía cạnh của QoS và các khía cạnh của hiệu quả tài nguyên để có thể đánh giá sâu hơn về hiệu quả của việc lưu trữ.

4) *Lưu trữ tại trạm cơ sở lớn (MBS caching)*: Lưu trữ tại trạm cơ sở lớn (MBS – Macro Base Station) là một trong những cơ chế lưu trữ đơn giản nhưng hiệu quả để thay thế cho việc đầu tư tuyến trục tốc độ cao nhưng rất tốn kém [39, 40]. Trong [39], Wei Han và các cộng sự từ Hong Kong University of Science and Technology, Hồng Kông, đã thiết kế một chiến lược lưu trữ nội dung trực tuyến dựa trên mã MDS. Kết quả nghiên cứu cho phép các ISP xem xét việc lưu trữ tại các MBS và/hoặc tại mạng lõi dung lượng cao nhằm đạt được lợi ích cao nhất có thể. Kết quả chỉ ra rằng để phục vụ số lượng MU cao nhất và hiệu quả phổ cao, lưu trữ tại MBS tốt hơn lưu trữ tại mạng lõi. Ngoài ra, các tác giả trong [40] đã xem xét nhiều khía cạnh của QoS nhằm đề xuất một khung giải pháp mmWave dựa trên lưu trữ thụ động tại MBS. Kết quả đã cho thấy giải pháp có khả năng cung cấp cho các MU có tốc độ di chuyển cao và thường hay chuyển trạm các dịch vụ truyền video với độ trễ truy xuất thấp, tính liên tục cao, và chất lượng hiển thị cao. Vấn đề còn tồn tại của [39, 40] đó là chưa tận dụng được dung lượng lưu trữ tại các SBS và các MU để cải tiến hiệu quả lưu trữ tại MBS.

5) *Lưu trữ đa tầng (Multi-tier caching)*: Trong lưu trữ đa tầng, nội dung có thể được lưu trữ tại MBS, SBS, và MU. Do đó, lưu trữ đa tầng sẽ cho hiệu quả cao nhất, nhưng phức tạp nhất [16, 18, 24, 25, 28, 36, 47, 48]. Cụ thể, Konglin Zhu và các

cộng sự từ trường Beijing University of Posts & Telecommunications, Trung Quốc, đã mở rộng vị trí lưu trữ từ MBS đến MU để giải quyết vấn đề xung đột tại tuyến trục của MBS [47]. Tác giả và cộng sự đã ứng dụng lý thuyết trò chơi nhằm cực đại lợi ích của nhóm người dùng di động có mối quan hệ xã hội với nhau vào bài toán trả phí cho lưu trữ. Giải pháp này cho phép thúc đẩy các SBS (hay FBS) và MU lưu trữ nội dung nhiều hơn bằng cách trả phí lưu trữ sao cho tối đa được lợi ích nhóm các người dùng di động có mối quan hệ với nhau với chi phí lưu trữ là thấp nhất. Lưu trữ nội dung tại FBS (hoặc MBS), tại các MU và việc phân phối nội dung được nghiên cứu bởi các tác giả đến từ University of Electronic Science & Technology of China và Northeastern University, Trung Quốc [36] (hoặc [25]). Các tác giả đã xây dựng và giải bài toán tối ưu việc phân phối và lưu trữ nội dung hợp tác để tìm ra tập hợp các vị trí lưu trữ tốt nhất tại FBS (hoặc MBS) và tại MU. Mục đích của giải pháp là để giảm độ trễ truy xuất trung bình và tăng tỷ lệ truy xuất thành công.

Trong các nghiên cứu [24, 28, 48], bằng cách xem xét đến tài nguyên năng lượng trong giải pháp truyền thông xanh, các tác giả từ trường Huazhong University of Science & Technology, Trung Quốc, Centre Tecnològic de Telecomunicacions de Catalunya, Tây Ban Nha, và University of Luxembourg, Luxembourg, đã cực đại tỷ lệ truy xuất thành công [28] và giảm tải tại các tuyến trục [24, 48] đồng thời đảm bảo hiệu quả sử dụng năng lượng cao. Điều này có thể đạt được thông qua việc có quan tâm đến sự di chuyển của MU [28] và kết hợp các cơ chế lưu trữ và truyền tại MBS và MU một cách hiệu quả [24, 48]. Quan trọng hơn, các tác giả từ trường University of Hong Kong, Hồng Kông và trường University of British Columbia, Canada, đã nghiên cứu cơ chế lưu trữ phức tạp được triển khai tại cả MBS, PCS, và FBS [16, 18]. Kết quả đạt được là tỷ lệ truy xuất thành công [16] và dung lượng hệ thống [18] được cực đại. Mặc dù [16, 18] đã xây dựng một kiến trúc lưu trữ khá tổng quát (chỉ thiếu lưu trữ ở MU), các tác giả vẫn chưa nghiên cứu sâu về các cơ chế lưu trữ và truyền theo: 1) hiệu quả sử dụng tài nguyên (dung lượng lưu trữ, băng thông, năng lượng và phổ), 2) dung lượng hệ thống và các khía cạnh khác của QoS như tỷ lệ truy xuất thành công cao, tính liên tục, chất lượng hiển thị cao, và tính ổn định chất lượng,

3) đặc tính hành vi và mối quan hệ xã hội của MU cũng như năng lực của thiết bị người dùng, nhằm cực đại hiệu quả truyền video trong 5G UDN.

1.4. Các thuật giải ứng dụng trong các bài toán tối ưu lưu trữ và truyền video

Như đã trình bày ở trên, tối ưu lưu trữ kết hợp chia sẻ tài nguyên là kỹ thuật đã và đang được nhiều nhà nghiên cứu thực hiện ứng dụng trong truyền video qua 5G UDN đạt hiệu quả cao với chi phí thấp. Các bài toán tối ưu lưu trữ này thường được giải bằng cách tìm các chỉ số xác định vị trí lưu trữ và chọn lựa/phân bổ tài nguyên trong hệ thống dưới dạng vector hoặc ma trận và các chỉ số thường được gán giá trị 1 (hoặc 0) có nghĩa là hệ thống có (hoặc không) lưu trữ hoặc/và có (hoặc không) chia sẻ tài nguyên. Để tìm kết quả tối ưu, các thuật giải tìm kiếm tối ưu đã được đề xuất để giải các bài toán trên như tìm kiếm vét cạn [18, 54-57], quy hoạch động [50], thuật giải học ngẫu nhiên [40], thuật giải trò chơi [43], thuật giải tham lam [58], và một số thuật giải heuristic [25, 59].

Các thuật giải vét cạn đã sử dụng ở các nghiên cứu trên mặc dù cho ra kết quả tối ưu chính xác nhưng sẽ tốn nhiều thời gian để thực thi và độ phức tạp trong tính toán cao. Do đó, rất khó để triển khai thuật giải vét cạn trong quy mô lớn của mạng 5G UDN. Trong khi đó, các phương pháp khác được đề xuất trong [18, 25, 39, 40, 43, 50, 58, 59] thì thời gian thực thi nhanh hơn khá nhiều với độ phức tạp thấp hơn so với thuật giải vét cạn nhưng độ chính xác lại thấp, khoảng 81% đến 87% và thậm chí là 50% trong trường hợp xấu [58, 59] khi so với thuật giải vét cạn. Ngoài ra, vấn đề khác của các thuật giải được nghiên cứu trong các bài báo trên là các ràng buộc được xem xét là tương đối đơn giản ở dạng tuyến tính và có ràng buộc là giới hạn dưới và trên.

Từ các lý do trên, thuật giải di truyền (GA – Genetic Algorithm), là một dạng phương pháp tìm kiếm thích nghi, sẽ được điều chỉnh để áp dụng cho phù hợp với không gian tìm kiếm và các ràng buộc phức tạp của bài toán tối ưu lưu trữ và chia sẻ tài nguyên ở các chương tiếp theo. Hơn nữa, các bài toán tối ưu này được giải để tìm ra nghiệm nhị phân là hoàn toàn phù hợp với đặc trưng nhị phân của từng cá thể

(nghiệm khả thi) của quần thể nghiệm được tiến hóa qua các thế hệ trong GA. Đồng thời, thuật giải đàn dơi (BA – Bat Algorithm) cũng được thực hiện để khẳng định tính hiệu quả của thuật giải GA. Ngoài ra, thuật giải EA cũng được sử dụng vì đơn giản, cho kết quả chính xác, phù hợp cho bài toán có không gian tìm kiếm nhỏ, cũng như được dùng để làm thước đo độ chính xác cho các thuật giải GA và BA.

1.4.1. Thuật giải vét cạn

Tìm kiếm vét cạn (exhaustive search) là một kỹ thuật giải quyết vấn đề tổng quát, trong đó, mô hình thuật giải xét hết tất cả các ứng cử viên có trong không gian tìm kiếm trong mọi trường hợp để kiểm tra xem từng ứng viên đó có đáp ứng yêu cầu của bài toán hay không.

Mặc dù tìm kiếm vét cạn dễ thực hiện và sẽ luôn tìm ra giải pháp nếu nó tồn tại, nhưng chi phí (độ phức tạp và thời gian tính toán) tỷ lệ thuận với số lượng ứng viên nghiệm (hay kích thước bài toán) trong không gian tìm kiếm, và trong nhiều trường hợp chi phí có xu hướng phát triển rất nhanh theo cấp số nhân (gọi là bùng nổ tổ hợp). Do đó, tìm kiếm vét cạn thường được sử dụng khi kích thước bài toán bị hạn chế hoặc khi có các kinh nghiệm về nghiệm cụ thể của bài toán có thể được sử dụng để làm giảm số lượng các ứng viên nghiệm xuống kích thước khả thi. Phương pháp này cũng được sử dụng khi tính đơn giản của việc thực hiện quan trọng hơn tốc độ.

1.4.2. Thuật giải đàn dơi

1.4.2.1. Giới thiệu về thuật giải đàn dơi

Dơi là động vật có thể bay được và chuyên sống về đêm. Do dơi có khả năng nhìn kém, nên chúng sử dụng một khả năng đặc biệt để di chuyển và xác định con mồi đó là sử dụng khả năng định vị bằng sóng âm. Với khả năng định vị bằng sóng âm siêu việt của mình, dơi sử dụng chúng để xây dựng bản đồ về môi trường xung quanh. Nhờ vậy, dơi có thể bay lượn mà không va chạm với các chướng ngại vật. Một khả năng nữa là bằng cách định vị bằng sóng âm, dơi có thể xác định được đâu là con mồi, đâu là vật cản (bức tường, cành cây, v.v.), và đâu là trái cây có thể ăn.

Những con dơi thường phát ra từ 10 đến 20 xung sóng âm mỗi giây nhưng khi phát hiện con mồi thì dơi sẽ tăng số lượng xung (có thể lên tới 200 xung mỗi giây) và giảm âm lượng. Dựa trên các đặc điểm định vị bằng sóng âm và hành vi của dơi trong tự nhiên, thuật giải BA đã được Xin-She Yang phát triển [60, 61] để giải các bài toán tối ưu toàn cục.

BA là một thuật giải siêu heuristic (metaheuristic) dựa trên hành vi định vị bằng sự phản hồi sóng âm của dơi. Thuật giải được mô hình hóa đơn giản dựa trên các giả định như sau:

- Các con dơi bay ngẫu nhiên với vận tốc cố định và ở các vị trí ngẫu nhiên khác nhau. Sóng âm của chúng phát ra có tần số, bước sóng và âm lượng khác nhau mà chúng sử dụng để tìm kiếm con mồi. Tốc độ (rate) phát xung và tần số (frequency) của xung được điều chỉnh dựa trên khoảng cách của dơi đến con mồi.
- Các con dơi sử dụng khả năng định vị bằng sự phản hồi của sóng âm (echolocation) để phát hiện cũng như phân biệt giữa con mồi (thức ăn) và các vật thể khác, và dơi cũng dùng sóng âm để xác định được khoảng cách.
- Âm lượng (loudness) của sóng âm do dơi phát ra thay đổi từ giá trị lớn nhất đến giá trị nhỏ nhất.

Trong luận án, các bài toán đều là những bài toán tìm kiếm nhị phân. Vì vậy, mô hình thuật toán đàn dơi trong bài toán tìm kiếm nhị phân (BBA – Binary Bat Algorithm) sẽ được áp dụng.

1.4.2.2. Thuật giải đàn dơi nhị phân

Gọi f_{min} và f_{max} lần lượt là tần số sóng âm cực tiểu và cực đại của dơi. Tốc độ phát xung có thể được giả định nằm trong đoạn $[0, 1]$, trong đó, 0 thể hiện không có xung nào được phát ra và 1 thể hiện tốc độ phát xung tối đa.

Gọi N_{PB} là kích thước của đàn dơi (số lượng dơi), giá trị nhị phân của chiều thứ j trong vị trí của dơi thứ i ở lần lặp thứ $iter$ trong không gian tìm kiếm d chiều là

x_{ij}^{iter} và vận tốc tương ứng là $v_{ij}^{iter}, j = 1, 2, \dots, d$. Những con dơi di chuyển trong không gian tìm kiếm để tìm sự tối ưu toàn cục. Các phương trình cập nhật vị trí mới của những con dơi để khám phá không gian tìm kiếm là:

$$v_{ij}^{iter+1} = v_{ij}^{iter} + (x_{ij}^{iter} - Gbest_{ij}^{iter})f_{ij} \quad (1.1)$$

$$V(v_{ij}^{iter}) = \left| \frac{\pi}{2} \arctan\left(\frac{\pi}{2} v_{ij}^{iter}\right) \right| \quad (1.2)$$

$$x_{ij}^{iter+1} = \begin{cases} \overline{x_{ij}^{iter}} & \text{if } rand < V(v_{ij}^{iter+1}) \\ x_{ij}^{iter} & \text{if } rand \geq V(v_{ij}^{iter+1}) \end{cases} \quad (1.3)$$

với $Gbest_{ij}^{iter}$ là vị trí tối ưu ở chiều thứ j trong vòng lặp hiện tại của dơi thứ i , $V(\cdot)$ là hàm truyền dạng chữ V [61], $rand$ có giá trị ngẫu nhiên phân bố đều trong khoảng $(0, 1)$ và f_{ij} là tần số xung ở chiều thứ j của dơi thứ i , được tính bởi:

$$f_{ij} = f_{min} + r_{ij}(f_{max} - f_{min}) \quad (1.4)$$

trong đó r_{ij} được phát ngẫu nhiên theo phân bố đều trong khoảng $(0, 1)$. Trong quá trình khởi tạo, các tần số được gán ngẫu nhiên cho các con dơi trong phạm vi đã chọn.

BBA tương tự như thuật giải bầy đàn (PSO) ở phương trình cập nhật vị trí và vận tốc. Âm lượng và tốc độ phát xung của dơi thay đổi theo khoảng cách của dơi đến con mồi trong quá trình tìm kiếm. Khi tiếp cận con mồi, dơi sẽ giảm âm lượng và tăng tốc độ phát xung. Cho âm lượng của dơi thứ i ở lần lặp $iter$ là L_i^{iter} và tốc độ phát xung của dơi thứ i ở lần lặp đầu tiên là E_i^0 , việc cập nhật âm lượng và tốc độ phát xung được tính như sau:

$$L_i^{iter+1} = \varepsilon L_i^{iter} \quad (1.5)$$

$$E_i^{iter+1} = E_i^0(1 - e^{-\varepsilon(iter)}) \quad (1.6)$$

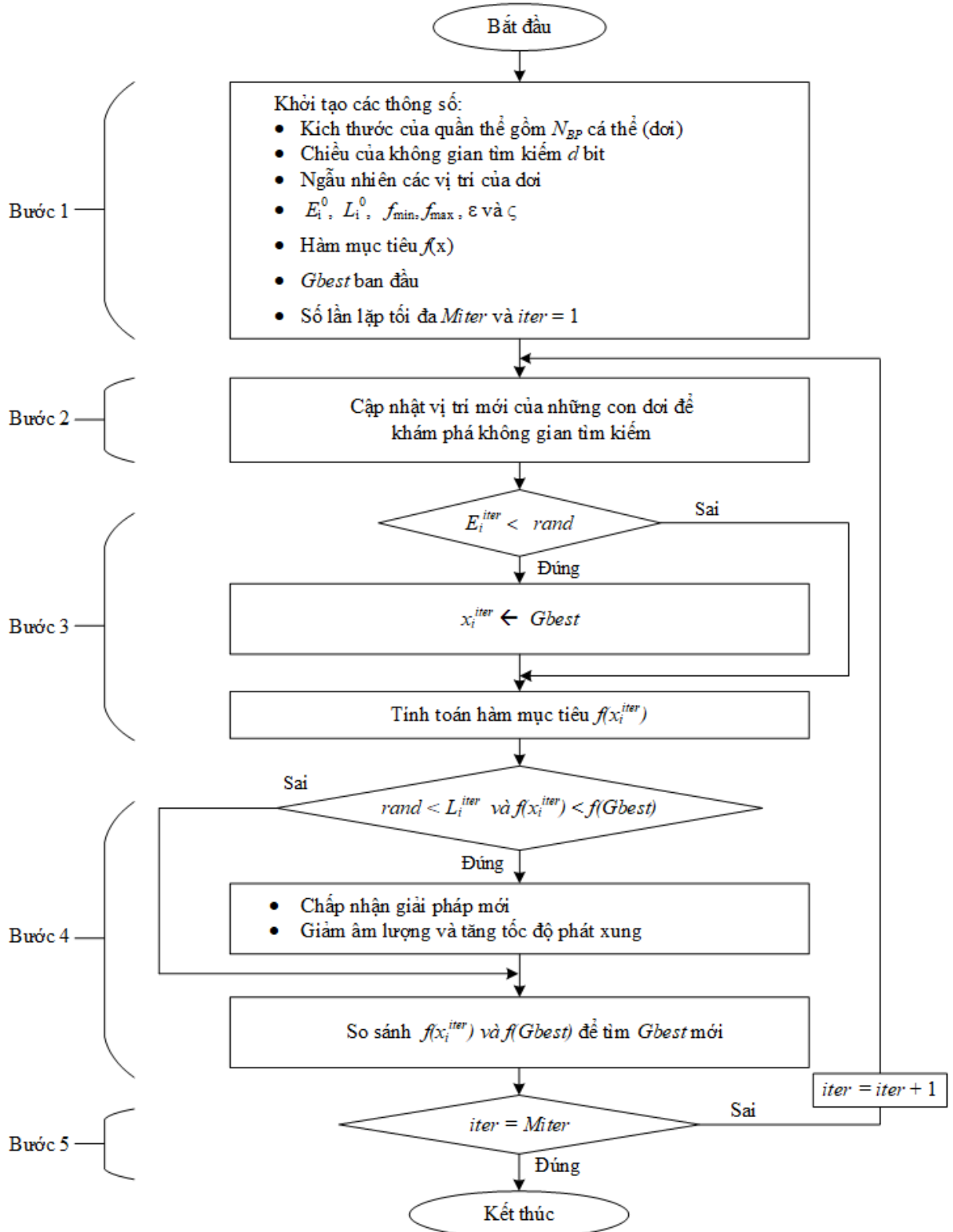
trong đó $0 < \varepsilon < 1$ là hệ số điều chỉnh giảm âm lượng, khi $iter \rightarrow \infty$ thì $L_i^{iter} \rightarrow 0$ (âm lượng tiến về 0) và ζ là hệ số mũ điều chỉnh tăng tần số phát xung, khi $iter \rightarrow \infty$ thì $E_i^{iter} \rightarrow E_i^0$ (tần số phát xung là cực đại).

Nếu giải pháp mới tốt hơn giải pháp trước đó, thì âm lượng và tốc độ của xung sẽ được cập nhật. Điều này cho thấy rằng thuật giải tiến tới giải pháp tối ưu hay nói cách khác là dơi tiếp cận gần đến con mồi.

Lưu đồ thực hiện BBA được mô tả như sau (Hình 1-3):

Bước 1: Đầu tiên, các thông số được khởi tạo. Trong bước này, quần thể gồm các cá thể (dơi) được khởi tạo ngẫu nhiên, trong đó mỗi cá thể là một chuỗi nhị phân có giá trị '0' và '1' đại diện cho các hướng di truyền của dơi. Các thông số khác cũng được khởi tạo như: hệ số điều chỉnh giảm âm lượng ε , hệ số mũ điều chỉnh tăng tần số phát xung ζ , tần số xung, tốc độ phát xung và âm lượng của xung. Ngoài ra, hàm mục tiêu được tính toán tương ứng với từng cá thể để chọn ra cá thể tốt nhất Gbest ban đầu.

Bước 2: Tiếp theo, vị trí của các con dơi được cập nhật theo các biểu thức (1.1), (1.2), (1.3) và (1.4).



Hình 1-3. Lưu đồ thuật giải đàn dơi nhị phân

Bước 3: Kế đến, lựa chọn con dơi có tần số phát xung cao và tính toán hàm mục tiêu tương ứng.

Bước 4: Lựa chọn con dơi có âm lượng giảm và có hàm mục tiêu tốt hơn con dơi tốt nhất ở vòng lặp trước đó. Đồng thời, giảm âm lượng và tăng tốc độ phát xung theo biểu thức (1.5) và (1.6).

Bước 5: Thực hiện cho đến khi hết số lần lặp *Miter*.

1.4.3. Thuật giải di truyền

Thuật giải di truyền (GA – Genetic Algorithm) được phát triển bởi John Holland và các đồng nghiệp của ông tại Đại học Michigan, là một phương pháp tìm kiếm toàn cầu ngẫu nhiên bắt chước quá trình tiến hóa sinh học tự nhiên. GA hoạt động dựa trên một tập hợp các giải pháp tiềm năng áp dụng nguyên tắc tồn tại của người phù hợp nhất để tạo ra các giá trị tốt hơn cho một giải pháp. Trong GA, ở mỗi thế hệ, một tập hợp các giá trị tốt hơn mới được tạo ra bằng quá trình chọn các cá thể theo mức độ phù hợp của chúng và lai ghép chúng với nhau bằng cách sử dụng các phương pháp mượn từ di truyền tự nhiên. Quá trình này dẫn đến sự tiến hóa của các cá thể trong quần thể phù hợp với môi trường của chúng hơn so với các cá thể đã dùng để lai ghép, giống như sự thích nghi trong tự nhiên.

Trong thuật giải di truyền, các cá thể được biểu diễn dưới dạng chuỗi nhiễm sắc thể², được mã hóa dưới dạng nhị phân '0' và '1', hoặc có thể biểu diễn khác dưới dạng số trị thực. Với phép biểu diễn này, giải pháp cho một bài toán được biểu diễn như là một vector bit và mỗi nhiễm sắc thể bao gồm nhiều gen, trong đó một gen đại diện cho một tham số thành phần của giải pháp. Với kiểu biểu diễn là số thực, các toán tử tiến hóa sẽ thực hiện trực tiếp trên các giá trị số thực.

1.4.3.1. Thuật giải di truyền so với các thuật giải khác

Từ các phân tích ở trên, có thể thấy rằng GA khác biệt đáng kể với các phương pháp tìm kiếm và tối ưu hóa khác. Bốn khác biệt đáng kể nhất là:

² Trong thuật giải di truyền, mỗi cá thể trong quần thể được biểu diễn dưới dạng chuỗi nhị phân ('0' và '1') đặc trưng hoặc dưới dạng số thực đặc trưng, được gọi là nhiễm sắc thể.

- GA cho phép tìm kiếm song song các điểm trong một quần thể. Điều này giúp tăng cơ hội tiếp cận tối ưu toàn cục và tránh hội tụ sớm tại các cực trị địa phương.
- GA không yêu cầu thông tin hoặc kiến thức phụ trợ khác mà chỉ có hàm mục tiêu và mức độ thích nghi tương ứng ảnh hưởng đến các hướng tìm kiếm.
- GA sử dụng quy tắc xác suất, không phải quy tắc xác định.
- GA làm việc dựa trên bộ mã của biến chứ không phải làm việc trực tiếp trên biến (ngoại trừ trong đó sử dụng các cá thể có giá trị thực).

1.4.3.2. Các thành phần của thuật giải di truyền

Thuật giải di truyền gồm có bốn quy luật cơ bản là chọn lọc, lai ghép, đột biến và sinh sản.

❖ Chọn lọc (Selection)

Chọn lọc là quá trình loại bỏ những cá thể xấu và giữ lại các cá thể tốt trong quần thể nhằm chuẩn bị cho quá trình lai tạo. Việc lựa chọn các cá thể có thể được xem như hai quá trình riêng biệt:

- 1) Sắp xếp các cá thể trong quần thể theo thứ tự độ thích nghi giảm dần.
- 2) Loại bỏ các cá thể cuối dãy, nghĩa là loại bỏ những cá thể có độ thích nghi thấp, và chỉ để lại các cá thể tốt nhất.

❖ Lai ghép (Crossover)

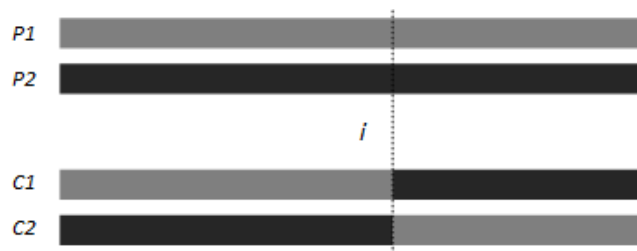
Lai ghép (lai tạo) là quá trình hình thành nhiễm sắc thể mới trên cơ sở ghép một hay nhiều đoạn gen từ hai nhiễm sắc thể cha-mẹ để hình thành nhiễm sắc thể mới mang đặc tính của cả cha và mẹ. Phép lai này có thể mô tả như sau: Chọn ngẫu nhiên hai hay nhiều cá thể trong quần thể. Giả sử chuỗi nhiễm sắc thể của cha và mẹ đều có chiều dài là m . Tìm điểm lai bằng cách tạo ngẫu nhiên một con số từ 1 đến $m-1$. Như vậy, điểm lai này sẽ chia hai chuỗi nhiễm sắc thể cha-mẹ thành hai nhóm nhiễm sắc thể con là $m1$ và $m2$. Hai chuỗi nhiễm sắc thể con lúc này sẽ là $m11+m22$ và

$m2l+m12$. Đưa hai chuỗi nhiễm sắc thể con vào quần thể để tiếp tục tham gia quá trình tiến hóa. Một số kỹ thuật lai ghép thường dùng:

- **Lai ghép đơn điểm (Single-point Crossover):** Đây là kỹ thuật lai ghép đơn giản nhất. Lai ghép đơn điểm nghĩa là chỉ có 1 vị trí lai ghép. Cho 2 chuỗi $P1$ và $P2$ có độ dài l như sau:

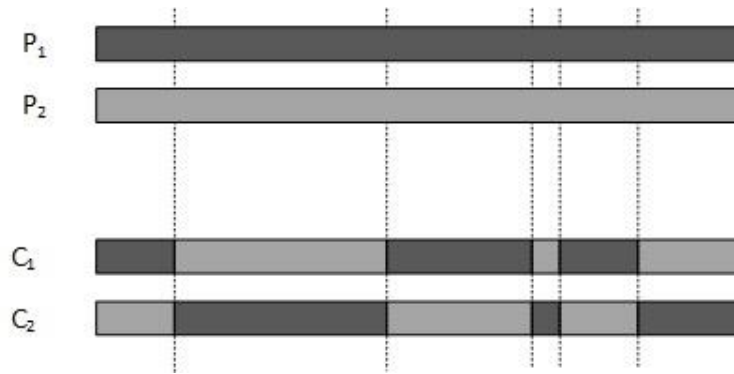


Một điểm lai ghép ngẫu nhiên được chọn tại vị trí i , khi đó, đầu đầu của chuỗi $P1$ $[1, i]$ sẽ ghép với phần sau của chuỗi $P2$ $[i+1, l]$ cho ra chuỗi con $C1$, và ngược lại phần đầu của chuỗi $P2$ $[1, i]$ ghép với phần sau của chuỗi $P1$ $[i+1, l]$ cho ra chuỗi $C2$ (Hình 1-4).



Hình 1-4. Lai ghép đơn điểm

- **Lai ghép đa điểm (Multi-point Crossover):** Lai ghép đa điểm nghĩa là sẽ có m vị trí lai ghép. Giả sử $k_i \in \{1, 2, \dots, l-1\}$, với k_i là điểm lai và l là độ dài nhiễm sắc thể được chọn ngẫu nhiên không có trùng lặp và được sắp xếp theo thứ tự tăng dần. Sau đó, các bit giữa các điểm lai liên tiếp được trao đổi giữa hai cha mẹ để tạo ra hai con cái mới. Phần giữa vị trí nhiễm sắc thể đầu tiên và điểm lai đầu tiên không được trao đổi giữa các cá thể. Quá trình này được minh họa trong Hình 1-5.



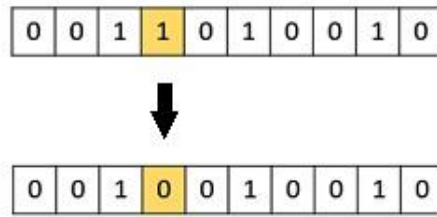
Hình 1-5. Lai ghép đa điểm ($m=5$)

Ngoài ra, còn có các phương pháp lai tạo khác có thể được sử dụng như lai tạo đồng nhất (Uniform Crossover), tái hợp trung gian (Intermediate Recombination) và tái hợp đường thẳng (Line Recombination) [62].

❖ Đột biến (Mutation)

Trong tiến hóa tự nhiên, đột biến là một quá trình ngẫu nhiên trong đó một hay nhiều gen được thay thế bằng gen khác để tạo ra cấu trúc di truyền mới nhằm duy trì sự đa dạng di truyền từ thế hệ này sang thế hệ tiếp theo. Trong thuật giải di truyền, đột biến được áp dụng ngẫu nhiên với xác suất thấp hơn nhiều so với xác suất lai ghép, thường là trong phạm vi 10^{-6} đến 10^{-3} tùy vào số cá thể trong quần thể cũng như độ dài của nhiễm sắc thể. Vai trò của đột biến trong thuật giải di truyền được coi là nỗ lực để tránh cực tiểu cục bộ bằng cách ngăn không cho quần thể nhiễm sắc thể trở nên quá giống nhau hoặc là những gen tốt bị mất thông qua hoạt động chọn lọc và lai ghép, khi đó sẽ làm chậm thậm chí ngừng hội tụ đến mức tối ưu toàn cục.

Ảnh hưởng của đột biến lên chuỗi nhị phân được minh họa trong Hình 1-6. Ở đây, đột biến nhị phân sẽ bù (đảo) giá trị của bit tại vị trí được chọn làm điểm đột biến. Đột biến thường được áp dụng cho toàn bộ các chuỗi (nhiễm sắc thể), và có thể một chuỗi nhất định có thể bị đột biến tại nhiều hơn một điểm.



Hình 1-6. Đột biến bù nhị phân

Một vài phương pháp đột biến thường dùng:

- Đột biến bù nhị phân (Inversion Mutation)
- Đột biến chèn (Insertion Mutation)
- Đột biến thay thế (Displacement Mutation)
- Đột biến tương hỗ (Reciprocal Exchange)
- Đột biến chuyển dịch (Shift Mutation)

❖ Quá trình sinh sản (Reinsertion)

Khi một quần thể mới đã được tạo ra bằng cách lựa chọn và lai ghép các cá thể từ quần thể cũ, các cá thể thích hợp trong quần thể mới có thể được xác định. Nếu kích thước của quần thể mới nhỏ hơn kích thước quần thể cũ thì hệ số khác biệt giữa kích thước quần thể mới và cũ được gọi là khoảng cách thế hệ. Trong trường hợp số lượng cá thể mới được tạo ra ở mỗi thế hệ là một hoặc hai, thuật giải di truyền được xem là ổn định hoặc tiến đến ổn định. Nếu một hoặc nhiều cá thể phù hợp nhất được cho phép chuyển qua các thế hệ kế tiếp thì thuật giải di truyền được cho là sử dụng chiến lược tìm kiếm tinh hoa.

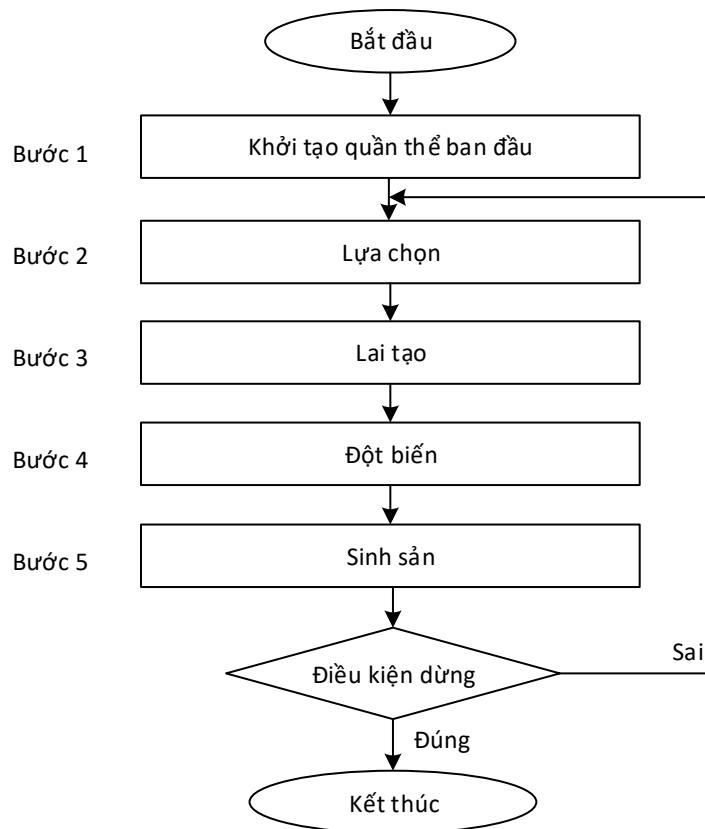
Để duy trì quy mô quần thể ban đầu, các cá thể mới phải được bổ sung vào quần thể cũ. Tương tự, nếu không phải tất cả các cá thể mới sẽ được sử dụng ở mỗi thế hệ hoặc nếu tạo ra nhiều con cái hơn kích thước của quần thể cũ thì mô hình tái tạo phải được sử dụng để xác định cá thể nào tồn tại trong quần thể mới. Một tính năng quan trọng của việc không tạo ra nhiều con cái mới hơn so với quy mô quần thể hiện tại ở mỗi thế hệ là thời gian tính toán thế hệ được giảm xuống, đáng kể nhất là

trong trường hợp thuật giải di truyền ở trạng thái ổn định và yêu cầu ít bộ nhớ hơn vì cần ít cá thể mới hơn được lưu trữ trong khi con cái mới được sinh ra.

Khi lựa chọn thành viên nào của quần thể cũ nên được thay thế, chiến lược rõ ràng nhất là thay thế các thành viên ít phù hợp nhất. Kế hoạch thay thế thành công nhất là kế hoạch đã chọn những thành viên lớn tuổi nhất trong quần thể để thay thế. Điều này được cho là phù hợp hơn với tái sản xuất thế hệ vì mọi lúc, mọi thành viên trong quần thể sẽ được thay thế. Do đó, để một cá nhân sống sót qua các thế hệ kế tiếp, nó phải đủ phù hợp để đảm bảo phát triển trong các thế hệ tương lai.

1.4.3.3. Các bước cơ bản của thuật giải di truyền

Thuật giải di truyền tổng quát gồm có 5 bước như mô tả trong Hình 1-7.



Hình 1-7. Lưu đồ thuật giải di truyền

Bước 1 - Khởi tạo: Trong bước này, quần thể gồm các cá thể ban đầu được khởi tạo ngẫu nhiên. Mỗi cá thể được đặc trưng bằng một chuỗi nhiễm sắc thể (một chuỗi nhị phân). Số cá thể trong quần thể, độ dài

của nhiễm sắc thể (số phần tử trong chuỗi nhị phân), xác suất lai ghép, xác suất đột biến và khoảng cách thế hệ được thiết lập. Sau đó, các giá trị hàm mục tiêu được tính toán tương ứng với các chuỗi nhị phân của từng cá thể.

Bước 2 - Chọn lọc: Dựa trên khoảng cách thế hệ đã khởi tạo ở trên, một nhóm các cá thể cho giá trị thích nghi cao hơn được lựa chọn, xếp hạng để đưa vào danh sách lai ghép.

Bước 3 - Lai ghép: Quá trình lai tạo được thực hiện bằng cách chọn ngẫu nhiên hai chuỗi nhị phân cha-mẹ trong nhóm đã được chọn lọc và xếp hạng với một xác suất lai ghép đã khởi tạo ở trên. Bằng cách này, hai cá thể con được sinh ra sẽ thừa hưởng một số đặc điểm di truyền từ cha mẹ của chúng.

Bước 4 - Đột biến: Quá trình đột biến được áp dụng bằng cách thay đổi một hay nhiều bit (gen) bất kỳ với một xác suất đột biến được khởi tạo ở trên để phục hồi đặc điểm di truyền tốt có khả năng bị mất trong các quá trình trước đó.

Bước 5 - Tính toán và sắp xếp lại (sinh sản): Cuối cùng, các giá trị hàm mục tiêu được tính lại theo từng chuỗi nhị phân thu được. Các kết quả tương ứng được đưa vào thế hệ hiện tại để kiểm tra xem rằng thuật giải di truyền kết thúc hay chưa (có thỏa mãn điều kiện hội tụ hay không). Nếu chưa kết thúc, các chuỗi nhị phân và giá trị hàm mục tiêu tương ứng này được chuyển đến bước 1: quá trình chọn lọc của thế hệ tiếp theo.

❖ Kết thúc thuật giải

Thuật giải di truyền (GA) là một phương pháp tìm kiếm ngẫu nhiên nên không có tiêu chí cụ thể để xác định đủ điều kiện kết thúc thuật giải. Có thể thấy rằng ban đầu, GA tiến triển rất nhanh với các giải pháp tốt hơn sau mỗi vài lần lặp lại, nhưng

càng về sau thì điều này có xu hướng bão hòa trong các giai đoạn sau khi các cải tiến là rất nhỏ. Do đó, cần phải xem xét, tính toán và cân nhắc điều kiện hội tụ dựa trên tính phù hợp và mức độ sai số cho phép của bài toán. Thông thường thì điều kiện dừng thuật giải di truyền là:

- Khi giá trị hàm mục tiêu đã đạt đến một giá trị cho trước.
- Khi không có sự cải thiện về giá trị thích nghi của các cá thể trong quần thể sau mỗi lần lặp (tạo ra thế hệ mới).
- Khi số lượng thế hệ đã đạt đến một hệ số cho trước.

Tuy nhiên, khi áp dụng vào từng bài toán cụ thể thì cần phải thử các điều kiện dừng thuật giải khác nhau để xem điều kiện nào phù hợp nhất với yêu cầu bài toán cần giải.

1.4.3.4. Các bài toán tối ưu lưu trữ và truyền video được giải bằng GA

Với các bài toán tối ưu lưu trữ và chia sẻ tài nguyên cho các ứng dụng truyền video trong 5G UDN, GA thường được chọn lựa như là một giải pháp hiệu quả để tính được kết quả tối ưu chính xác hoặc xấp xỉ với độ phức tạp về thời gian và dung lượng tính toán thấp (hoặc có thể được quy định linh hoạt bởi người dùng) [63-68]. Cụ thể, Wei Xiang và cộng sự đã đề xuất một cơ chế mới để truyền video có khả năng phục hồi lỗi trên các mạng kém tin cậy [63]. Cơ chế đề xuất đóng gói luồng bit có thể mở rộng 2 chiều và phân bổ tối ưu các luồng bit phụ vào hai đường truyền với tỷ lệ tổn thất không bằng nhau. Bài toán tối ưu hóa mối quan hệ giữa méo (distortion) và tốc độ mã hóa (rate) 2 chiều được xây dựng để giảm thiểu độ méo dự kiến và được thực hiện bằng GA. Ngoài ra, tác giả cũng chỉ ra rằng GA là một phương pháp nhanh để giải bài toán tối ưu và GA có độ phức tạp tính toán thấp so với phương pháp nhân tử Lagrangian thông thường.

Trong [64], tác giả Xu Du và cộng sự đã giải quyết vấn đề tối ưu hóa ràng buộc phi tuyến cho cả mật độ lưu trữ tối ưu và phân bố tỷ lệ tốc độ mã hóa cho tất cả các đặc tả của từng phân đoạn video bằng cách áp dụng GA, nhằm cung cấp cho

người dùng dịch vụ truyền video với chất lượng trải nghiệm cao trong khi sử dụng hiệu quả tài nguyên lưu trữ. Với [65], Sunday Oladayo Olandejo và cộng sự đã áp dụng GA để giải bài toán tối ưu cơ chế phân bổ tài nguyên nhận biết độ trễ thông minh đa tầng trong môi trường không đồng nhất của mạng 5G. Nguyen-Son Vo và cộng sự trong [66] đã dùng GA để tìm kiếm trong không gian nhị phân nhằm cực tiểu độ méo của video thu được. Từ đó, đảm bảo chất lượng trải nghiệm của người dùng (QoE – Quality of Experience) sao cho các đoạn video đến đúng thời điểm và đảm bảo phát lại liên tục; tỷ số tín hiệu cực đại trên nhiễu (PSNR – Peak Signal-to-Noise Ratio) cao cho chất lượng phát lại trung bình cao và dao động chất lượng thấp. Các tác giả trong [67] áp dụng GA cho bài toán tối ưu tài nguyên định hướng QoE để đạt hiệu suất cao cho các ứng dụng và dịch vụ video hợp tác trong mạng 5G hai tầng. Trong [68] tác giả De-Thu Huynh và cộng sự đã dùng thuật giải GA thích nghi có xem xét đến mối quan hệ xã hội và mối quan hệ vật lý giữa tất cả người dùng di động để giải bài toán tối ưu hóa hiệu quả năng lượng và chia sẻ tài nguyên phổ tần trong mạng 5G.

1.5. Tổng kết chương 1

Chương 1 đã giới thiệu tổng quan về 5G UDN với những ưu điểm nổi bật cũng như những thách thức cần nghiên cứu, các cơ chế lưu trữ và chia sẻ tài nguyên trong 5G UDN cũng được trình bày một cách tổng quát. Ngoài ra, các ưu điểm và nhược điểm của các mô hình, cơ chế lưu trữ và chia sẻ trong các công trình nghiên cứu liên quan cũng được phân tích và đánh giá chi tiết. Các phân tích và đánh giá này cho thấy:

- 1) Các công trình trong nước chủ yếu tập trung nghiên cứu tối ưu lưu trữ hoặc tối ưu chi phí cho nhà cung cấp dịch vụ hoặc nhà phát triển nội dung và chỉ thực hiện trên đơn tầng ở trạm nền chính (MBS) hoặc tại trạm nền nhỏ (SBS), nhưng chưa có các nghiên cứu định hướng đến lợi ích người dùng (MU) và đa tầng trong mạng 5G UDN;

- 2) Các công trình nghiên cứu ngoài nước đã thực hiện trên đa tầng (MBS, FBS, MU), nhưng chưa nghiên cứu đồng thời tối ưu lưu trữ đa tầng kết hợp chia sẻ tài nguyên có quan tâm đến mối quan hệ xã hội của người dùng để nâng cao hiệu quả của hệ thống.

Từ những vấn đề còn tồn đọng trong các nghiên cứu trong và ngoài nước như trên, luận án đề xuất mô hình lưu trữ đa tầng kết hợp với chia sẻ tài nguyên tối ưu theo định hướng người dùng nhằm nâng cao hiệu quả truyền video trong 5G UDN. Liên quan đến định hướng người dùng, luận án xem xét đến mối quan hệ xã hội và nhu cầu của người dùng. Luận án cũng xem xét đến độ phân giải của các video nhằm đồng thời sử dụng tài nguyên lưu trữ hiệu quả và cung cấp các video có độ phân giải phù hợp với thiết bị người dùng. Ngoài ra, Chương 1 cũng giới thiệu các thuật giải như toán vét cạn, thuật giải di truyền và thuật giải đàn dơi, là những thuật giải sẽ được triển khai, so sánh và áp dụng để giải các bài toán tối ưu cho các mô hình lưu trữ được đề xuất trong Chương 2 và Chương 3.

CHƯƠNG 2. LƯU TRỮ VÀ CHIA SẺ TÀI NGUYÊN TỐI ƯU THEO MỐI QUAN HỆ XÃ HỘI NGƯỜI DÙNG NÂNG CAO DUNG LƯỢNG TRUYỀN VIDEO TRONG 5G UDN

***Tóm tắt:** Trong Chương 2, bài toán lưu trữ và chia sẻ tài nguyên tối ưu theo mối quan hệ xã hội người dùng (SCS – Social-aware Caching and Resource Sharing) được xây dựng và giải để chỉ định FBS (Femtocell Base Station) nào lưu trữ video nào và chỉ định SU (Spectrum owner User) nào chia sẻ tài nguyên phổ tần kênh truyền xuống cho cặp truyền thông D2D (TX-RX) nào, sao cho dung lượng phân phối trung bình đến các MU (Mobile User) là cực đại. Nhằm nâng cao hiệu quả của cơ chế SCS, bài toán tối ưu có xét đến mối quan hệ xã hội của từng cặp TX-RX và có xét đến ràng buộc về dung lượng lưu trữ tại các FBS. Hơn nữa, do SU chia sẻ tài nguyên phổ tần kênh truyền xuống của nó cho truyền thông D2D từ TX (Transmisster) đến RX (Receiver), việc TX dùng lại phổ tần kênh truyền xuống này sẽ gây can nhiễu lên SU. Vì vậy, bài toán SCS cũng xem xét đến ràng buộc về giới hạn dưới của tỉ số tín hiệu trên can nhiễu và nhiễu trắng (SINR – Signal to Interference plus Noise Ratio) tại SU để đảm bảo chất lượng cho SU. Những ưu điểm của cơ chế SCS được kiểm chứng thông qua các kết quả mô phỏng và so sánh với các cơ chế thông thường khác. Các công trình công bố liên quan đến Chương 2 gồm [C1] trong phần “Danh mục các công trình đã công bố của nghiên cứu sinh” ở cuối luận án.*

2.1. Giới thiệu cơ chế SCS

Trong mạng 5G, số lượng người dùng di động ngày càng tăng nhanh đòi hỏi lượng kết nối truy cập dữ liệu ngày càng lớn, song song đó là yêu cầu về chất lượng ngày càng cao, đặc biệt là chất lượng của các dịch vụ có tốc độ và dung lượng dữ liệu cao như các dịch vụ truyền video. Điều này sẽ gây tắc nghẽn nghiêm trọng tại các tuyến trục của các trạm cơ sở lớn (MBS – Macro Base Station) trong mạng 5G. Giải pháp nâng cấp các tuyến trục của các MBS cũng như của các trạm cơ sở nhỏ (SCS – Small cell Base Station) sẽ hạn chế tắc nghẽn, tuy nhiên, sẽ rất tốn kém vì phải đầu tư trang thiết bị hạ tầng và thay đổi kiến trúc/thiết kế hệ thống. Do đó, như đã trình

bày ở Chương 1, cơ chế lưu trữ và chia sẻ tài nguyên là một giải pháp thay thế linh động hơn, không làm thay đổi kiến trúc/thiết kế hệ thống, chi phí thấp và hiệu quả dễ hạn chế tắc nghẽn. Giải pháp này sẽ mang nội dung video đến gần với người dùng đi động (MU – Mobile User) hơn bằng sự kết hợp giữa 1) lưu trữ nội dung video đa tầng tại MBS, tại SCS và tại các thiết bị MU và 2) chia sẻ tài nguyên (nội dung video và phổ tần) giữa các MU thông qua truyền thông từ thiết bị đến thiết bị (D2D – Device-to-Device).

Các kỹ thuật lưu trữ thông thường từ đơn tầng [15, 17, 20, 26, 27, 32-34, 37, 39-43, 45, 46] đến đa tầng [16, 18, 24, 25, 28, 36, 47, 48] đều được triển khai với mục đích là xác định vị trí lưu trữ tối ưu cho nội dung video để phục vụ MU tốt nhất như dung lượng phân phối cao, trì hoãn thấp, tỷ lệ truy cập nội dung thành công cao,... Tuy nhiên, những kỹ thuật lưu trữ này không xem xét đến các yếu tố phân phối video liên quan đến môi trường truyền thông không dây (tài nguyên phổ tần và đặc tính kênh truyền), các khía cạnh người dùng (mối quan hệ xã hội), thông tin của video (độ nổi tiếng, kích thước video). Các yếu tố này nếu được quan tâm sẽ nâng cao hiệu suất của kỹ thuật lưu trữ.

Liên quan đến kỹ thuật chia sẻ tài nguyên giữa các MU thông qua truyền thông D2D, truyền thông D2D đã được phát triển như một thành phần mới nội để nâng cao hiệu quả sử dụng tài nguyên, vùng phủ sóng và chất lượng dịch vụ (QoS – Quality of Services) trong mạng 5G [69]. Có thể kể đến một số ít như: truyền thông D2D được khai thác để đạt được dung lượng hệ thống và hiệu suất năng lượng cao nhất bằng cách chia sẻ và sử dụng lại phổ tần của MU khác, cụ thể là của người dùng chấp nhận chia sẻ tài nguyên phổ tần kênh truyền xuống (SU – Spectrum owning User) [70, 71]. Hơn nữa, xét đến độ trễ hàng đợi, một cơ chế tối ưu tài nguyên có nhận biết can nhiễu (interference-aware³) đã được đề xuất để tối đa hiệu suất sử dụng phổ tần và năng lượng trong khi đảm bảo sự ổn định của mạng và tỷ số tín hiệu trên can nhiễu và

³ Nhận biết can nhiễu là một thuật ngữ được dùng cho các giải pháp triệt can nhiễu trong truyền thông tin vô tuyến. Nhận biết can nhiễu càng tốt thì việc triệt can nhiễu sẽ tốt, dẫn đến công suất của tín hiệu can nhiễu tại đầu thu càng thấp, và do đó nâng cao chất lượng tín hiệu thu được.

hiệu trắng (SINR – Signal to Interference plus Noise Ratio⁴). Tỷ số này càng cao thì chất lượng tín hiệu mong muốn tại đầu thu càng tốt) của người dùng chia sẻ tài nguyên (SU) [8]. Không sử dụng lại phổ tần được chia sẻ bởi các SU trong cùng một tế bào mạng (cell) như trong [8, 70, 71], việc chia sẻ tài nguyên cho truyền thông D2D trong đa tế bào mạng đã được nghiên cứu để phục vụ các MU di chuyển giữa các tế bào mạng với QoS cao nhưng độ phức tạp thấp [72]. Ngoài ra, truyền thông D2D cũng được kết hợp với kỹ thuật lưu trữ tại các MU để đạt được hiệu quả hệ thống cao hơn như được nghiên cứu trong [15, 33, 42, 73]. Tuy nhiên, các kỹ thuật kết hợp này vẫn chưa tận dụng được tài nguyên dung lượng lưu trữ tại các MBS và SBS song song với lưu trữ tại các MU cũng như chưa tận dụng khai thác các thông tin MU và thông tin nội dung video để đạt được hiệu quả lưu trữ và phân phối nội dung cao nhất.

Có thể nhận thấy rằng việc tận dụng thông tin MU và thông tin nội dung video lưu trữ cùng với các thông tin liên kết vật lý (đặc tính kênh truyền không dây) sẽ cải thiện QoS của hệ thống. Ví dụ, khi dự đoán rằng một video sẽ có tần suất truy cập cao (độ nổi tiếng cao) và kích thước video lớn (truyền với thời gian dài), thì trong số các MU đã lưu video này, MU nào có mối quan hệ xã hội và kênh truyền tốt với MU gọi yêu cầu video sẽ được chỉ định để gửi nội dung qua truyền thông D2D. Trong chương này, trong khi các thông tin về video và thông tin về đặc tính kênh truyền không dây có thể thu thập được hoặc đã có mô hình cụ thể, mô hình mối quan hệ xã hội được áp dụng sẽ là mô hình Buffet Ẩn Độ (IBM) [74, 75]. Mô hình IBM cung cấp các yếu tố xã hội cho phép các MU lân cận có cùng sở thích với số lần liên lạc và thời gian liên lạc đủ lớn sẽ được chọn lựa để thiết lập mạng truyền thông D2D. Sau đó, tần suất truy cập video và khoảng cách của các cặp truyền thông D2D được tận dụng để đề xuất các kỹ thuật truyền thông D2D kết hợp với lưu trữ theo mối quan hệ xã hội nhằm phục vụ MU với QoS cao [46, 76, 77].

⁴ SINR là tỷ số giữa công suất tín hiệu mong muốn thu được và công suất của các tín hiệu nhiễu gồm nhiễu trắng trên toàn miền tần số và can nhiễu.

Từ những phân tích trên, trong Chương 2, cơ chế lưu trữ và chia sẻ tài nguyên tối ưu theo mối quan hệ xã hội người dùng (SCS – Social-aware Caching and Resource Sharing) được đề xuất để phân phối video trong 5G UDN. Cơ chế SCS về cơ bản sẽ giúp giảm tải cho mạng lõi và cho các tuyến trực của cả MBS và SBS (cụ thể là FBS⁵) nhờ vào sự trợ giúp của truyền thông D2D từ thiết bị phát (TX – Transmitter) có lưu trữ video đến thiết bị thu (RX – Receiver) có yêu cầu truy xuất video. Quan trọng hơn, cơ chế SCS cung cấp cho MU khả năng truy cập video linh động với dung lượng phân phối cao nhờ vào việc tận dụng: 1) dung lượng lưu trữ sẵn có của MBS, FBS và MU; 2) tài nguyên phổ tần kênh truyền xuống của SU để chia sẻ cho truyền thông D2D; và 3) mối quan hệ xã hội của các cặp người dùng D2D (TX-RX) và thông tin nội dung video. Vấn đề được đặt ra là làm thế nào để lưu trữ các video vào các FBS, đồng thời, làm thế nào để SU chia sẻ tài nguyên phổ tần kênh truyền xuống của nó cho truyền thông D2D một cách tốt nhất với dung lượng phân phối video là cực đại.

Vấn đề trên được giải quyết bằng cách xây dựng bài toán tối ưu SCS dựa vào mối quan hệ xã hội người dùng. Bài toán tối ưu SCS được xây dựng dựa trên việc tính toán các thông số của mạng liên quan đến đặc tính kênh truyền không dây, dung lượng lưu trữ, tài nguyên phổ tần và các thông tin của video (độ nổi tiếng và độ dài). Liên quan đến lưu trữ, giả sử rằng tất cả các video đều được lưu ở MBS nhưng một video được lưu vào TX hay không sẽ phụ thuộc vào độ nổi tiếng của video và dung lượng còn trống của TX. Do đó, giải bài toán tối ưu SCS sẽ xác định số lượng bản sao tối ưu của mỗi video và vị trí lưu trữ tối ưu của chúng trong các FBS. Liên quan đến việc chia sẻ tài nguyên phổ tần kênh truyền xuống, giải bài toán tối ưu SCS sẽ xác định SU nào chia sẻ tài nguyên kênh truyền xuống của nó với cặp TX-RX nào cho truyền thông video D2D từ TX đến RX. Để việc chia sẻ tài nguyên kênh truyền xuống được hiệu quả, bài toán tối ưu SCS có xét đến 1) mối quan hệ xã hội của các cặp TX-RX và 2) ngưỡng dưới tỷ số tín hiệu trên can nhiễu và nhiễu trắng (SINR) tại

⁵ Trong luận án này, trạm cơ sở nhỏ (SBS) được sử dụng là trạm cơ sở nhỏ femto (FBS – Femtocell Base Station) do FBS thường đặt ở gần MU hơn so với các loại SBS khác.

SU để đảm bảo chất lượng (đảm bảo sự công bằng) cho SU vì khi TX dùng lại phổ tần được chia sẻ bởi SU, tín hiệu phát từ TX sẽ gây can nhiễu lên SU. Cuối cùng, từ kết quả mô phỏng có được, các kiểm chứng và phân tích sẽ được trình bày để đánh giá hiệu suất của cơ chế SCS so với các cơ chế khác.

2.2. Mô hình truyền video trong 5G UDN với cơ chế SCS

2.2.1. Các ký hiệu và giả thiết

Bảng 2-1 trình bày các ký hiệu được sử dụng trong Chương 2. Mô hình truyền video trong 5G UDN với cơ chế SCS được xây dựng dựa trên các giả thiết sau:

- MBS thu thập được tất cả các thông số hệ thống bao gồm dung lượng lưu trữ của FBS và TX, đặc tính của các kênh truyền không dây, mối quan hệ xã hội của các cặp TX-RX, thông tin về video v.v., để xây dựng và giải bài toán tối ưu SCS.
- Các thông số hệ thống không thay đổi trong quá trình xây dựng và giải bài toán tối ưu SCS cũng như truyền video. Một số ứng dụng hiệu quả của cơ chế SCS đó là truyền video tại những khu vực có lượng lớn MU như sân vận động, buổi hòa nhạc hoặc phòng họp, khuôn viên và tòa nhà văn phòng, khuôn viên trường v.v.
- Nhờ dung lượng lưu trữ lớn, MBS có khả năng lưu tất cả các video. Trong khi đó, TX của các cặp D2D sẽ quyết định lưu một video hay không tùy thuộc vào dung lượng lưu trữ còn lại và độ nổi tiếng của video.
- Một SU có thể chia sẻ tài nguyên phổ tần của nó cho nhiều hơn một cặp TX-RX miễn là ngưỡng tỷ số tín hiệu trên can nhiễu và nhiễu trắng (SINR) của SU được đảm bảo.
- Các vấn đề liên quan đến tần số phát sóng, hiệu năng của ăng-ten, môi trường truyền dẫn v.v. không được xem xét nhằm đơn giản hóa mô hình hệ thống và các tính toán liên quan. Ngoài ra, can nhiễu từ MBS đến các FBS và giữa các

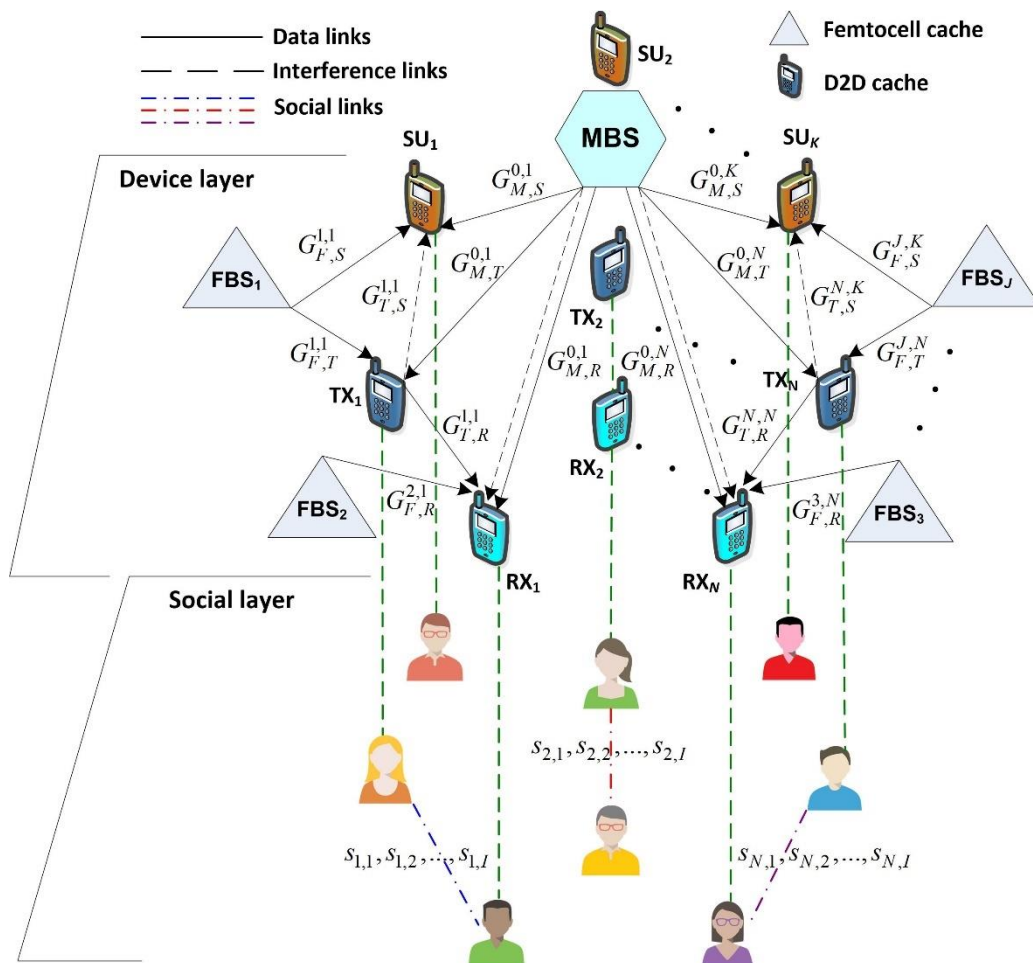
FBS được triệt tiêu bằng phương pháp đa truy cập F-ALOHA [78, 79].

Bảng 2-1. Các ký hiệu sử dụng cho bài toán SCS

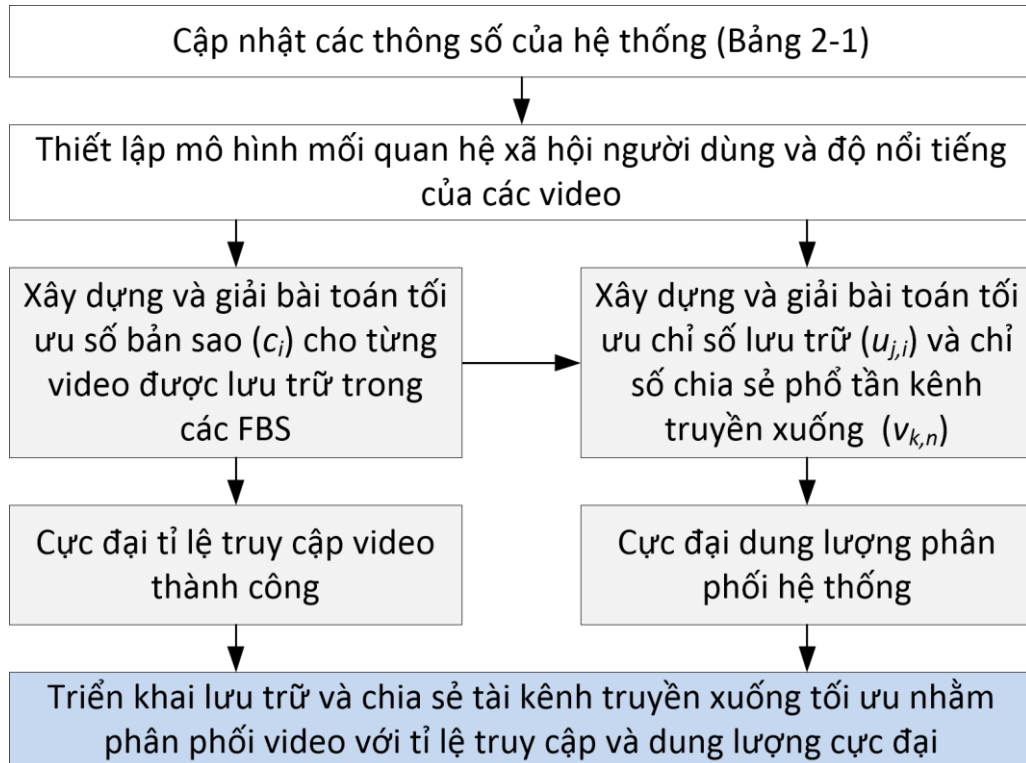
Ký hiệu	Diễn giải
I	Số video
J	Số trạm cơ sở nhỏ femto (FBS)
K	Số người dùng di động chấp nhận chia sẻ tài nguyên phổ tần kênh truyền xuống (SU)
N	Số cặp truyền thông D2D, mỗi cặp gồm một người dùng di động lưu trữ video (TX – Transmitter) và một người dùng nhận video (RX – Receiver).
T_{min}^i	Thời gian tối thiểu để truyền video thứ i , $i = 1, 2, \dots, I$ (tùy thuộc vào độ dài của mỗi video)
$s_{n,i}$	Xác suất dựa trên mối quan hệ xã hội để xem xét khả năng cặp TX-RX thứ n có truyền thành công video thứ i hay không
$u_{j,i}$	Chỉ số lưu trữ video tại FBS, $u_{j,i} = 1$ nghĩa là FBS thứ j quyết định lưu video thứ i và ngược lại.
$v_{k,n}$	Chỉ số chia sẻ tài nguyên phổ tần kênh truyền xuống: $v_{k,n} = 1$ nghĩa là SU thứ k đồng ý chia sẻ phổ tần kênh truyền xuống của nó cho cặp D2D thứ n và ngược lại.
P_M^0	Công suất phát của trạm cơ sở lớn (MBS)
P_F^j	Công suất phát của trạm cơ sở nhỏ femto (FBS) thứ j
P_T^n	Công suất phát của TX thứ n trong cặp truyền thông D2D thứ n .
$G_{X,Y}^{x,y}$	Độ lợi kênh giữa X và Y , trong đó x và y là các chỉ số của X và Y
N_0	Công suất nhiễu trắng Gaussian (AWGN – Additive White Gaussian Noise)
W	Băng thông hệ thống
r_i	Tỷ lệ truy xuất (hay độ nổi tiếng) của video thứ i
β_n	Phần trăm dung lượng lưu trữ còn trống của TX thứ n
$p_{n,i}$	Xác suất để TX thứ n quyết định lưu video thứ i
γ_0	Giới hạn dưới tỉ số tín hiệu trên can nhiễu và nhiễu trắng (SINR) tại SU
R_S	Dung lượng phân phối từ MBS và FBS đến SU
R_T	Dung lượng phân phối từ MBS và FBS đến TX

2.2.2. Mô hình hệ thống với cơ chế SCS

Trong chương này, mô hình truyền video trong 5G UDN với cơ chế SCS được mô tả như Hình 2-1. Mô hình được phân ra thành lớp thiết bị và lớp mối quan hệ xã hội. Lớp thiết bị gồm 1 MBS, J FBS, $(K + 2N)$ MU gồm K SU và N cặp TX-RX, và I video. Lớp xã hội mô tả mối quan hệ xã hội của các cặp TX-RX trong truyền thông D2D. Trong mô hình này, khi MBS dự đoán được sẽ có lượng lớn MU yêu cầu truy xuất các video, nó sẽ thực thi cơ chế SCS như được mô tả trong Hình 2-2, cụ thể gồm 3 bước như sau:



Hình 2-1. Mô hình hệ thống của SCS



Hình 2-2. Lưu đồ hoạt động cơ chế SCS

Bước 1 – Cập nhật các thông số: Trong bước này, MBS sẽ cập nhật các thông số của hệ thống như số FBS (J) để lưu trữ video, số SU (K) để chia sẻ tài nguyên phổ tần kênh truyền xuống, số TX-RX (N) cho truyền thông D2D, số video (I), mối quan hệ xã hội của các cặp TX-RX, bảng thông hệ thống, đặc tính kênh truyền và các thông số khác như được liệt kê trong Bảng 2-1.

Bước 2 – Xây dựng và giải bài toán tối ưu SCS: Dựa trên các thông số ở Bước 1, MBS sẽ xây dựng bài toán tối ưu SCS và giải bài toán này nhằm tìm ra số lượng bản sao lưu tối ưu của mỗi video để cực đại số bản sao lưu trữ trung bình trong hệ thống (tăng tỷ lệ truy suất thành công video); các chỉ số lưu trữ tối ưu $u_{j,i}$, với $j = 1, 2, \dots, J$ và $i = 1, 2, \dots, I$, và các chỉ số chia sẻ tài nguyên phổ tần kênh truyền xuống tối ưu $v_{k,n}$, $k = 1, 2, \dots, K$ và $n = 1, 2, \dots, N$; nhằm cực đại dung lượng phân phối của hệ thống. Trong đó:

$u_{j,i} = 1$ nghĩa là FBS thứ j quyết định lưu trữ video thứ i , ngược lại thì $u_{j,i} = 0$

và

$v_{k,n} = 1$ nghĩa là SU thứ k quyết định chia sẻ tài nguyên kênh truyền xuống của nó cho cặp TX-RX thứ n , ngược lại thì $v_{k,n} = 0$

Bước 3 – Triển khai SCS: Cuối cùng, dựa vào kết quả tối ưu sau khi giải bài toán SCS, MBS sẽ cộng tác với các FBS và các TX để phân phối các video được lưu trữ đến các SU và TX (phân phối bởi MBS và FBS) cũng như đến các RX (phân phối bởi TX) với dung lượng cực đại.

2.3. Tính toán các thông số hệ thống với cơ chế SCS

Từ hệ thống với mục tiêu nêu trên, ta có thể nhận thấy các thông số cần được mô hình và tính toán cho bài toán tối ưu SCS bao gồm: mối quan hệ xã hội của các cặp TX-RX và tỷ số tín hiệu trên nhiễu (SNR - Signal-to-Noise Ratio) hoặc tỷ số tín hiệu trên can nhiễu và nhiễu trắng (SINR) của các kênh truyền không dây để làm cơ sở tính toán dung lượng kênh truyền. Từ đó, ta có thể tính được dung lượng phân phối trung bình của hệ thống – là hàm mục tiêu của bài toán tối ưu SCS.

2.3.1. Mối quan hệ xã hội giữa các cặp TX-RX

Chắc chắn rằng mối quan hệ xã hội của các cặp TX-RX sẽ ảnh hưởng đến hiệu suất chia sẻ tài nguyên phổ tần cho truyền thông D2D. Nói cách khác, các SU quyết định chia sẻ tài nguyên phổ tần kênh truyền xuống của nó hay không, phụ thuộc không chỉ vào công suất phát của TX, độ lợi kênh truyền giữa TX và RX, độ lợi kênh truyền giữa TX và SU để xác định mức ảnh hưởng của can nhiễu từ TX lên SU, xác suất các video được lưu tại TX, mà còn phụ thuộc vào mối quan hệ xã hội giữa TX và RX.

Việc xác định mối quan hệ xã hội giữa các người dùng di động là rất khó vì nó phụ thuộc rất nhiều yếu tố như: khoảng cách địa lý, vùng miền, thời lượng kết nối, số lần kết nối thông qua các mạng truyền thông tin và mạng xã hội, v.v. Do đó, trong

giới hạn của luận án, để đơn giản cho việc áp dụng mối quan hệ xã hội của người dùng cũng như sự phù hợp của nó vào bài toán lưu trữ và truyền video trong mạng thông tin di động, mối quan hệ xã hội theo mô hình IBM (Indian Buffet Model) [46, 74, 75, 80-82] dựa trên phân bố Gamma được áp dụng. Mô hình IBM cho phép chọn lựa các cặp TX-RX phù hợp để thực hiện truyền thông D2D bằng cách sử dụng lại tài nguyên phổ tần kênh truyền xuống được chia sẻ bởi SU. Trong mô hình IBM, đối với cặp D2D thứ n , mối quan hệ xã hội giữa TX thứ n và RX thứ n được tính toán dựa trên lịch sử về thời lượng và số lần kết nối của chúng.

Trong truyền thông D2D có quan tâm đến mối quan hệ xã hội, mối quan hệ xã hội giữa TX và RX của cặp D2D thứ n được dùng để tính xác suất truyền thành công video thứ i trong thời gian T_{min}^i , nghĩa là các cặp D2D có mối quan hệ xã hội cao hơn thì sẽ có xác suất truyền video thành công cao hơn. Để tính toán xác suất này, ta đặt Z_n là số lần kết nối của cặp TX-RX thứ n và Z_m là thời gian của mỗi lần kết nối, $m = 1, 2, \dots, Z_n$, thời gian kết nối trung bình M_n của cặp TX-RX thứ n được tính như sau:

$$M_n = \frac{\sum_{m=1}^{Z_n} Z_m}{Z_n} \quad (2.1)$$

Ngoài ra, để có được phân bố về thời gian kết nối, ta cần tìm phương sai biểu thị độ biến động thời gian kết nối giữa các cặp TX-RX khác nhau. Nếu hai cặp TX-RX có cùng thời gian kết nối trung bình thì cặp nào có phương sai thấp hơn sẽ thích hợp hơn cho việc truyền video. Phương sai của thời gian kết nối V_n của cặp TX-RX thứ n là

$$V_n = \frac{\sum_{m=1}^{Z_n} (Z_m - M_n)^2}{Z_n} \quad (2.2)$$

với M_n và V_n từ công thức (2.1) và (2.2), ta có phân bố Gamma cho mô hình phân bố thời gian kết nối như sau:

$$X \sim \Gamma(\kappa_n, \theta_n) = \Gamma\left(\frac{M_n^2}{V_n}, \frac{V_n}{M_n}\right) \quad (2.3)$$

và hàm mật độ xác suất (probability density function – PDF) của thời gian kết nối được xác định bởi:

$$f(x; \kappa_n, \theta_n) = \frac{1}{\theta_n^{\kappa_n}} \frac{1}{\Gamma(\kappa_n)} x^{\kappa_n-1} e^{-\frac{x}{\theta_n}} \quad (2.4)$$

trong đó $\Gamma(\kappa_n) = \int_0^{\infty} t^{\kappa_n-1} e^{-t} dt$

Do đó, xác suất truyền thành công video thứ i từ TX đến RX của cặp D2D thứ n trong thời gian T_{min}^i thì được tính như sau [81]:

$$s_{n,i} = 1 - \int_0^{\delta T_{min}^i} f(u; \kappa_n, \theta_n) du = 1 - \frac{\gamma\left(\kappa_n, \frac{\delta T_{min}^i}{\theta_n}\right)}{\Gamma(\kappa_n)} \quad (2.5)$$

với $\gamma\left(\kappa_n, \frac{\delta T_{min}^i}{\theta_n}\right) = \int_0^{\frac{\delta T_{min}^i}{\theta_n}} t^{\kappa_n-1} e^{-t} dt$ và $\delta \geq 1$ được thêm vào để linh hoạt điều chỉnh độ dài tối thiểu của video cho phù hợp với mối quan hệ xã hội của tất cả các cặp TX-RX nhằm đạt hiệu suất hệ thống cao nhất.

2.3.2. Mô hình kênh truyền không dây

Trong hệ thống này, sự hiện diện đồng thời của cả MSB và FSB có thể gây ra can nhiễu giữa các tầng và can nhiễu đồng tầng (cross-tier and co-tier interferences). Để thuận tiện trong mô hình kênh truyền không dây, cơ chế phân kênh và giao thức F-ALOHA [78, 79] được áp dụng cho truyền thông từ FBS đến MU để kiểm soát can nhiễu. Việc áp dụng F-ALOHA sẽ dẫn đến sự phức tạp cho hệ thống. Tuy nhiên, giả thiết này là hợp lý và được áp dụng cũng như phát triển trong các nghiên cứu gần đây [83-87] khi cấu hình, năng lực xử lý và tính toán của các thiết bị mạng và thiết bị đầu

cuối được cải tiến không ngừng và nhanh chóng để đáp ứng nhu cầu ngày càng cao của người dùng. Đây được xem như là một giả thiết để không làm phức tạp bài toán quản lý và phân bổ phổ tần của toàn hệ thống mà chỉ tập trung vào cơ chế chia sẻ phổ tần cho truyền thông D2D có quan tâm đến can nhiễu giữa thiết bị phát của các cặp D2D lên thiết bị chia sẻ phổ tần, can nhiễu giữa MBS và thiết bị thu của các cặp D2D, và can nhiễu giữa thiết bị phát của các cặp D2D và thiết bị thu của các cặp D2D dùng chung phổ tần được chia sẻ.

Như vậy, vẫn tồn tại can nhiễu khi SU chia sẻ tài nguyên phổ tần kênh truyền xuống của nó với các cặp TX-RX cho truyền thông D2D. Trong quá trình chia sẻ và dùng chung phổ tần, tín hiệu được truyền từ MBS đến SU có thể gây can nhiễu lên RX và tín hiệu truyền từ TX đến RX sẽ gây can nhiễu lên SU. Chất lượng của tín hiệu thu được tại đầu thu sẽ phụ thuộc vào tỷ số giữa công suất tín hiệu mong muốn thu được và tổng công suất của các tín hiệu không mong muốn (can nhiễu và nhiễu trắng). Đại lượng quan trọng trong tỷ số này đó là độ lợi kênh truyền $G_{X,Y}^{x,y}$ từ đầu phát đến đầu thu, trong đó:

$X \in \{M, F, T\}$ là viết tắt của $\{MBS, FBS, TX\}$

$Y \in \{S, T, R\}$ là viết tắt của $\{SU, TX, RX\}$.

$x \in \{j, n\}, j = 1, 2, \dots, J; n = 1, 2, \dots, N; j = 0$ chỉ định MBS,

$y \in \{k, n\}, k = 1, 2, \dots, K$

và theo [79], ta có

$$G_{X,Y}^{x,y} = h_{X,Y}^{x,y} g_{X,Y}^{x,y} \quad (2.6)$$

trong đó $h_{X,Y}^{x,y}$ là hệ số fading phân phối mũ có trị trung bình là 1 ($\sim \exp(1)$)

và $g_{X,Y}^{x,y} = \|d\|^{-\eta}$ với η là hệ số suy hao công suất theo hệ số mũ và d là khoảng cách giữa X và Y.

2.3.3. Dung lượng phân phối hệ thống

Dung lượng phân phối hệ thống (system delivery capacity) được xác định bằng tổng dung lượng truyền từ MBS, FBS và TX đến MU (gồm SU, TX và RX). Dung lượng phân phối của hệ thống được tính thông qua tỷ số tín hiệu trên nhiễu trắng (SNR) hoặc tỷ số tín hiệu trên can nhiễu và nhiễu trắng (SINR) của các kênh từ MBS, FBS và TX đến MU, cụ thể được trình bày như sau.

2.3.3.1. Dung lượng phân phối đến SU

Trong mô hình hệ thống với cơ chế SCS, SU thứ k có thể nhận video thứ i được lưu trữ tại FBS thứ j nếu chỉ số lưu trữ video $u_{j,i} = 1$. Do đó, SNR tại SU thứ k từ FBS thứ j được thể hiện như sau:

$$\gamma_{F,S}^{j,k,i} = \frac{u_{j,i} P_F^j G_{F,C}^{j,k}}{N_0} \quad (2.7)$$

trong đó:

P_F^j là công suất truyền của FBS j

$G_{F,S}^{j,k}$ là độ lợi kênh truyền giữa FBS j và SU k

N_0 là công suất nhiễu trắng Gaussian (AWGN)

Ngược lại, nếu tất cả các FBS không lưu video thứ i , nghĩa là $u_{j,i} = 0, \forall j = 1, 2, \dots, J$, thì SU thứ k sẽ nhận video thứ i từ MBS như thông thường. Trong trường hợp này, bởi vì SU thứ k chia sẻ tài nguyên phổ tần của nó với cặp TX-RX thứ n để truyền video thứ i qua truyền thông D2D. Nghĩa là khi $v_{k,n} = 1$, SINR tại SU thứ k sẽ bị suy giảm do ảnh hưởng can nhiễu bởi công suất phát P_T^n của TX thứ n . Hơn nữa, khi xem xét thêm mối quan hệ xã hội của cặp TX-RX thứ n và khả năng lưu trữ của TX thứ n , can nhiễu này sẽ phụ thuộc vào 1) xác suất $s_{n,i}$ mà video thứ i được truyền thành công từ TX n đến RX n dựa trên mối quan hệ xã hội của cặp TX-RX thứ n và 2) xác suất $p_{n,i}$ mà video thứ i được lưu tại TX thứ n . Vì vậy, SINR tại SU thứ k từ MBS được tính như sau:

$$\gamma_{M,S}^{0,k,i} = \frac{\prod_{j=1}^J (1 - u_{j,i}) P_M^0 G_{M,S}^{0,k}}{N_0 + \sum_{n=1}^N s_{n,i} \nu_{k,n} p_{n,i} P_T^n G_{T,S}^{n,k}} \quad (2.8)$$

Trong công thức (2.8), $u_{j,i} = 0, j=1, 2, \dots, J$, nghĩa là nếu tất cả các FBS không lưu video thứ i , thì SU thứ k sẽ lấy video thứ i từ MBS, với:

P_M^0 là công suất truyền của MBS

$G_{M,S}^{0,k}$ là độ lợi kênh truyền giữa MBS và SU k

P_T^n là công suất truyền của TX

$G_{T,S}^{n,k}$ là độ lợi kênh truyền giữa TX n và SU k

$s_{n,i}$ được tính từ công thức (2.5)

$p_{n,i}$ là xác suất để TX thứ n lưu trữ video thứ i , nó phụ thuộc vào độ nổi tiếng (popularity hay access rate) của video thứ i (r_i) và phần trăm dung lượng còn trống của TX thứ n (β_n), được xác định bởi:

$$p_{n,i} = ar_i + b\beta_n \quad (2.9)$$

với $a, b \in [0,1]$, $a + b = 1$ và độ nổi tiếng của video thứ i , đặc trưng cho hành vi người dùng đối với video thứ i , được tính theo phân bố Zipf-like [88], như sau:

$$r_i = \frac{i^{-\alpha}}{\sum_{i=1}^I i^{-\alpha}} \quad (2.10)$$

với $\alpha \geq 0$ thể hiện cho độ lệch nổi tiếng giữa các video khác nhau. Nghĩa là nếu $\alpha = 0$ thì tất cả các video có cùng độ nổi tiếng là $\frac{1}{I}$, khi giá trị α cao hơn có nghĩa là độ lệch nổi tiếng giữa các video càng cao.

Mặc dù Zipf-like (2.10) được đề xuất để mô tả độ nổi tiếng của nội dung từ rất lâu (năm 1999), cho đến nay, mô hình mô tả độ nổi tiếng của nội dung chắc chắn sẽ có nhiều khác biệt. Việc mô hình hóa để đáp ứng những đặc tính khác biệt này có thể sẽ trở nên phức tạp bởi sẽ có nhiều trọng số ảnh hưởng đến độ lệch nổi tiếng giữa các

nội dung thay vì chỉ có một (đó là α) như đối với Zipf-like. Tuy nhiên, trong quá trình tìm hiểu các công trình nghiên cứu liên quan về lưu trữ, đa số các tác giả vẫn chọn lựa Zipf-like. Một lý do nữa để chọn lựa Zipf-like là nhờ vào tính đơn giản mà ta có thể dễ dàng đánh giá được phản ứng của mô hình đề xuất/hệ thống khi thay đổi trọng số ảnh hưởng đến độ nổi tiếng trong mô hình Zipf-like.

Với băng thông hệ thống là W (được tính bằng MHz) và số video là I , dựa vào định luật Shannon, ta tính được tổng dung lượng phân phối đến các SU như sau:

$$R_S = W \left\{ \sum_{i=1}^I r_i \sum_{k=1}^K \left[\log_2(1 + \gamma_{M,S}^{0,k,i}) + \sum_{j=1}^J \log_2(1 + \gamma_{F,S}^{j,k,i}) \right] \right\} \quad (2.11)$$

2.3.3.2. Dung lượng phân phối đến TX

Như đã đề cập, TX sẽ nhận được video từ MBS và FBS. Nhờ vào việc áp dụng cơ chế phân kênh và giao thức F-ALOHA [78, 79] để kiểm soát can nhiễu liên tầng và can nhiễu đồng tầng (cross-tier and co-tier interferences) giữa MBS và FBS nên các TX không bị ảnh hưởng bởi can nhiễu. Tương tự như với các SU, dung lượng phân phối từ FBS thứ j và MBS tới TX thứ n được tính thông qua SNR tại TX thứ n tương ứng như sau:

$$\gamma_{F,T}^{j,n,i} = \frac{u_{j,i} P_F^j G_{F,T}^{j,n}}{N_0} \quad (2.12)$$

và

$$\gamma_{M,T}^{0,n,i} = \frac{\prod_{j=1}^J (1 - u_{j,i}) P_M^0 G_{M,T}^{0,n}}{N_0} \quad (2.13)$$

với $G_{F,T}^{j,n}$ và $G_{M,T}^{0,n}$ là độ lợi kênh từ FBS j và MBS tới TX n

Từ (2.12) và (2.13), dung lượng phân phối từ FBS thứ j và MBS tới TX thứ n được tính như sau:

$$R_T = W \left\{ \sum_{i=1}^I r_i \sum_{n=1}^N \left[\log_2(1 + \gamma_{M,T}^{0,n,i}) + \sum_{j=1}^J \log_2(1 + \gamma_{F,T}^{j,n,i}) \right] \right\} \quad (2.14)$$

2.3.3.3. Dung lượng phân phối đến RX

Không giống như SU và TX, dung lượng phân phối đến RX không chỉ từ MBS và FBS, mà còn từ TX qua truyền thông D2D bằng cách dùng lại tài nguyên phổ tần kênh truyền xuống được chia sẻ bởi SU. Trong trường hợp này, SINR tại RX thứ n từ TX thứ n phải xét đến can nhiễu gây ra bởi MBS và của các TX thứ $l \neq n$, $l = 1, 2, \dots, N$ (do một SU có thể chia sẻ tài nguyên phổ tần kênh truyền xuống của nó cho nhiều cặp TX-RX). Xét thêm xác suất truyền video thứ i thành công dựa trên mối quan hệ xã hội của cặp TX-RX thứ n ($s_{n,i}$), chỉ số chia sẻ tài nguyên phổ tần kênh truyền xuống ($v_{k,n}$) và xác suất lưu trữ video thứ i tại TX thứ n ($p_{n,i}$), SINR này được tính như sau:

$$\gamma_{T,R}^{n,k,i} = \frac{s_{n,i} v_{k,n} p_{n,i} P_T^n G_{T,R}^{n,n}}{N_0 + P_M^0 G_{M,R}^{0,n} + \sum_{l=1, l \neq n}^N s_{l,i} v_{k,l} p_{l,i} P_T^l G_{T,R}^{l,l}} \quad (2.15)$$

với $G_{T,R}^{n,n}$ và $G_{M,R}^{0,n}$ là độ lợi kênh từ TX thứ n và từ MBS tới RX thứ n .

Nếu như các điều kiện về mối quan hệ xã hội của cặp TX-RX thứ n , SU thứ k chia sẻ tài nguyên kênh truyền cho cặp TX-RX thứ n , hoặc video thứ i lưu trữ tại TX thứ n không xảy ra, trong khi đó, video thứ i lại được lưu trữ tại FBS thứ j , thì SNR tại RX thứ n từ FBS thứ j được tính như sau:

$$\gamma_{F,R}^{j,n,k,i} = \frac{u_{j,i} (1 - s_{n,i} v_{k,n} p_{n,i}) P_F^j G_{F,R}^{j,n}}{N_0} \quad (2.16)$$

với $G_{F,R}^{j,n}$ là độ lợi kênh từ FBS thứ j tới RX thứ n .

Hơn nữa, nếu tất cả các FBS cũng không lưu video thứ i , thì SNR tại RX thứ n từ MBS được tính như sau:

$$\gamma_{M,R}^{0,n,k,i} = \frac{\prod_{j=1}^J (1 - u_{j,i})(1 - s_{n,i} v_{k,n} p_{n,i}) P_M^0 G_{M,R}^{0,n}}{N_0} \quad (2.17)$$

Như vậy, tổng dung lượng từ MBS, từ FBS và từ TX đến RX được tính như sau:

$$R_R = W \left\{ \sum_{i=1}^I r_i \sum_{n=1}^N \sum_{k=1}^K \left[\log_2(1 + \gamma_{M,R}^{0,n,k,i}) \right. \right. \\ \left. \left. + \log_2(1 + \gamma_{T,R}^{n,k,i}) + \sum_{j=1}^J \log_2(1 + \gamma_{F,R}^{j,n,k,i}) \right] \right\} \quad (2.18)$$

Cuối cùng, từ các công thức (2.11), (2.14) và (2.18), ta có dung lượng phân phối trung bình hệ thống cho mỗi MU được tính như sau:

$$R = \frac{R_S + R_T + R_R}{K + 2N} \quad (2.19)$$

2.4. Bài toán tối ưu SCS và thuật giải vét cạn

Bài toán tính tối ưu SCS sẽ bao gồm hàm mục tiêu (2.19) và 2 ràng buộc: 1) dung lượng lưu trữ của FBS và 2) ngưỡng SINR để đảm bảo chất lượng cho SU. Đầu tiên, khi xem xét khả năng lưu trữ của FBS, giới hạn tổng dung lượng lưu trữ (C^*) của tất cả các FBS (được tính bằng số lượng bản lưu trong FBS của tất cả các video) phải được quan tâm. Do vậy, bài toán đặt ra làm thế nào khai thác C^* một cách hiệu quả để tối đa số lượng bản sao trung bình được lưu trong FBS, và như vậy, sẽ tương đương với việc đạt được tỉ lệ truy cập (hit rate) video thành công cao bằng cách tìm số lượng bản sao (c_i) tối ưu của video thứ i . Bài toán này được trình bày như sau:

$$\max_{c_i} \sum_{i=1}^I r_i c_i \quad (2.20)$$

$$s. t. \begin{cases} 0 \leq c_i \leq J, i = 1, 2, \dots, I \\ \sum_{i=1}^I c_i \leq C^*, I \leq C^* \leq I \times J \end{cases} \quad (2.21)$$

Bài toán quy hoạch tuyến tính trong (2.20) và (2.21) có thể được giải bằng phương pháp điểm trong (interior point) [89, 90]. Một số lưu ý trong quá trình giải bài toán (2.20) và (2.21) đó là:

- Khi $\alpha = 0$, các giá trị r_i là như nhau cho các video. Khi đó, nghiệm được chọn là $c_i = 1, \forall i = 1, 2, \dots, I$ để đảm bảo số lượng bản sao của từng video tỷ lệ theo độ nổi tiếng của video đó.
- Về mục đích, bài toán (2.20) để tìm ra số bản sao lưu trữ tối ưu (c_i) cho từng video phụ thuộc vào độ nổi tiếng của video để cực đại số bản sao lưu trữ trung bình. Số bản sao lưu trữ trung bình càng lớn thì khả năng truy cập video thành công càng cao.
- Sau khi giải bài toán tối ưu (2.20), ta sẽ có được c_i . Giá trị c_i này được dùng làm ràng buộc về số bản sao video được lưu trữ trong các FBS cho bài toán tối ưu (2.22).

Ngoài ra, để giới hạn ngưỡng SINR (γ_0) cho các SU, ta dựa vào công thức (2.8) và đặt $\frac{P_M^0 G_{M,S}^{0,k}}{N_0 + \sum_{n=1}^N s_{n,i} v_{k,n} p_{n,i} P_T^n G_{T,S}^{n,k}} \geq \gamma_0$. Điều này có nghĩa là khi tăng γ_0 thì SINR của SU tăng để đảm bảo chất lượng cho SU.

Xem xét c_i bằng cách giải (2.20), (2.21) và xem xét γ_0 như là 2 ràng buộc của bài toán tối ưu SCS, khi đó, bài toán tối ưu SCS được trình bày như sau:

$$\max_{u_{j,i}, v_{k,n}} R \quad (2.22)$$

$$s. t. \begin{cases} \sum_{j=1}^J u_{j,i} \leq c_i, i = 1, 2, \dots, I \\ \sum_{n=1}^N s_{n,i} v_{k,n} p_{n,i} P_T^n G_{T,C}^{n,k} \leq \frac{P_M^0 G_{M,S}^{0,k}}{\gamma_0} - N_0, k = 1, 2, \dots, K, i = 1, 2, \dots, I \end{cases} \quad (2.23)$$

Ta thấy rằng, việc tìm kết quả tối ưu của $u_{j,i}$ và $v_{k,n}$ trong (2.22) và (2.23) thực chất là tìm 2 ma trận nhị phân U_{JI}^* (ma trận J hàng I cột) và V_{KN}^* (ma trận K hàng N cột) bằng cách tìm kiếm tương ứng trong 2 không gian nghiệm:

$$\mathcal{U} = \{U_{JI}^1, U_{JI}^2, \dots, U_{JI}^j, \dots, U_{JI}^J\} \text{ và } \mathcal{V} = \{V_{KN}^1, V_{KN}^2, \dots, V_{KN}^k, \dots, V_{KN}^{2KN}\}.$$

Ví dụ: nếu phần tử tại hàng thứ j và cột thứ i của ma trận U_{JI}^j có giá trị bằng 1, điều này có nghĩa là FBS thứ j lưu trữ video thứ i và nếu phần tử tại hàng thứ k và cột thứ n của ma trận V_{KN}^k có giá trị bằng 1, điều này có nghĩa là SU thứ k chia sẻ tài nguyên phổ tần kênh truyền xuống của nó cho cặp TX-RX thứ n .

Bài toán này được giải bằng cách sử dụng thuật giải vét cạn [54] để tìm nghiệm U_{JI}^* và V_{KN}^* được trình bày trong Thuật giải 2.1. Trong thuật giải này, Bước 1 là để khởi tạo 2 không gian tìm kiếm ma trận khả thi $\mathcal{U}^* \in \mathcal{U}$ và $\mathcal{V}^* \in \mathcal{V}$ được chọn từ tất cả các phần tử trong không gian tìm kiếm \mathcal{U} và \mathcal{V} thỏa (2.23). Trong các Bước 2 \rightarrow 8, với mỗi ma trận U_{JI} trong \mathcal{U}^* và mỗi ma trận V_{KN} trong \mathcal{V}^* , ta tính giá trị của hàm mục tiêu (2.19) để có được $R(U_{JI}, V_{KN})$. Kết thúc Bước 8, ta có \mathcal{R} gồm $L_{\mathcal{U}^*} \times L_{\mathcal{V}^*}$ giá trị của hàm mục tiêu và nghiệm tương ứng. Ở đây, $L_{\mathcal{U}^*}$ và $L_{\mathcal{V}^*}$ là số ma trận khả thi trong \mathcal{U}^* và số ma trận khả thi trong \mathcal{V}^* . Cuối cùng, ta tìm giá trị cực đại R^* trong \mathcal{R} và kết quả tối ưu tương ứng cho chỉ số lưu trữ U_{JI}^* và chỉ số chia sẻ tài nguyên phổ tần kênh truyền xuống V_{KN}^* trong Bước 9 và Bước 10.

Thuật giải 2.1: Tìm kiếm vét cạn

- 1 **Input:** tham số hệ thống được liệt kê ở Bảng 2-2
 Khởi tạo 2 không gian tìm kiếm ma trận khả thi $\mathcal{U}^* \in \mathcal{U}$ và $\mathcal{V}^* \in \mathcal{V}$ thỏa điều kiện (2.23)
 - 2 **Output:**
 \mathcal{R}^* : cực đại dung lượng phân phối trung bình đến các MU
 $\{U_{JI}^*, V_{KN}^*\}$
 - 3 **For** mỗi ma trận U_{JI} trong \mathcal{U}^* **do**
 - 4 **For** mỗi ma trận V_{KN} trong \mathcal{V}^* **do**
 - 5 $R(U_{JI}, V_{KN}) = R$
 - 6 $\mathcal{R} \leftarrow \mathcal{R} \cup R(U_{JI}, V_{KN})$
 - 7 End for
 - 8 **End for**
 - 9 $R^* = \max \mathcal{R}$
 - 10 $\{U_{JI}^*, V_{KN}^*\} = \operatorname{argmax} \mathcal{R}$
-

Thuận lợi của Thuật giải 2.1 trên là cách triển khai rất đơn giản để tìm kết quả chính xác toàn cục. Tuy nhiên, thuật giải 2.1 có độ phức tạp lớn $\mathcal{O}(2^{JI+KN})$ và thời gian thực thi sẽ lâu và điều này chắc chắn sẽ khó khả thi khi được triển khai trong mạng có quy mô lớn của 5G UDN. Trong trường hợp này, nếu số lượng các ứng viên nghiệm quá lớn (khi tăng các thông số kích thước của hệ thống), MBS có thể chia không gian các ứng viên nghiệm ra thành các không gian con và yêu cầu các trạm cơ sở con femto (FBS) thực thi thuật toán vét cạn trong các không gian con. Việc chia này dựa vào năng lực tính toán của các FBS, nghĩa là, FBS nào có năng lực tính toán càng cao thì không gian con sẽ lớn hơn so với các FBS khác để đảm bảo các kết quả tính toán được trả về đồng thời và nhanh nhất.

2.5. Đánh giá hiệu suất cơ chế SCS

2.5.1. Thiết lập thông số hệ thống

Bảng 2-2. Các tham số đầu vào cho bài toán SCS

Tham số	Giá trị	Lý giải/Tham chiếu
I	5 video	Các thông số kích thước hệ thống được chọn tương đối nhỏ để thực hiện dễ dàng bằng thuật giải vét cạn nhằm cho kết quả chính xác và có những đánh giá ban đầu về hiệu năng mô hình.
J	3 FBS	
K	3 SU	
N	5 cặp TX-RX	
(θ_n)	(5; 10; 20; 15; 25)	Các giá trị được chọn để đảm bảo tính hợp lý của phân bố Gamma cho mô hình thời gian kết nối đề xuất trong [81].
(κ_n)	(1; 4; 3; 2; 5)	
(β_n)	(0,1; 0,5; 0,9; 0,3; 0,7)	Phần trăm dung lượng lưu trữ còn trống của TX thứ n được chọn từ 0,1 đến 0,9 để đảm bảo tính phân hóa giữa các TX.
(T_{min}^i)	(1; 15; 10; 5; 20)s	Thời gian tối thiểu để truyền video thứ i được chọn từ rất thấp (1s) đến vừa phải (20s) để dễ dàng hiệu chỉnh bởi hệ số δ trong quá trình tính toán và mô phỏng, nhưng vẫn đảm bảo tính phân hóa giữa các video.
δ	10	Hệ số được chọn để linh hoạt điều chỉnh T_{min}^i .
α	2	Giá trị được chọn không quá cao cũng không quá thấp để cho thấy hiệu quả của bài toán tối ưu.
γ_0	5dB	Giá trị phù hợp để đảm bảo chất lượng của SU trong quá trình chia sẻ phổ tần kênh truyền xuống.
N_0	10^{-13} W	Giá trị được chọn thấp để không ảnh hưởng nhiều đến chất lượng hệ thống.
W	5MHz	Giá trị được chọn phù hợp để đảm bảo dung lượng phân phối trung bình của hệ thống cho ứng dụng truyền video.
P_M^0	5W	Giá trị công suất phát được chọn phù hợp với thực tế cho các trạm MBS, FBS và thiết bị di động.
P_F^j	1W	
P_T^n	0.1W	
$\{a, b\}$	0,5; 0,5	a và b được chọn bằng nhau để cho thấy quyết định lưu trữ một video hay không phụ thuộc như nhau vào độ nổi tiếng của video và dung lượng lưu trữ khả dụng của thiết bị di động.
C^*	$0,7 \times I \times J$	70% tổng dung lượng lưu trữ của tất cả các FBS được sử dụng là phù hợp.

Để thực hiện mô phỏng đánh giá hệ thống, các tham số đầu vào được thể hiện trong Bảng 2-2. Các tham số này được chọn lựa dựa trên 3 tiêu chí gồm: 1) đã được đề xuất sử dụng bởi các công trình nghiên cứu trước đó, 2) phù hợp với thực tế và 3) dựa trên kinh nghiệm sau nhiều lần thực thi. Nhằm thuận tiện trong tính toán nhưng không làm mất đi tính tổng quát, tôi bỏ qua ảnh hưởng của hệ số fading (fading coefficient) và chỉ quan tâm tới suy hao theo khoảng cách đường truyền với hệ số mũ suy hao là $\eta = 4$ và khoảng cách từ MBS đến MU trong phạm vi 300m đến 1500m, từ FBS đến MU trong phạm vi 50m đến 250m, từ SU đến TX trong phạm vi 50m đến 100m, và từ TX đến RX trong phạm vi 1m đến 50m.

2.5.2. *Đánh giá hiệu suất cơ chế SCS*

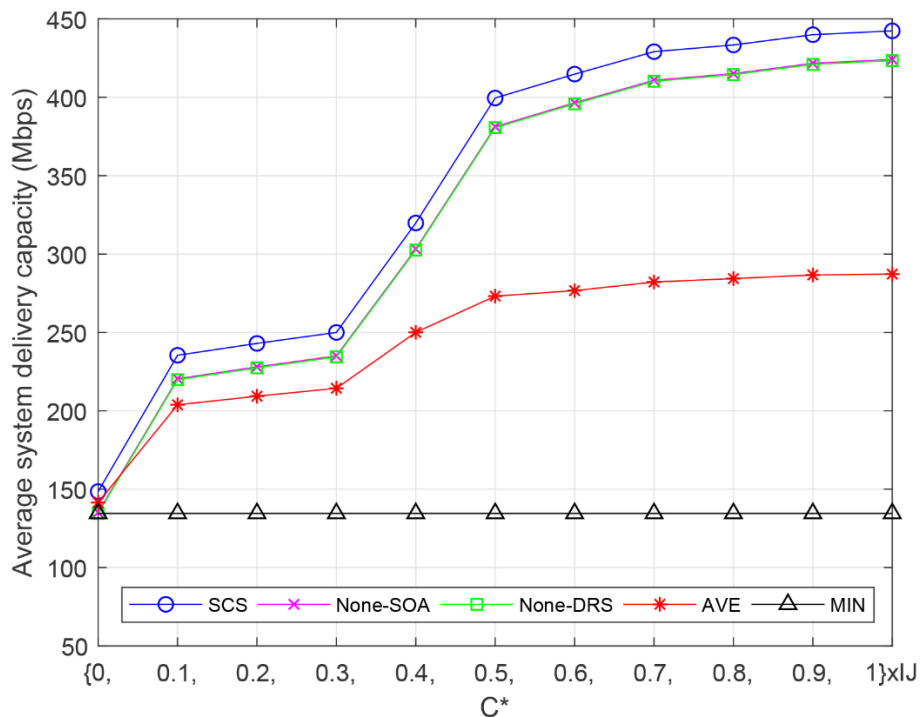
Cơ chế SCS được đánh giá dựa trên tiêu chí về dung lượng hệ thống (được tính bằng bps) theo các thông số đầu vào của hệ thống như: dung lượng lưu trữ của FBS, độ dài (thời lượng) video, độ lệch về sự nổi tiếng giữa các video, số lượng cặp TX-RX, số lượng FBS, ngưỡng SINR tại SU và khoảng cách giữa MBS và MU. Trong mỗi đánh giá, cơ chế SCS được so sánh với 4 cơ chế khác gồm:

- 1) Không xét đến yếu tố xã hội (None-SOA): Cơ chế này được dùng để so sánh với cơ chế có xét đến yếu tố xã hội nhằm làm rõ giả thuyết là nếu biết được mối quan hệ giữa các người dùng, việc chia sẻ tài nguyên và phục vụ người dùng sẽ hiệu quả hơn. Để tính toán trong trường hợp không xét yếu tố xã hội, các đặc điểm của mối quan hệ xã hội được thay đổi bằng cách tạo ra một ma trận ngẫu nhiên chứa tất cả các hoán vị của $\{\kappa_n\}$ trong công thức công thức (2.3) và Bảng 2-1. Sau khi có các đặc điểm mới, tức là, thời lượng và số lần kết nối mới theo các hoán vị khác nhau, đặc tính xấu nhất của mối quan hệ xã hội được sử dụng cho None-SOA.
- 2) Không chia sẻ tài nguyên phổ tần kênh truyền xuống (None-DRS): Nghĩa là cơ chế None-DRS không cho phép SU chia sẻ tài nguyên phổ tần kênh truyền xuống của nó cho các cặp TX-RX.

- 3) Trung bình của dung lượng phân phối (AVE): Là khả năng phân phối của hệ thống được tính trung bình trên tổng số cặp ma trận khả thi (tổng số ứng viên nghiệm khả thi thỏa các ràng buộc của bài toán tối ưu) được tạo ra trong Bước 1 của Thuật giải 2.1.
- 4) Cực tiểu của dung lượng phân phối (MIN): là khả năng phân phối hệ thống được tính là cực tiểu của \mathcal{R} thay vì cực đại \mathcal{R} trong Bước 9 của Thuật giải 2.1. Kịch bản MIN được sử dụng để cho thấy độ lệch giữa kết quả tối ưu và kết quả xấu nhất. Độ lệch này cũng cho phép ta đánh giá mức độ hiệu quả của giải pháp tối ưu ngoài việc so sánh với None-SOA, None-DRS và AVE.

2.5.2.1. Đánh giá SCS theo C^*

Trước tiên, tôi đánh giá cơ chế SCS, None-SOA, None-DRS, AVE và MIN theo dung lượng lưu trữ của FBS bằng cách đặt $C^* = \{0; 0,1; 0,2; \dots; 1\} \times I \times J$, ở đây, $C^* = I \times J$ có nghĩa là tất cả các video được lưu trong tất cả các FBS. Mục tiêu là để tìm ra dung lượng lưu trữ thích hợp tại các FBS thay vì lưu tất cả các video trong tất cả các FBS.

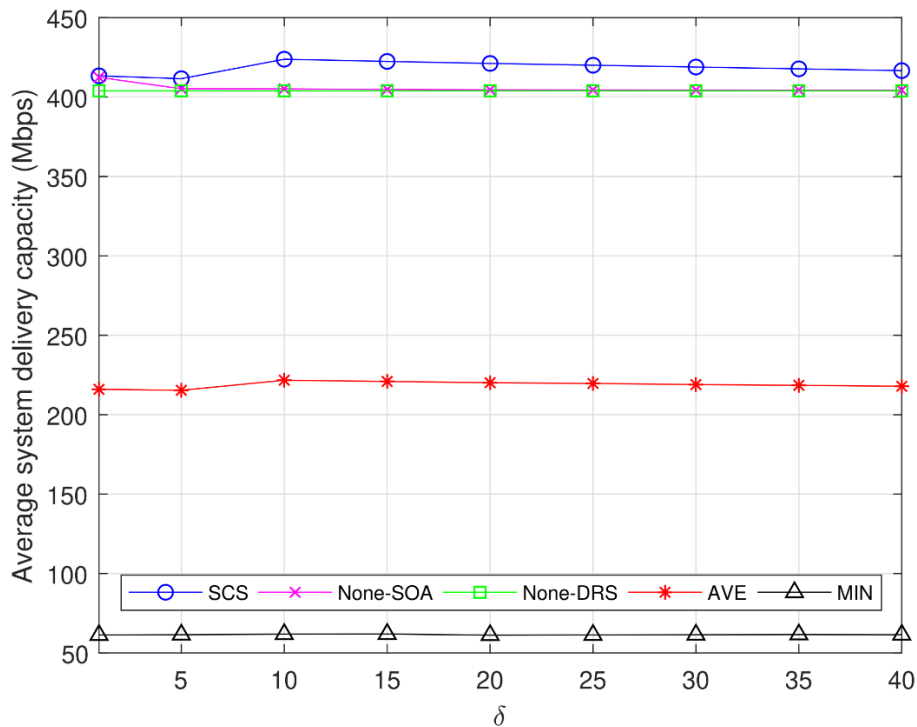


Hình 2-3. Dung lượng hệ thống theo khả năng lưu trữ của FBS

Như kết quả được hiển thị trong Hình 2-3, để có dung lượng lưu trữ thích hợp trong các FBS, tôi chọn $C^* = 0,7 \times I \times J$. Khi tăng C^* , cơ chế SCS luôn đạt hiệu suất cao hơn so với các cơ chế khác. Tương tự, trường hợp None-SOA và None-DRS cho dung lượng phân phối tốt hơn so với AVE và MIN, trong khi MIN là trường hợp xấu nhất. Kết quả trường hợp None-SOA và None-DRS tương đương nhau cho thấy rằng có quan tâm đến mối quan hệ xã hội người dùng và có chia sẻ tài nguyên đóng vai trò quan trọng như nhau trong việc nâng cao hiệu suất hệ thống.

2.5.2.2. Đánh giá SCS theo δ

Tiếp theo, tôi sẽ đánh giá dung lượng phân phối hệ thống của cơ chế SCS, None-SOA, None-DRS, AVE và MIN theo độ dài (thời lượng) video bằng cách thay đổi δ trong công thức (2.5) từ 1 đến 40.



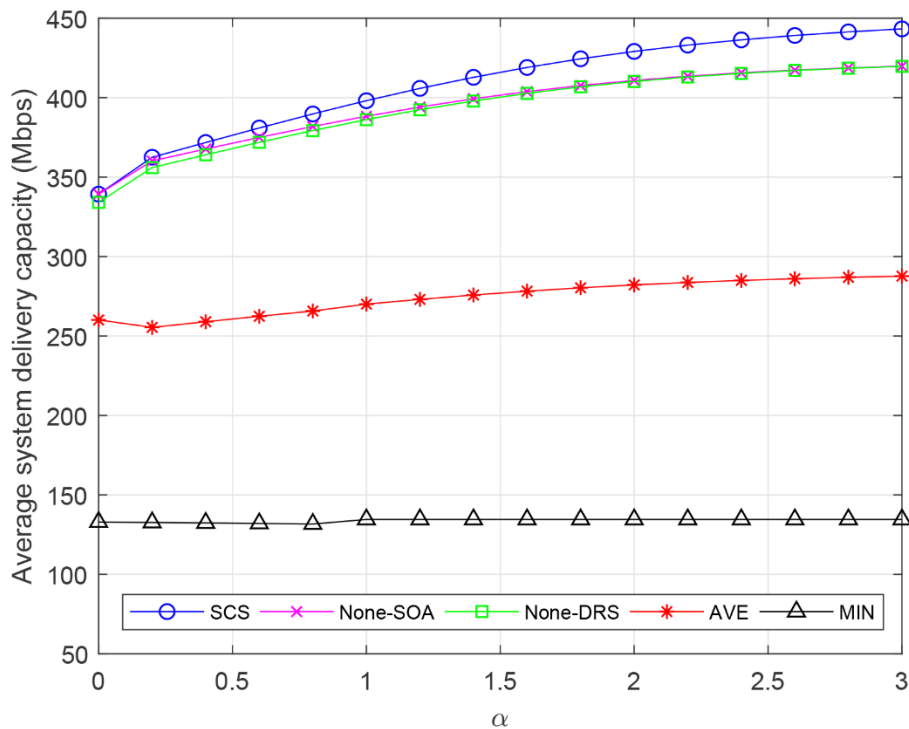
Hình 2-4. Dung lượng hệ thống theo hệ số δ

Kết quả Hình 2-4 cho thấy dung lượng phân phối hệ thống tăng nếu chúng ta tăng δ từ 1 đến 10, nhưng sẽ giảm nếu chúng ta tiếp tục tăng δ từ 10 đến 40. Lý do là nếu thời lượng video quá thấp thì xác suất thiết lập truyền thông D2D cũng quá cao và ngược lại. Cả xác suất thiết lập truyền thông D2D quá cao (nghĩa là, tác động nhiễu

quá cao lên SU) lẫn xác suất thiết lập truyền thông D2D quá thấp (nghĩa là không khai thác truyền thông D2D) đều khiến khả năng cung cấp dung lượng hệ thống thấp. Vì vậy, dung lượng phân phối hệ thống đạt giá trị cao nhất ở một giá trị δ nhất định là, tức là $\delta = 10$. Điều thú vị là ta có thể điều chỉnh thời lượng của video để đáp ứng các mối quan hệ xã hội của các cặp TX-RX nhằm đạt được dung lượng hệ thống cao nhất. Rõ ràng, hiệu suất của None-DRS không bị ảnh hưởng bởi δ . Cơ chế SCS cho kết quả tốt hơn so với các cơ chế khác và dung lượng hệ thống của nó có xu hướng giảm xuống bằng với None-SOA và None-DRS khi δ quá thấp (hoặc quá cao). Dung lượng hệ thống của None-SOA và None-DRS là tương tự nhau và vẫn tốt hơn so với AVE và MIN. Và cuối cùng, MIN luôn cho dung lượng hệ thống thấp nhất.

2.5.2.3. Đánh giá SCS theo α

α là hệ số mũ liên quan đến độ lệch về sự nổi tiếng giữa các video. Trong đánh giá này, tôi cho α thay đổi giá trị từ 0 đến 3. Kết quả được thể hiện như Hình 2-5.



Hình 2-5. Dung lượng hệ thống theo độ nổi tiếng của video

Trong hình này, khi tăng α , kết cho thấy rằng việc khai thác độ lệch nổi tiếng giữa các video có thể cải thiện hiệu năng hệ thống. Nếu tất cả các video có cùng độ nổi tiếng, tức là $\alpha = 0$, cơ chế SCS không có được lợi ích cao khi so với None-SOA và None-DRS. Tuy nhiên, SCS đạt dung lượng phân phối hệ thống cao hơn khi tăng α . Kết quả cho thấy rằng việc tập trung phục vụ một số video có độ nổi tiếng cao hơn (khi α tăng) sẽ mang lại dung lượng hệ thống cao hơn. Và qua đánh giá này, cơ chế SCS vẫn cho dung lượng hệ thống tốt nhất so với None-SOA, None-DRS, AVE và MIN.

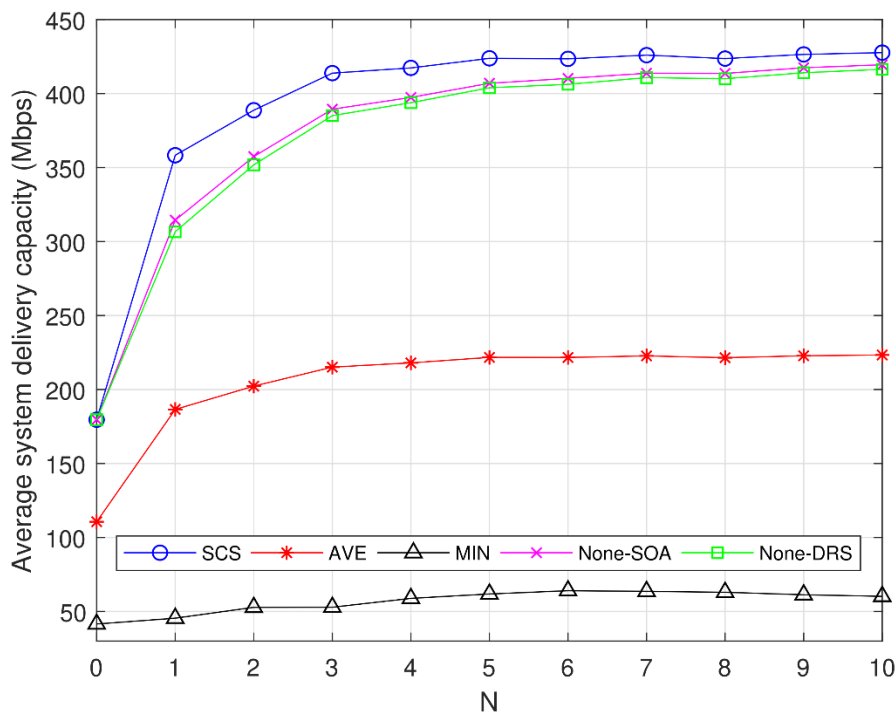
2.5.2.4. Đánh giá SCS theo N

Tiếp theo, dung lượng phân phối hệ thống của cơ chế SCS được đánh giá bằng cách thay đổi số lượng cặp TX-RX N từ 0 đến 10. Để làm như vậy, θ_n , κ_n và β_n được mở rộng thành:

$$\{\theta_n\} = \{5; 10; 20; 15; 25; 10; 5; 20; 25; 15\},$$

$$\{\kappa_n\} = \{1; 4; 3; 2; 5; 3; 2; 4; 1; 5\},$$

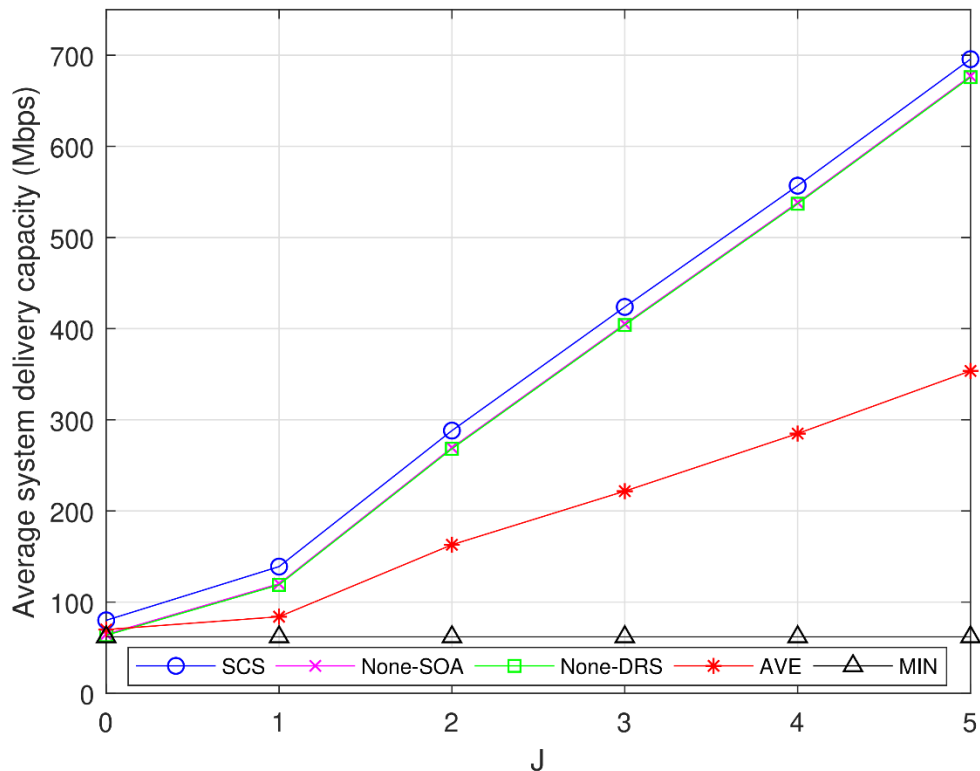
$$\{\beta_n\} = \{0,1; 0,5; 0,9; 0,3; 0,7; 1; 0,2; 0,4; 0,6; 0,8\}.$$



Hình 2-6. Dung lượng hệ thống theo số cặp D2D

Trong Hình 2-6, ta dễ dàng thấy rằng nếu $N = 0$, hệ thống không thu được bất kỳ lợi ích nào từ việc chia sẻ tài nguyên phổ tần kênh truyền xuống và mối quan hệ xã hội vì không có truyền thông D2D. Do đó, SCS, None-SOA và None-DRS có cùng kết quả và dung lượng phân phối hệ thống thấp. Nếu N tăng, dung lượng phân phối hệ thống sẽ cao hơn và khi N tăng đến một mức nào đó thì dung lượng phân phối hệ thống trở nên bão hòa. Tình trạng bão hòa là do các ràng buộc đối với SINR của SU làm giới hạn số lượng các cặp truyền thông của D2D. Trong trường hợp có nhiều cặp D2D dày đặc của 5G UDN, chúng ta cần thận trọng chọn số lượng cặp TX-RX sao cho SINR của SU được đảm bảo và dung lượng phân phối hệ thống đủ cao trước khi đạt đến tình huống bão hòa. Kết quả chứng minh rằng SCS vẫn cho dung lượng hệ thống tốt nhất so với None-SOA, None-DRS, AVE và MIN.

2.5.2.5. Đánh giá SCS theo J



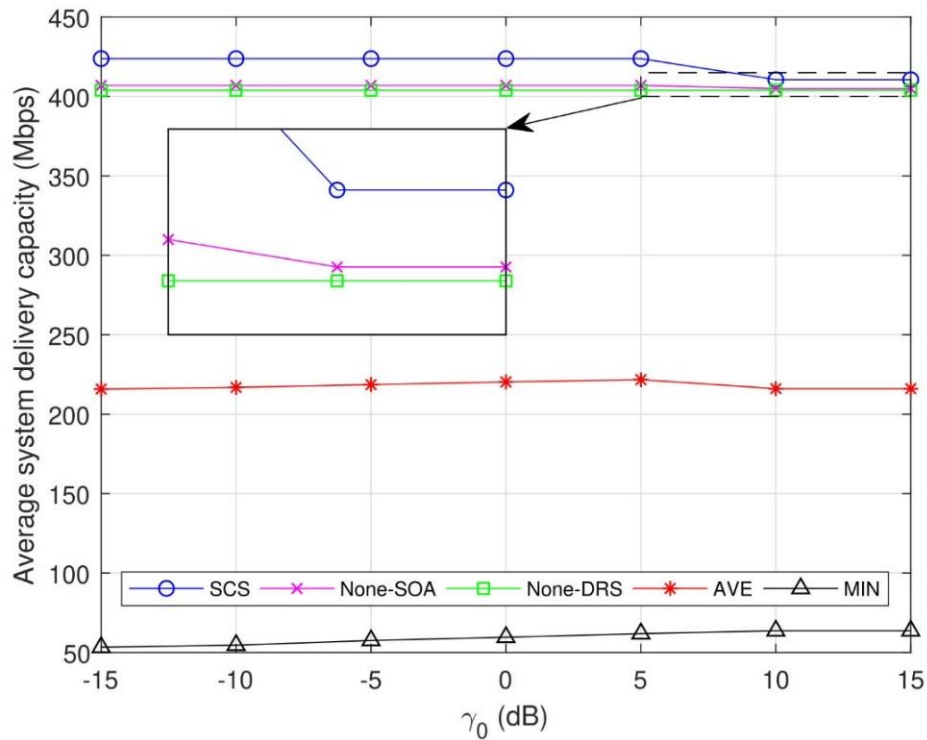
Hình 2-7. Dung lượng hệ thống theo số lượng FBS

Hình 2-7 biểu thị dung lượng phân phối hệ thống theo số lượng FBS J . Kết quả cho thấy số lượng FBS ảnh hưởng đáng kể đến dung lượng phân phối hệ thống. Không giống như khi tăng số lượng cặp D2D, việc tăng số lượng FBS trong 5G UDN

làm cho dung lượng phân phối hệ thống được cải thiện nhanh chóng mà không sớm bị bão hòa vì các điều kiện đối với can nhiễu liên tầng và đồng tầng không được xem xét nhờ sử dụng mô hình phân kênh và F-ALOHA [79, 80]. Dĩ nhiên, dung lượng hệ thống sẽ tăng chậm khi mật độ FBS đủ lớn, nghĩa là khó có khả năng tìm được vị trí lưu trữ mới tốt hơn khi tăng số lượng FBS đến một giá trị quá lớn. Trong thực tế, dung lượng hệ thống đạt bão hòa cũng còn do nhiều yếu tố khác, ví dụ như chi phí triển khai FBS, hiệu quả sử dụng năng lượng, hiệu quả sử dụng phổ tần và dung lượng lưu trữ tại FBS (được thực hiện trong Chương 3). Cần lưu ý rằng mặc dù cơ chế SCS đạt được dung lượng phân phối hệ thống cao nhất so với các cơ chế khác, nhưng sự khác biệt giữa SCS đối với None-SOA và None-DRS không thay đổi vì cả None-SOA và None-DRS đều sử dụng cùng vị trí lưu trữ tối ưu so với SCS.

2.5.2.6. Đánh giá SCS theo γ_0

Trong phần này, tôi kiểm chứng hiệu năng hệ thống dưới tác động bởi ngưỡng SINR (γ_0) của SU như được mô tả trong Hình 2-8.

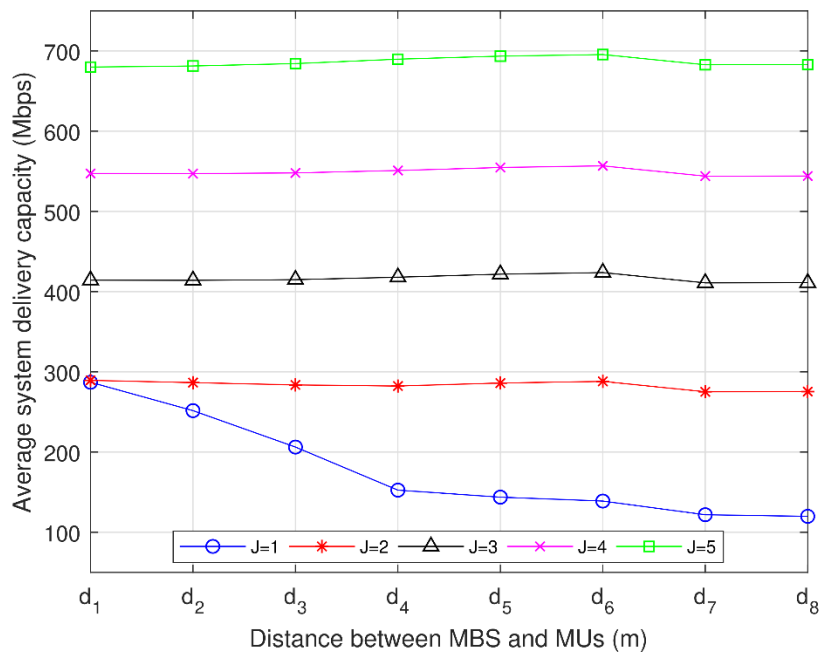


Hình 2-8. Dung lượng hệ thống theo Gamma

Như kết quả được mô tả trong Hình 2-8, khi γ_0 tăng, dung lượng phân phối hệ thống của SCS và None-SOA giảm và giảm dần bằng với dung lượng phân phối trong trường hợp None-DRS. Lý do là nếu γ_0 thấp, nhiều cặp TX-RX được thiết lập nhờ tài nguyên kênh truyền xuống được chia sẻ bởi SU nên dung lượng phân phối hệ thống cao hơn. Ngược lại, nếu γ_0 cao, số cặp TX-RX sẽ ít đi để giảm can nhiễu lên SU nên dung lượng phân phối hệ thống thấp hơn. Dung lượng phân phối hệ thống của MIN tăng vì khi γ_0 tăng, nhiễu ma trận ứng viên gây ra tác động nhiễu cao cho SU được loại khỏi không gian tìm kiếm \mathcal{V} . Rõ ràng thấy rằng, kết quả của None-DRS không bị ảnh hưởng bởi γ_0 vì không có hiệu ứng nhiễu từ truyền thông D2D trên SU. Trong trường hợp này, cơ chế SCS cũng cho kết quả tốt hơn None-SOA, None-DRS, AVE và MIN.

2.5.2.7. Đánh giá hệ thống theo khoảng cách giữa MBS và các MU

Cuối cùng, tôi đánh giá dung lượng phân phối hệ thống theo khoảng cách giữa MBS và MU. Cụ thể, khoảng cách (tính bằng mét) giữa MBS và MUs được thay đổi trong các phạm vi khác nhau: $d_1 \in [5; 10]$ m, $d_2 \in [10; 20]$ m, $d_3 \in [20; 100]$ m, $d_4 \in [100; 500]$ m, $d_5 \in [200; 1000]$ m, $d_6 \in [300; 1500]$ m, $d_7 \in [400; 2000]$ m, $d_8 \in [500; 2500]$ m.

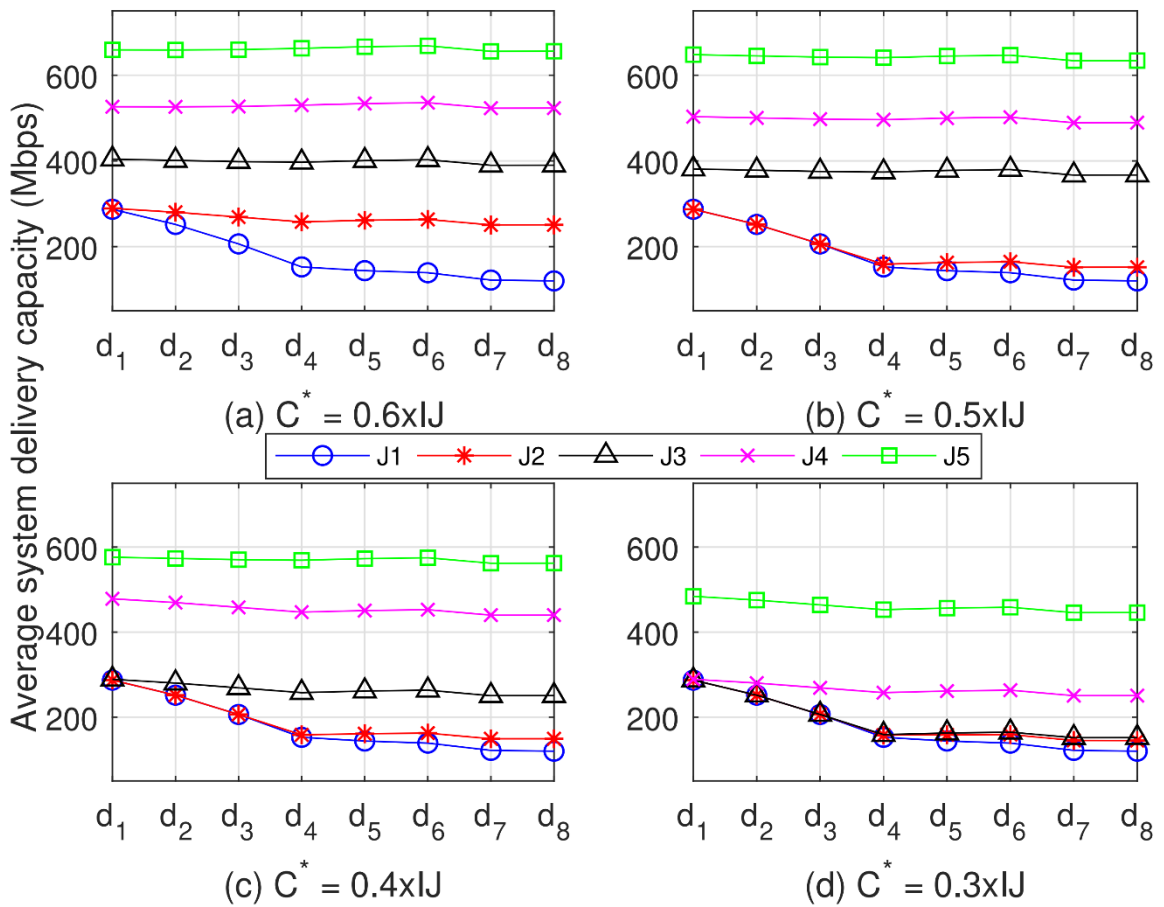


Hình 2-9. Dung lượng hệ thống theo khoảng cách giữa MBS và các MU

Trong đánh giá này, tôi chỉ xem xét dung lượng phân phối hệ thống của SCS với số lượng FBS J khác nhau từ 1 đến 5.

Kết quả trong Hình 2-9 cho thấy rằng nếu số lượng FBS thấp (nghĩa là $J = 1$ và $J = 2$), các MU có nhiều khả năng chỉ được MBS phục vụ cho nên dung lượng phân phối hệ thống thấp. Dễ dàng thấy rằng nếu các MU quá gần MBS, khi đó tất cả các MU hầu như chỉ được phục vụ bởi MBS nên các MU đều nhận cùng dung lượng phân phối của hệ thống. Tuy nhiên, nếu số lượng FBS tăng (tức là, $J = 3$, $J = 4$ và $J = 5$), các MU có nhiều khả năng được phục vụ thêm bởi FBS, nên khi đó dung lượng phân phối hệ thống cao hơn. Do ta thiết lập khả năng lưu trữ của FBS khá cao ($C^* = 0.7 \times I \times J$), nên dù các MU ở rất gần với MBS nhưng MBS không phục vụ các MU đó mà chúng đã được các FBS phục vụ. Một điều thú vị ở đây là, tồn tại khu vực tốt nhất (tức là, $d_6 \in [300, 1500]$ m) trong đó các MU được phục vụ với dung lượng phân phối cao nhất khi chúng không ở quá gần và không quá xa MBS. Nói cách khác, các MU trong khu vực này được phục vụ tối ưu bởi sự cộng tác của tất cả MBS, FBS và TX.

Để hiểu rõ hơn, tôi kiểm chứng thêm các kịch bản mô phỏng khi dung lượng lưu trữ của FBS thấp hơn, cụ thể là $C^* = \{0,6; 0,5; 0,4; 0,3\} \times I \times J$ như trong Hình 2-10. Có thể thấy rằng nếu dung lượng lưu trữ của FBS quá thấp ($C^* = 0,3 \times I \times J$) (Hình 2-10d), thì các MU chủ yếu được MBS phục vụ khi chúng ở gần MBS hơn. Đặc biệt, trong trường hợp $d_1 \in [5, 10]$ m, hiệu suất của SCS là như nhau đối với tất cả các trường hợp $J = 1$, $J = 2$, $J = 3$ và $J = 4$ vì tất cả các MU chỉ được phục vụ bởi MBS. Rõ ràng, nếu C^* tăng, khả năng các MU được phục vụ bởi MBS sẽ giảm. Điều quan trọng là sự gia tăng của cả C^* và J trong 5G UDN có thể giảm tải lưu lượng tuyến trực của MBS.



Hình 2-10. Dung lượng hệ thống theo khoảng cách giữa MBS và các MU với C^* thấp hơn

Ở những kết quả trong Hình 2-9 và Hình 2-10, việc thay đổi khoảng cách chỉ thực hiện giữa MBS và MU (TX và TX) mà không quan tâm đến sự dịch chuyển khoảng cách tương đối giữa FBS và MU. Nếu đưa vào sự dịch chuyển tương đối về khoảng cách giữa các phần tử trong toàn hệ thống đồng thời phải xem xét đến các dung lượng truyền dẫn trung gian thì rất khó thực hiện mô phỏng. Vì vậy, để đơn giản, dựa trên giả sử trong thực tế rằng mật độ FBS sẽ cao, nếu các MU dịch chuyển xa MBS thì trong quá trình dịch chuyển này, các MU sẽ chuyển từ “được phục vụ bởi FBS này” sang “được phục vụ bởi FBS khác” với kết quả tương đương nhau. Hiệu năng sẽ suy giảm khi mật độ FBS quá thấp (không có sự chuyển vùng được phục vụ bởi các FBS khác nhau).

2.6. Kết luận chương 2

Trong Chương 2, tôi đã đề xuất một cơ chế lưu trữ và chia sẻ tài nguyên (SCS) tối ưu có xét đến mối quan hệ xã hội của người dùng để phân phối video với dung lượng cao trong 5G UDN. Cụ thể, việc lưu trữ nội dung video tại MBS, FBS và TX được sử dụng để đưa video đến gần hơn với các MU và tài nguyên phổ tần kênh truyền của SU được chia sẻ với các cặp TX-RX trong truyền thông D2D. Để hiệu quả hơn, mối quan hệ xã hội của các cặp TX-RX và độ phổ biến của các video được khai thác trong cơ chế SCS. Cơ chế SCS được phân tích cẩn thận bằng cách tính đến ngưỡng SINR của các SU để đảm bảo QoS của chúng. Ngoài ra, kết quả cũng cho thấy rằng: 1) tồn tại một giá trị thời lượng trung bình cho nhóm các video được quan tâm và 2) tồn tại khu vực tốt nhất mà ở đó hệ thống đạt được dung lượng phân phối cao nhất. Những phát hiện này có thể giúp cả nhà cung cấp dịch vụ Internet và nhà cung cấp nội dung hiểu rõ hơn về đặc điểm của 5G UDN để phục vụ các MU hiệu quả hơn.

Tuy nhiên, cơ chế SCS vẫn còn những hạn chế như: 1) chưa quan tâm đến độ phân giải của thiết bị người dùng, từ đó, xác định nhu cầu người dùng để lưu trữ video với các phiên bản khác nhau (chất lượng/độ phân giải khác nhau) nhằm đồng thời sử dụng hiệu quả dung lượng lưu trữ và đáp ứng được nhu cầu người dùng; 2) các cặp thu phát video D2D được định sẵn, nghĩa là, chưa cho phép thiết bị thu chọn lựa thiết bị phát có lưu trữ video được yêu cầu; và 3) thông số đánh giá hệ thống là dung lượng phân phối video (được tính bằng bps) không đặc tả được chất lượng video tại đầu thu một cách tường minh. Những hạn chế này sẽ được khắc phục trong Chương 3.

CHƯƠNG 3. LƯU TRỮ ĐA PHÂN GIẢI VÀ CHIA SẺ TÀI NGUYÊN TỐI ƯU THEO NHU CẦU NGƯỜI DÙNG NÂNG CAO CHẤT LƯỢNG TRUYỀN VIDEO TRONG 5G UDN

***Tóm tắt:** Trong Chương 3 sẽ trình bày cơ chế lưu trữ đa phân giải và chia sẻ tài nguyên tối ưu (CRS – Multi-Resolution Caching and Resource Sharing Optimization) theo nhu cầu người dùng nhằm nâng cao chất lượng truyền video trong mạng 5G siêu dày đặc (UDN – Ultra-dense Network). Cơ chế CRS một mặt vẫn cho phép người dùng di động (MU – Mobile User) truy xuất video linh hoạt từ 3 tầng gồm: trạm cơ sở chính (MBS – Macro Base Station), trạm cơ sở nhỏ femto (FBS – Femto Base Station) và từ người dùng khác có lưu trữ video (CU – Caching User) thông qua truyền thông từ thiết bị đến thiết bị (D2D – Device-to-Device), mặt khác khắc phục những vấn đề của cơ chế SCS trong Chương 2 liên quan đến cách thức lưu trữ, chia sẻ tài nguyên và thông số đánh giá hệ thống. Như vậy, bài toán CRS được xây dựng để xác định đồng thời: 1) video được phân phối từ tầng nào, 2) video với độ phân giải nào sẽ được lưu trữ trong FBS nào, và 3) các bộ 3 (SU, CU, NU) nào được chọn. Mục tiêu của cơ chế CRS là cực đại chất lượng phát lại (playback) video và sử dụng hiệu quả dung lượng lưu trữ tại FBS theo nhu cầu người dùng. Công trình nghiên cứu liên quan đến Chương 3 là [C2] và [C3] được trình bày tại “Danh mục các công trình đã công bố của nghiên cứu sinh” ở cuối luận án.*

3.1. Giới thiệu cơ chế CRS

Trong mạng 5G UDN, tài nguyên bộ nhớ đa tầng tại trạm cơ sở lớn (MBS – Macro Base Station), trạm cơ sở nhỏ (SBS – Small Cell Base Station) và thiết bị di động đều có thể được tận dụng để lưu trữ nhằm đưa nội dung video đến gần với người dùng di động (MU – Mobile User) [16, 18, 54, 91]. Không như các kỹ thuật khác, kỹ thuật lưu trữ không làm thay đổi kiến trúc mạng với chi phí thấp [92]. Kỹ thuật lưu trữ làm giảm lưu lượng tại các tuyến trục (backhaul link) của MBS và SBS vốn thường xảy ra quá tải do lượng lớn MU truy cập [50, 67, 93, 94]. Nhờ vào việc rút ngắn khoảng cách truyền thông kết hợp với việc sử dụng và chia sẻ hiệu quả tài

nguyên về nội dung, phổ tần, băng thông và năng lượng một cách hiệu quả, kỹ thuật lưu trữ sẽ giúp cho MU dễ dàng truy cập nội dung với chất lượng dịch vụ (QoS – Quality of Service) cao hơn [18, 25, 39, 40, 58, 59].

Liên quan đến kỹ thuật lưu trữ đa tầng, tác giả P. Lin cùng các cộng sự [91] đã đề xuất một cơ chế lưu trữ cộng tác giữa MBS và MU để giảm độ trễ phân phối, đồng thời hạn chế dư thừa nội dung. Phức tạp hơn nhưng đạt kết quả tốt hơn, tác giả J. Wen [16] và tác giả X. Li [18] cùng các cộng sự đã triển khai kỹ thuật lưu trữ trong mạng 5G đa tầng bao gồm MBS và một hoặc hai loại SBS, để đạt được xác suất truy cập thành công và dung lượng hệ thống cao bằng cách tìm vị trí lưu trữ tối ưu. Ngoài những kết quả đạt được trong [16, 18, 91], các tác giả trong [54] đã tận dụng tài nguyên phổ tần kênh truyền xuống của MU để truyền nội dung video trực tiếp từ thiết bị đến thiết bị (D2D – Device-to-Device). Tuy nhiên, hạn chế của các kỹ thuật trong [16, 18, 54, 91] đó là: 1) không cho phép mở rộng chọn lựa một cách linh động bộ 3 người dùng gồm người dùng sở hữu phổ tần kênh truyền xuống (SU – Spectrum owning User), người dùng lưu trữ video (CU – Caching User) và người dùng thông thường (NU – Normal User) yêu cầu truy xuất video, để thiết lập phiên truyền thông video D2D tốt nhất; 2) không cho phép chọn lựa lưu trữ đa phân giải với các phiên bản video có chất lượng khác nhau; và 3) không quan tâm đến nhu cầu người dùng. Những hạn chế này khiến hệ thống không thể để đạt được hiệu suất hệ thống cao, nghĩa là không tận dụng và phân bổ tài nguyên hợp lý với QoS cao.

Trong chương này, để khắc phục những hạn chế trên, tôi đề xuất cơ chế lưu trữ đa phân giải và chia sẻ tài nguyên tối ưu (CRS – Multi-Resolution Caching and Resource Sharing Optimization) theo nhu cầu người dùng nhằm nâng cao chất lượng truyền video trong 5G UDN. Cơ chế CRS cho phép MU truy xuất video một cách linh động từ 3 tầng gồm: tầng MBS, tầng SBS (cụ thể là FBS – Femtocell Base Station) và tầng thiết bị người dùng. Tại tầng MBS, dung lượng lưu trữ được giả sử là rất lớn để lưu tất cả các video đang được MU quan tâm. Tại tầng FBS, do dung lượng lưu trữ có giới hạn, các nội dung video với độ phân giải khác nhau sẽ được chọn lựa lưu trữ hoặc không lưu trữ trong các FBS một cách tối ưu. Tại tầng thiết bị

người dùng, MU được chia ra làm 3 loại gồm SU, CU và NU. Một phiên truyền thông video sẽ được thiết lập từ một bộ 3 (SU, CU, NU), trong đó, CU có lưu trữ video, NU yêu cầu truy xuất video, SU chia sẻ tài nguyên phổ tần kênh truyền xuống của nó để CU sử dụng lại phổ tần này cho việc truyền video đến NU thông qua truyền thông D2D. Trong mô hình này, SU và CU sẽ truy xuất video từ MBS và FBS, nhưng NU sẽ được phép truy xuất video từ cả MBS, FBS và CU. Vấn đề được đặt ra đó là 1) làm thế nào để lưu trữ các video với độ phân giải khác nhau vào các FBS và làm thế nào để chia sẻ tài nguyên từ SU và CU cho NU, hay nói cách khác là thiết lập bộ 3 (SU, CU, NU) một cách tốt nhất để đạt được hiệu suất hệ thống cao nhất.

Vấn đề nêu trên được giải quyết bằng việc xây dựng bài toán tối ưu CRS dựa vào mô hình hệ thống. Giải bài toán tối ưu CRS sẽ cho phép tìm được cơ chế lưu trữ video tối ưu (video với độ phân giải nào sẽ được lưu trữ tại FBS nào) và cơ chế chọn lựa bộ chọn lựa bộ 3 (SU, CU, NU) tối ưu để SU chia sẻ tài nguyên phổ tần kênh truyền xuống cho CU thực hiện truyền video có độ phân giải phù hợp đến NU qua truyền thông D2D. Để đạt được kết quả tối ưu cho cơ chế lưu trữ và chia sẻ tài nguyên, bài toán CRS có quan tâm đến 3 ràng buộc quan trọng. Ràng buộc thứ nhất là ràng buộc về dung lượng yêu cầu bởi người dùng được xác định từ độ phân giải/khả năng hiển thị video của thiết bị người dùng, ràng buộc này sẽ giới hạn băng thông của hệ thống một cách hợp lý (tránh lãng phí), đủ đáp ứng nhu cầu người dùng. Ràng buộc thứ hai là ràng buộc về dung lượng lưu trữ của FBS, ràng buộc này sẽ cho phép các FBS phải chọn lựa lưu trữ các video có độ phân giải phù hợp, đáp ứng nhu cầu người dùng nhằm sử dụng hiệu quả dung lượng lưu trữ. Và ràng buộc thứ ba liên quan đến việc đảm bảo QoS cho SU. Như đã đề cập, SU chia sẻ tài nguyên phổ tần kênh truyền xuống của nó để thiết lập phiên truyền thông video từ CU đến NU. Trong quá trình truyền này, CU sử dụng lại phổ tần kênh truyền xuống của SU nên tín hiệu phát từ CU sẽ gây can nhiễu lên SU. Để đảm bảo QoS cho SU (hay nói cách khác là đảm bảo công bằng cho SU), bài toán CRS có xem xét đến ràng buộc về ngưỡng tỷ số tín hiệu trên can nhiễu và nhiễu trắng (SINR – Signal to Interference plus Noise Ratio) tại SU trong quá trình chọn lựa bộ 3 (SU, CU, NU).

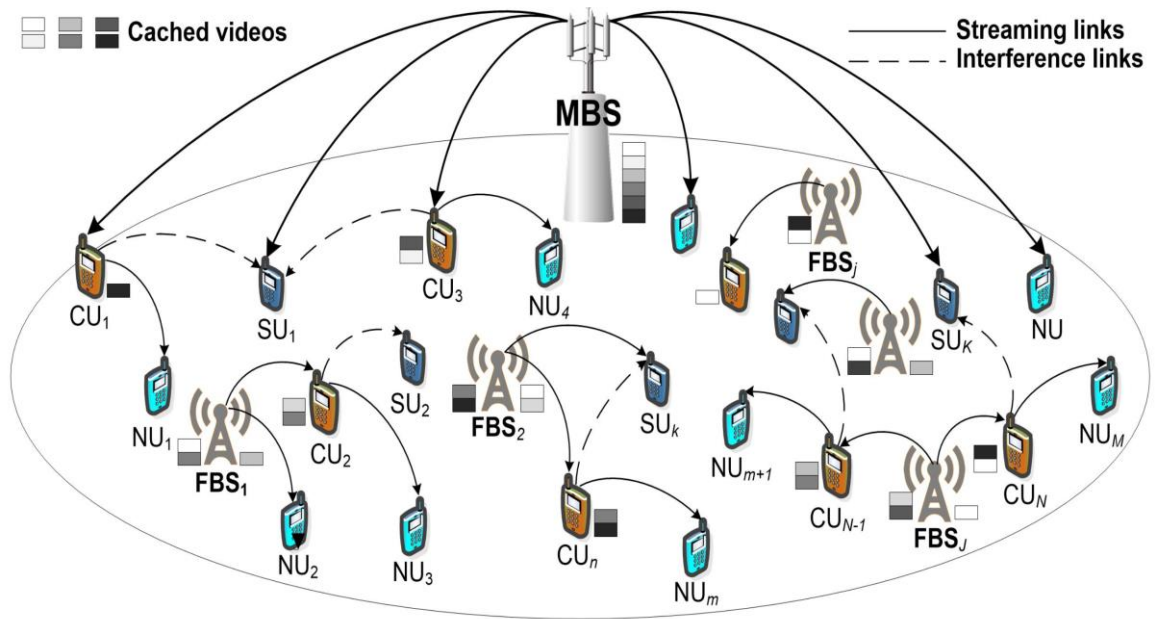
Việc giải bài toán tối ưu CRS được thực hiện bằng cách thiết lập chỉ số lưu trữ và chỉ số chia sẻ tài nguyên dưới dạng một vectơ hoặc ma trận nhị phân. Các chỉ số được gán giá trị 1 (hoặc 0) có nghĩa là một FBS sẽ có (hoặc không) lưu trữ video với độ phân giải phù hợp và một SU sẽ có (hoặc không) chia sẻ tài nguyên phổ tần kênh truyền xuống để CU truyền video với độ phân giải phù hợp đến NU. Các phương pháp tìm kiếm để giải bài toán tối ưu cho kết quả nhị phân có thể được áp dụng như quy hoạch động [40], thuật giải học ngẫu nhiên [39], thuật giải học trò chơi [58], thuật giải tham lam [25], thuật giải heuristic [59], thuật giải heuristic tham lam [18] và thuật giải vét cạn (EA – Exhaustive Algorithm) [54-57, 95].

Trong thuật giải vét cạn đã sử dụng, đặc biệt ở Chương 3, kết quả tối ưu đạt được là chính xác nhưng tốn nhiều thời gian thực thi cũng như dung lượng tính toán. Do đó, rất khó để triển khai thuật giải vét cạn khi gia tăng các phần tử trong 5G UDN. Trong khi đó, các phương pháp khác được đề xuất trong [18, 25, 39, 40, 43, 50, 58, 59] có thời gian thực thi nhanh hơn khá nhiều với dung lượng tính toán phù hợp nhưng độ chính xác lại thấp. Ví dụ, độ chính xác là khoảng 81% đến 87% với thuật giải heuristic tham lam [59] hoặc thậm chí khoảng 50% với thuật giải tham lam trong trường hợp xấu nhất [58]. Vì vậy, trong chương này, một phương pháp tìm kiếm thích nghi, cụ thể là thuật giải di truyền (GA – Genetic Algorithms) có điều chỉnh cho phù hợp với cách thức giải bài toán tối ưu trong không gian ma trận nhị phân, được áp dụng để giải bài toán tối ưu CRS. Song song với GA, thuật giải vét cạn cũng được triển khai (nhưng với lượng nhỏ các phần tử trong 5G UDN) để làm cơ sở kiểm chứng độ chính xác và thời gian thực thi cũng như chọn lựa các thông số phù hợp cho GA. Từ đó, GA sẽ được triển khai với số lượng các phần tử trong 5G UDN lớn hơn. Những ưu nhược điểm của cơ chế CRS được kiểm chứng thông qua tính toán mô phỏng và đối sánh với các cơ chế thông thường khác.

3.2. Mô hình truyền video trong 5G UDN với cơ chế CRS

Trong Chương 3, mô hình truyền video trong 5G UDN với cơ chế CRS được đề xuất như Hình 3-1. Các thông số tính toán liên quan đến mô hình được liệt kê và diễn giải trong Bảng 3-1. Trong mô hình này, có một trạm cơ sở lớn (MBS), J trạm

cơ sở nhỏ femto (FBS), và các người dùng di động (MU) có nhu cầu truy xuất nội dung video bao gồm K người dùng sẵn sàng chia sẻ tài nguyên phổ tần kênh truyền xuống (SU), N người dùng đã lưu trữ một số video có độ phân giải khác nhau (CU), M người dùng thông thường (NU) và I video. Video thứ i , $i = 1, 2, \dots, I$, có V_i mức phân giải được mã hóa ở các tốc độ khác nhau tương ứng với các mức chất lượng khác nhau. Các video này được gửi từ MBS đến các MU thông qua kênh truyền di động thông thường, từ FBS đến các MU bằng cách sử dụng các cơ chế phân kênh và F-ALOHA để tránh can nhiễu [79] và từ CU đến NU qua truyền thông từ thiết bị đến thiết bị (D2D) bằng cách tái sử dụng tài nguyên phổ tần kênh truyền xuống được chia sẻ bởi SU.

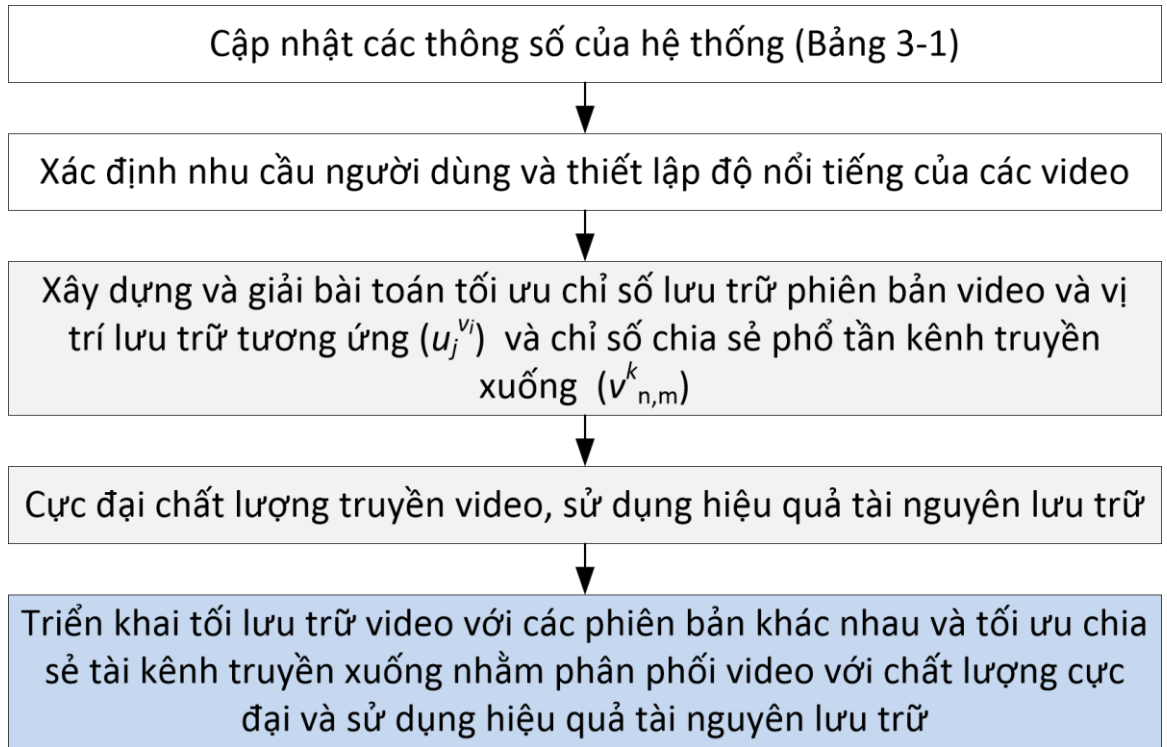


Hình 3-1. Mô hình hệ thống của CRS

Bảng 3-1. Các ký hiệu sử dụng cho bài toán CRS

Ký hiệu	Diễn giải
I	Số video
J	Số trạm cơ sở nhỏ femto (FBS)

K	Số người dùng di động chấp nhận chia sẻ tài nguyên phổ tần kênh truyền xuống (SU)
N	Số người dùng di động đã lưu trữ các video (CU) có độ phân giải khác nhau
M	Số người dùng di động thông thường (NU)
V_i	Số mức phân giải của video thứ $i, i=1, 2, \dots, I$
$w_{n,m}^k$	Chỉ số chia sẻ tài nguyên phổ tần kênh truyền xuống, khi $w_{n,m}^k = 1$ thì SU thứ k sẽ chia sẻ kênh truyền xuống của nó cho CU thứ n và NU thứ m thực hiện truyền video qua truyền thông D2D
$u_j^{v_i}$	Chỉ số lưu trữ video tại FBS, khi $u_j^{v_i} = 1$ thì FBS thứ j lưu phiên bản video v_i và ngược lại
P_M^0	Công suất phát của trạm cơ sở lớn (MBS)
P_F^j	Công suất phát của trạm cơ sở nhỏ femto (FBS) thứ j
P_T^n	Công suất phát của thiết bị CU thứ n trong truyền thông D2D.
G_0^k	Độ lợi kênh từ MBS tới SU k
N_0	Công suất nhiễu trắng Gaussian (AWGN – Additive White Gaussian Noise)
W	Băng thông hệ thống
L_F^i	Tổng dung lượng cần có để lưu video thứ i với các mức phân giải khác nhau trong các FBS theo tính toán
L_{max}^i	Giới hạn trên về dung lượng bộ nhớ của các FBS để lưu trữ video thứ i với các mức phân giải khác nhau, cụ thể $L_{max}^i = r_i \times I \times J \times \max\{L_i^{v_i}, v_i = 1, 2, \dots, V_i\},$ trong đó $r_i = i^{-\alpha} (\sum_{i=1}^I i^{-\alpha})^{-1}$ là độ nổi tiếng của video với hệ số mũ phản ánh độ lệch nổi tiếng α được mô hình hóa dưới dạng phân phối Zipf-like [88] và $L_i^{v_i}$ là kích thước của video thứ i với mức phân giải v_i
C	Tổng dung lượng (bps) được yêu cầu bởi các MU theo tính toán
C^*	Giới hạn trên về tổng dung lượng được yêu cầu bởi các MU
μ và δ	μ và δ với $0 < \mu, \delta \leq 1$ là hệ số được sử dụng để điều chỉnh giới hạn trên L_{max}^i và C^*



Hình 3-2. Lưu đồ hoạt động cơ chế CRS

Trong mô hình truyền video với cơ chế CRS, giả sử các tham số hệ thống không thay đổi trong quá trình truyền phiên bản video dài nhất trong số tất cả các phiên bản được yêu cầu. Cũng cần nhấn mạnh thêm rằng cơ chế CRS sẽ phục vụ hiệu quả cho các MU trong khu vực đông đúc như, sân vận động, phòng hòa nhạc, hội trường, bảo tàng, tòa nhà văn phòng làm việc, bệnh viện, khuôn viên trường học, v.v.... Bất cứ khi nào MBS dự đoán có sự gia tăng yêu cầu truy cập các video nào đó, thì nó sẽ thực thi cơ chế CRS như được mô tả trong Hình 3-2, cụ thể gồm 3 bước như sau:

- Đầu tiên, MBS thu thập các thông số của hệ thống như đặc tính kênh truyền; dung lượng lưu trữ của FBS và CU; thông tin video (các mức phân giải, kích thước, và mức độ phổ biến của từng video); dung lượng yêu cầu bởi các MU; và tỉ số tín hiệu trên can nhiễu và nhiễu trắng (SINR) của SU cùng các thông số liên quan khác được trình bày trong Bảng 3-1.

- Tiếp theo, dựa vào các thông số của hệ thống, các biểu thức liên quan đến hàm mục tiêu và các ràng buộc được tính toán để xây dựng bài toán tối ưu có ràng buộc CRS. Bài toán CRS sẽ được giải để tìm ra chỉ số $u_j^{v_i}$ tối ưu nhằm xác định các phiên bản video phù hợp và vị trí tối ưu để lưu trữ chúng trong các FBS ($u_j^{v_i} = 1$ nghĩa là video thứ i với mức phân giải v_i được lưu trữ tại FBS j ; và ngược lại thì $u_j^{v_i} = 0$). Đồng thời, việc giải bài toán CRS cũng xác định được chiến lược chia sẻ tài nguyên phổ tần kênh truyền xuống tối ưu, cụ thể là tìm ra chỉ số $w_{n,m}^k$ tối ưu để xác định các các bộ 3 (SU, CU, NU) cho truyền thông video D2D ($w_{n,m}^k = 1$ nghĩa là SU k đồng ý chia sẻ phổ tần kênh truyền xuống của nó cho CU n thực hiện truyền video đến NU m qua truyền thông D2D; ngược lại thì $w_{n,m}^k = 0$). Phân bố của các CU được chia sẻ tài nguyên phổ tần kênh truyền xuống từ SU k , $k = 1, 2, \dots, K$, được mô hình bởi quá trình Poisson đồng nhất (Poisson Point Process) với mật độ λ_c^k [96]. Bài toán tối ưu CRS cũng tính đến các ràng buộc về dung lượng mà các MU yêu cầu để phát lại (playback) video, về dung lượng lưu trữ của FBS để sử dụng tài nguyên lưu trữ một cách hiệu quả và về tỉ số tín hiệu trên can nhiễu và giao thoa (SINR) để đảm bảo chất lượng dịch vụ (QoS) của SU.
- Cuối cùng, tất cả các video với mức phân giải khác nhau được lưu trong FBS cũng như tất cả các bộ 3 (SU, CU, NU) được chỉ định để sẵn sàng truyền từ MBS, FBS và CU đến các MU với chất lượng phát lại cao đồng thời đáp ứng được nhu cầu về dung lượng của các MU, sử dụng hiệu quả dung lượng lưu trữ của FBS và đảm bảo tỉ số tín hiệu trên can nhiễu và giao thoa (SINR) của SU (hay nói cách khác là đảm bảo chất lượng/sự công bằng cho SU).

3.3. Tính toán các thông số hệ thống với cơ chế CRS

Để xây dựng bài toán tối ưu CRS, ta phải tính toán chất lượng trung bình của các video với các mức phân giải khác nhau nhận được tại các MU, tức là cần phải tính xác suất đạt dung lượng kênh tương ứng với các mức phân giải của các video được thu tại các SU, CU và NU. Gọi $u_{j,i}^{v_i}$ là chỉ số lưu trữ của video thứ i phiên bản thứ v_i trên FBS thứ j . Nếu $u_j^{v_i} = 1$ thì FBS thứ j sẽ lưu trữ video phiên bản thứ v_i của video thứ i ; ngược lại $u_j^{v_i} = 0$ là không lưu. Và gọi $w_{n,m}^k$ là chỉ số chia sẻ tài nguyên phổ tần kênh truyền xuống của SU thứ k cho CU thứ n truyền video đến NU thứ m qua truyền thông D2D. Nếu $w_{n,m}^k = 1$ thì SU thứ k đồng ý chia sẻ tài nguyên phổ tần kênh truyền xuống của nó với CU thứ n và NU thứ m ; ngược lại $w_{n,m}^k = 0$ thì không chia sẻ. $u_{j,i}^{v_i}$ và $w_{n,m}^k$ chính là những biến chỉ số cần phải tìm trong bài toán tối ưu CRS. Các thông số cho mô hình truyền video trong 5G UDN với cơ chế CRS được tính toán ở phần sau.

3.3.1. Xác suất đạt dung lượng tại SU

Trong cơ chế CRS, ngoài việc chia sẻ tài nguyên phổ tần kênh truyền xuống, SU thứ k cũng sẽ có yêu cầu video thứ i với mức phân giải v_i từ FBS thứ j hoặc từ MBS. Để xác định được xác suất đạt dung lượng tại SU thứ k , ta cần tính dung lượng tại SU thứ k từ FBS thứ j và từ MBS lần lượt như sau:

$$C_j^{k,v_i} = W \log_2 \left(1 + \frac{u_j^{v_i} P_j G_j^k}{N_0} \right) \quad (3.1)$$

và

$$C_0^{k,v_i} = W \log_2 \left(1 + \frac{\prod_{j=1}^J (1 - u_j^{v_i}) P_0 G_0^k}{N_0 + I_{C,S}^{k,v_i}} \right) \quad (3.2)$$

Trong đó

- W là băng thông hệ thống
- P_j và P_0 là công suất phát của FBS thứ j và MBS

- G_j^k và G_0^k là độ lợi kênh từ FBS thứ j và MBS đến SU thứ k , mô hình bởi hệ số Rayleigh fading có phân phối mũ độc lập đồng nhất với trị trung bình bằng 1 và bởi hệ số suy hao công suất theo khoảng cách d với hệ số mũ suy hao η (tương tự như trong Chương 2)
- N_0 là công suất nhiễu trắng Gaussian
- $I_{C,S}^{k,v_i} = \sum_{n=1}^N \sum_{m=1}^M w_{n,m}^k p_n^{v_i} P_C^k G_n^k$ là can nhiễu từ các CU tác động lên SU thứ k , trong đó:
 - P_C^k là công suất phát của CU khi nhận chia sẻ kênh truyền từ SU thứ k ,
 - G_n^k là độ lợi kênh giữa CU thứ n và SU thứ k ,
 - $p_n^{v_i}$ là xác suất mà CU thứ n lưu trữ video thứ i với mức phân giải v_i , được tính theo công thức sau:

$$p_n^{v_i} = ar_i + b\theta_n^{v_i} \quad (3.3)$$

với $a, b \in [0, 1]$, r_i là độ nổi tiếng của video thứ i được mô hình bởi phân phối Zipf-like [88] được tính bởi:

$$r_i = \frac{i^{-\alpha}}{\sum_{i=1}^I i^{-\alpha}} \quad (3.4)$$

trong đó:

- $\alpha \geq 0$ là hệ số mũ phản ánh độ lệch nổi tiếng giữa các video
- và $\theta_n^{v_i}$ được tính theo:

$$\theta_n^{v_i} = r_n \left(1 - \frac{L_i^{v_i} - \min\{L_i^{v_i}, v_i = 1, 2, \dots, V_i\}}{L_i^{v_i}} \right) \quad (3.5)$$

trong đó:

- r_n là phần trăm bộ nhớ còn trống của CU thứ n

- $L_i^{v_i}$ là kích thước của video thứ i với mức phân giải v_i

Đặt $C_{th}^{v_i}$ là ngưỡng dung lượng (tính bằng Kb/giây) cần thiết để video thứ i với phiên bản v_i được truyền để phát lại, xác suất đạt dung lượng được tính như sau:

- ❖ Xác suất đạt dung lượng tại SU thứ k từ FBS thứ j :

$$p_j^{k,v_i} = Pr\{C_j^{k,v_i} \geq C_{th}^{v_i}\} = \exp\left(\frac{-\xi_j^{k,v_i} N_0}{u_j^{v_i} P_j}\right) \quad (3.6)$$

- ❖ Và xác suất đạt dung lượng tại SU thứ k từ MBS:

$$\begin{aligned} p_0^{k,v_i} &= Pr\{C_0^{k,v_i} \geq C_{th}^{v_i}\} \\ &= \exp\left\{-\xi_0^{k,v_i} \left[\lambda_C^{k,v_i} \left(\frac{P_C^k}{\prod_{j=1}^J (1-u_j^{v_i}) P_0} \right)^{\frac{2}{\eta}} \right]\right\} \end{aligned} \quad (3.7)$$

với

- $\xi_j^{k,v_i} = (d_j^k)^\eta \left(2^{\frac{C_{th}^{v_i}}{W}} - 1\right)$, $\xi_0^{k,v_i} = \pi(d_0^k)^2 \Gamma\left(1 + \frac{2}{\eta}\right) \Gamma\left(1 - \frac{2}{\eta}\right) \left(2^{\frac{C_{th}^{v_i}}{W}} - 1\right)^{2/\eta}$
- d_j^k và d_0^k là khoảng cách từ FBS thứ j và MBS đến SU thứ k
- $\lambda_C^{k,v_i} = \sum_{n=1}^N \sum_{m=1}^M w_{n,m}^k p_n^{v_i}$ là mật độ CU trong bán kính của tế bào mạng

Trong đó, công thức (3.6) và (3.7) có được tương ứng từ [97] và [96].

Như vậy, xác suất đạt dung lượng để truyền video thứ i với mức phân giải v_i từ FBS thứ j hoặc từ MBS đến SU thứ k được tính như sau:

$$p_{0,j}^{k,v_i} = 1 - (1 - p_j^{k,v_i})(1 - p_0^{k,v_i}) \quad (3.8)$$

3.3.2. Xác suất đạt dung lượng tại CU

Tương tự như tính dung lượng đạt được tại SU, nhưng do CU không bị ảnh hưởng bởi can nhiễu (vì nó không chia sẻ tài nguyên phổ tần kênh truyền cho MU khác), nên dung lượng tại CU thứ n từ FBS thứ j và MBS được tính lần lượt bằng:

$$C_{j,n}^{v_i} = W \log_2 \left(1 + \frac{u_j^{v_i} P_j G_j^n}{N_0} \right) \quad (3.9)$$

và

$$C_{0,n}^{v_i} = W \log_2 \left(1 + \frac{\prod_{j=1}^J (1 - u_j^{v_i}) P_0 G_0^n}{N_0} \right) \quad (3.10)$$

với G_j^n và G_0^n là độ lợi kênh truyền từ FBS thứ j và MBS tới CU thứ n .

Như vậy, ta có được xác suất đạt dung lượng tương ứng tại CU thứ n như sau:

❖ Từ FBS thứ j tới CU thứ n :

$$p_{j,n}^{v_i} = Pr\{C_{j,n}^{v_i} \geq C_{th}^{v_i}\} = \exp\left(\frac{-\xi_j^{n,v_i} N_0}{u_j^{v_i} P_j}\right) \quad (3.11)$$

❖ Và từ MBS tới CU thứ n :

$$p_{0,n}^{v_i} = Pr\{C_{0,n}^{v_i} \geq C_{th}^{v_i}\} = \exp\left(\frac{-\xi_0^{n,v_i} N_0}{\prod_{j=1}^J (1 - u_j^{v_i}) P_0}\right) \quad (3.12)$$

với

$$\bullet \xi_j^{n,v_i} = (d_j^n)^\eta \left(2^{\frac{C_{th}^{v_i}}{W}} - 1 \right) \text{ và } \xi_0^{n,v_i} = \pi (d_0^n)^\eta \left(2^{\frac{C_{th}^{v_i}}{W}} - 1 \right)$$

- d_j^n và d_0^n là khoảng cách từ FBS thứ j và MBS tới CU thứ n .

Cuối cùng, xác suất đạt dung lượng để truyền video thứ i với mức phân giải v_i từ FBS thứ j hoặc MBS đến CU thứ n là:

$$p_{0,j,n}^{v_i} = 1 - (1 - p_{j,n}^{v_i})(1 - p_{0,n}^{v_i}) \quad (3.13)$$

3.3.3. Xác suất đạt dung lượng tại NU

Khác với SU và CU, dung lượng đạt được tại NU thứ m , đến từ các CU thứ n , FBS thứ j hoặc MBS, được tính tương ứng là:

$$C_{n,m}^{k,v_i} = W \log_2 \left(1 + \frac{w_{n,m}^k p_n^{v_i} P_C^k G_n^m}{N_0 + P_0 G_0^m + I_{C,C}^{k,v_i}} \right), \quad (3.14)$$

$$C_{j,m}^{k,v_i} = W \log_2 \left(1 + \frac{u_j^{v_i} (1 - w_{n,m}^k p_n^{v_i}) P_j G_j^m}{N_0} \right) \quad (3.15)$$

và

$$C_{0,m}^{k,v_i} = W \log_2 \left(1 + \frac{\prod_{j=1}^J (1 - u_j^{v_i}) (1 - w_{n,m}^k p_n^{v_i}) P_0 G_0^m}{N_0} \right) \quad (3.16)$$

trong đó,

- G_n^m , G_0^m và G_j^m là độ lợi kênh từ CU thứ n , MBS, và FBS thứ j đến NU thứ m
- $I_{C,C}^{k,v_i} = \sum_{n' \neq n}^N \sum_{m' \neq m}^M w_{n',m'}^k p_{n'}^{v_i} P_C^k G_{n',m'}^k$ là can nhiễu từ các CU khác tác động lên CU thứ n khi chúng dùng chung tài nguyên phổ tần của SU thứ k .

Từ các dung lượng được tính toán trong (3.14), (3.15) và (3.16), ta có được các xác suất đạt dung lượng tương ứng là:

$$\begin{aligned}
p_{n,m}^{k,v_i} &= Pr\{C_{n,m}^{k,v_i} \geq C_{th}^{v_i}\} \\
&= exp\left\{-\xi_{n,m}^{v_i} \left[\lambda_M \left(\frac{P_0}{w_{n,m}^k p_n^{v_i} P_C^k} \right)^{\frac{2}{\eta}} + \lambda_C^{k,v_i} \right]\right\}
\end{aligned} \tag{3.17}$$

$$\begin{aligned}
p_{j,m}^{k,v_i} &= Pr\{C_{j,m}^{k,v_i} \geq C_{th}^{v_i}\} \\
&= exp\left[\frac{-\xi_{j,m}^{v_i} N_0}{u_j^{v_i} (1 - w_{n,m}^k p_n^{v_i}) P_j} \right]
\end{aligned} \tag{3.18}$$

và

$$\begin{aligned}
p_{0,m}^{k,v_i} &= Pr\{C_{0,m}^{k,v_i} \geq C_{th}^{v_i}\} \\
&= exp\left[\frac{-\xi_{0,m}^{v_i} N_0}{\prod_{j=1}^J (1 - u_j^{v_i}) (1 - w_{n,m}^k p_n^{v_i}) P_0} \right]
\end{aligned} \tag{3.19}$$

Trong đó, công thức (3.17) có được từ [96], với:

- $\xi_{j,m}^{v_i} = (d_j^m)^\eta \left(2^{\frac{C_{th}^{v_i}}{W}} - 1 \right)$. $\xi_{0,m}^{v_i} = (d_0^m)^\eta \left(2^{\frac{C_{th}^{v_i}}{W}} - 1 \right)$
- $\xi_{n,m}^{v_i} = \pi (d_n^m)^2 \Gamma\left(1 + \frac{2}{\eta}\right) \Gamma\left(1 - \frac{2}{\eta}\right) \left(2^{\frac{C_{th}^{v_i}}{W}} - 1 \right)^{2/\eta}$
- d_n^m, d_j^m và d_0^m là khoảng cách từ CU thứ n , FBS thứ j và MBS đến NU thứ m
- $\lambda_C^{k,v_i} = \sum_{n'=1}^N \sum_{m'=1}^M w_{n',m'}^k p_{n'}^{v_i}$ trong phạm vi bán kính của tế bào mạng
 $n' \neq n$ $m' \neq m$

Cuối cùng, xác suất đạt dung lượng tại NU thứ m được tính bởi:

$$p_{0,j,n,m}^{k,v_i} = 1 - (1 - p_{n,m}^{k,v_i})(1 - p_{j,m}^{k,v_i})(1 - p_{0,m}^{k,v_i}) \quad (3.20)$$

3.3.4. *Chất lượng trung bình của video nhận được*

Để đạt được chất lượng trung bình của các video với các mức phân giải khác nhau thu được, mô hình mối quan hệ giữa méo và tốc độ mã hóa (RD – Rate-Distortion) theo hàm số mũ [63] được áp dụng. Trong mô hình RD của một video, càng nhiều bit được dùng để mã hóa trong một giây (tốc độ mã hóa) thì video này được giải mã với chất lượng tốt hơn (méo thấp hơn). Với từng chuẩn mã hóa, mô hình đường cong RD của một video sẽ khác nhau đặc trưng cho tính hiệu quả của chuẩn mã hóa đó. Nghĩa là một chuẩn mã hóa sẽ tốt hơn nếu cùng một tốc độ mã hóa, khi giải mã sẽ cho ra video có méo thấp hơn. Đường cong này có được nhờ vào công cụ phân tích video thực nghiệm theo một chuẩn mã hóa video cụ thể. Mô hình đường cong RD của video được sử dụng nhiều trong các bài toán tối ưu tốc độ mã hóa nhằm đáp ứng với các điều kiện khác nhau của môi trường truyền dẫn. Một trong những ứng dụng được các nhà cung cấp nội dung quan tâm đó là cho phép nhiều phiên bản video có chất lượng khác nhau (được mã hóa ở tốc độ khác nhau) tồn tại nhằm phục vụ hiệu quả với từng ứng dụng/dịch vụ đến từng nhóm người dùng qua từng môi trường truyền dẫn.

Mô hình RD [63] cho phép biểu thị các đường cong theo thực nghiệm RD (experimental RD) của video được mã hóa theo tiêu chuẩn mã hóa video hiệu quả cao [98]. Trong mô hình này, nếu video thứ i với mức phân giải v_i mã hóa ở tốc độ (hoặc dung lượng) $C_{th}^{v_i}$, độ méo (distortion) tương ứng được tính bởi:

$$D_i(C_{th}^{v_i}) = \gamma_i (C_{th}^{v_i})^{\beta_i} \quad (3.21)$$

trong đó γ_i và β_i là các tham số phụ thuộc được chọn sao cho công thức (3.21) đáp ứng các đường cong RD thực nghiệm.

Dựa trên các xác suất đạt dung lượng đã trình bày ở (3.8), (3.13) và (3.20) trên, ta có thể tính toán các giá trị chất lượng trung bình của các video với các mức phân giải khác nhau nhận tại các MU (SU, CU, và NU) như sau:

$$\bar{Q} = \frac{\sum_{j=1}^J (Q_S^j + Q_C^j + Q_N^j)}{3J} \quad (3.22)$$

Với

$$Q_S^j = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^I \frac{r_i}{V_i} \sum_{v_i=1}^{V_i} p_{0,j}^{k,v_i} Q(D_i(C_{th}^{v_i})) \quad (3.23)$$

$$Q_C^j = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^I \frac{r_i}{V_i} \sum_{v_i=1}^{V_i} p_{0,j,n}^{v_i} Q(D_i(C_{th}^{v_i})) \quad (3.24)$$

Và

$$Q_N^j = \frac{1}{KMN} \sum_{k=1}^K \sum_{N=1}^N \sum_{m=1}^M \sum_{i=1}^I \frac{r_i}{V_i} \sum_{v_i=1}^{V_i} p_{0,j,n,m}^{k,v_i} Q(D_i(C_{th}^{v_i})) \quad (3.25)$$

trong đó $Q(D_i(C_{th}^{v_i})) = 10 \log_{10} \frac{255^2}{D_i(C_{th}^{v_i})}$ là tỉ số tín hiệu cực đại trên nhiễu (PSNR).

3.4. Bài toán tối ưu CRS và thuật giải di truyền

3.4.1. Bài toán tối ưu CRS

Để xây dựng bài toán tối ưu CRS, ta tính toán thêm dung lượng lưu trữ trên FBS (L_F^i) được sử dụng để lưu trữ video i với các mức phân giải khác nhau và tổng dung lượng (C) mà hệ thống cung cấp cho các MU và xem xét chúng như các ràng buộc trong quá trình giải bài toán tối ưu CRS. Cụ thể, L_F^i và C tương ứng được tính bởi:

$$L_F^i = \sum_{j=1}^J \sum_{v_i=1}^{V_i} u_j^{v_i} L_i^{v_i} \quad (3.26)$$

và

$$C = C_S + C_C + C_N \quad (3.27)$$

trong đó

$$C_S = \sum_{j=1}^J \sum_{k=1}^K \sum_{i=1}^I \frac{r_i}{V_i} \sum_{v_i=1}^{V_i} (C_j^{k,v_i} + C_0^{k,v_i}) \quad (3.28)$$

$$C_C = \sum_{j=1}^J \sum_{n=1}^N \sum_{i=1}^I \frac{r_i}{V_i} \sum_{v_i=1}^{V_i} (C_{j,n}^{v_i} + C_{0,n}^{v_i}) \quad (3.29)$$

và

$$C_N = \sum_{j=1}^J \sum_{k=1}^K \sum_{N=1}^N \sum_{m=1}^M \sum_{i=1}^I \frac{r_i}{V_i} \sum_{v_i=1}^{V_i} (C_{n,m}^{k,v_i} + C_{j,m}^{k,v_i} + C_{0,m}^{k,v_i}) \quad (3.30)$$

Ta đặt

- L_{max}^i là giới hạn trên về dung lượng bộ nhớ của các FBS để lưu trữ video thứ i với các mức phân giải khác nhau;
- C^* là giới hạn trên về tổng dung lượng được yêu cầu bởi các MU;
- γ_0 là giới hạn dưới tỉ số tín hiệu trên can nhiễu và nhiễu trắng (SINR) tại SU khi SU đồng ý chia sẻ tài nguyên phổ tần kênh truyền xuống.

Cuối cùng, bài toán tối ưu CRS được mô tả như sau:

$$\max_{u_j^{v_i}, w_{n,m}^k} \bar{Q} \quad (3.31)$$

$$s. t. \begin{cases} \sum_{v_i=1}^{V_i} u_j^{v_i} \leq 1, i = 1, 2, \dots, I, j = 1, 2, \dots, J, \\ \sum_{m=1}^M w_{n,m}^k \leq 1, k = 1, 2, \dots, K, n = 1, 2, \dots, N, \\ \sum_{n=1}^N w_{n,m}^k \leq 1, k = 1, 2, \dots, K, m = 1, 2, \dots, M, \\ \sum_{k=1}^K w_{n,m}^k \leq 1, n = 1, 2, \dots, N, m = 1, 2, \dots, M, \\ L_F^i \leq \mu L_{max}^i, i = 1, 2, \dots, I, \\ C \leq \delta C^* \\ I_{C,S}^{k,v_i} \leq \frac{P_0 G_0^k}{\gamma_0} - N_0, k = 1, 2, \dots, K, i = 1, 2, \dots, I, v_i = 1, 2, \dots, V_i \end{cases} \quad (3.32)$$

Trong đó:

- \bar{Q} (3.31) được tính theo (3.22), (3.23), (3.24) và (3.25).
- Ràng buộc đầu tiên trong (3.32) là giới hạn mỗi video thứ i chỉ được lưu trữ 1 phiên bản với mức phân giải thích hợp tại FBS thứ j .
- Bộ 3 ràng buộc tiếp theo cho phép băng tần của SU thứ k được chia sẻ cho nhiều cặp CU-NU (n, m), nhưng không cho phép một CU phát đến nhiều NU cũng như một NU nhận từ nhiều CU.
- Các ràng buộc thứ tư và thứ năm là điều chỉnh L_{max}^i và C^* để thuận tiện cho việc đánh giá, $L_{max}^i = r_i \times I \times J \times \max\{L_i^{v_i}, v_i = 1, 2, \dots, V_i\}$, $0 < \mu \leq 1$ và $0 < \delta \leq 1$.
- Ràng buộc cuối cùng, xuất phát từ (3.2) bằng cách đặt $\frac{P_0 G_0^k}{N_0 + I_{C,S}^{k,v_i}} \geq \gamma_0$, nhằm hạn chế ảnh hưởng can nhiễu của truyền thông D2D giữa CU và NU lên SU k .

3.4.2. Thuật giải di truyền cho bài toán CRS

Để giải bài toán tối ưu CRS, trong chương này, một số thư viện trong bộ công cụ “Thuật giải di truyền trong Matlab” [62] được sử dụng. Tuy nhiên, công cụ này có hạn chế là chỉ có thể giải các bài toán có ràng buộc cận trên và cận dưới đơn giản của biến chứ không thể giải các bài toán có ràng buộc phức tạp như ràng buộc trong bài toán CRS nêu trên. Do đó, giải pháp là chuyển bài toán tối ưu có ràng buộc (3.32) thành bài toán tối ưu không có ràng buộc (3.35) bằng phương pháp hàm phạt (penalty function) [99-102], cụ thể như sau:

Đầu tiên, các ràng buộc trong (3.32) được chuyển về, ta có:

$$\left\{ \begin{array}{l} \Delta u_{i,j} = 1 - \sum_{v_i=1}^{V_i} u_j^{v_i} \geq 0, i = 1,2, \dots, I, j = 1,2, \dots, J, \\ \Delta w_{k,n} = 1 - \sum_{m=1}^M w_{n,m}^k \geq 0, k = 1,2, \dots, K, n = 1,2, \dots, N, \\ \Delta w_{k,m} = 1 - \sum_{n=1}^N w_{n,m}^k \geq 0, k = 1,2, \dots, K, m = 1,2, \dots, M \\ \Delta w_{n,m} = 1 - \sum_{k=1}^K w_{n,m}^k \geq 0, n = 1,2, \dots, N, m = 1,2, \dots, M \\ \Delta L_F^i = \mu L_{\max}^i - L_F^i \geq 0, i = 1,2, \dots, I, \\ \Delta C = \delta C^* - C \geq 0, \\ \Delta I_{C,S}^{k,i,v_i} = \frac{P_0 G_0^k}{\gamma_0} - N_0 - I_{C,S}^{k,v_i} \geq 0, k = 1,2, \dots, K, i = 1,2, \dots, I, v_i = 1,2, \dots, V_i. \end{array} \right. \quad (3.33)$$

Từ các thành phần trong (3.33), hàm phạt (F) được tính như sau:

$$\begin{aligned} F = & \lambda_1 \sum_{i=1}^I \sum_{j=1}^J (\min\{0, \Delta u_{i,j}\})^2 \\ & + \lambda_2 \sum_{k=1}^K \sum_{n=1}^N (\min\{0, \Delta w_{k,n}\})^2 \\ & + \lambda_3 \sum_{k=1}^K \sum_{m=1}^M (\min\{0, \Delta w_{k,m}\})^2 \\ & + \lambda_4 \sum_{n=1}^N \sum_{m=1}^M (\min\{0, \Delta w_{n,m}\})^2 \\ & + \lambda_5 \sum_{i=1}^I (\min\{0, \Delta L_F^i\})^2 \\ & + \lambda_6 (\min\{0, \Delta C\})^2 \\ & + \lambda_7 \sum_{k=1}^K \sum_{i=1}^I \sum_{v_i=1}^{V_i} (\min\{0, \Delta I_{C,S}^{k,i,v_i}\})^2 \end{aligned} \quad (3.34)$$

Trong đó, các hệ số $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7$ đặc trưng cho mức độ vi phạm ràng buộc. Các hệ số này được sử dụng để trừng phạt các cá thể vi phạm ràng buộc trong thế hệ hiện tại. Hình phạt được thực hiện bằng cách giảm hàm mục tiêu \bar{Q} (3.31) một lượng F (3.34) sao cho nhiều cá thể hơn đáp ứng các ràng buộc trong thế hệ tiếp theo cho đến khi các cá thể đều thỏa mãn các ràng buộc, nghĩa là $F = 0$ (ngưng phạt).

Như vậy, thay vì cực đại hàm mục tiêu với nhiều ràng buộc (3.32), thông qua các bước chuyển đổi (3.33) và (3.34), ta có được bài toán tối ưu không ràng buộc có thể giải được bằng thuật toán di truyền như sau:

$$\max_{u_j^{v_i}, w_{n,m}^k} \bar{Q}_F = \bar{Q} - F. \quad (3.35)$$

Việc cực đại (3.35) sẽ tương đương với cực đại (3.31) và thỏa mãn các ràng buộc (3.32) bởi vì khi cực đại (3.35) thì \bar{Q}_F sẽ tiến đến \bar{Q} và F sẽ tiến đến 0 như được mô tả trong (3.33) và (3.34) để thỏa các ràng buộc.

Để giải (3.35), việc tìm $u_{j,i}^{v_i}$ và $w_{n,m}^k$ để cực đại \bar{Q}_F sẽ tương đương với việc tìm các hoán vị của J ma trận ($\mathcal{U} = \{U_{IV_i}^1, U_{IV_i}^2, \dots, U_{IV_i}^j, \dots, U_{IV_i}^J\}$) và K ma trận ($\mathcal{W} = \{W_{NM}^1, W_{NM}^2, \dots, W_{NM}^k, \dots, W_{NM}^K\}$), trong đó $U_{IV_i}^j$ là ma trận nhị phân I hàng V_i cột và W_{NM}^k là ma trận nhị phân N hàng M cột. Ví dụ, một phần tử ở hàng thứ i ($i = 1, 2, \dots, I$) và cột thứ v_i ($v_i = 1, 2, \dots, V_i$) của ma trận $U_{IV_i}^j$ có giá trị bằng 1 (hoặc 0) có nghĩa là video thứ i ở mức phân giải v_i được (hoặc không) lưu trữ tại FBS thứ j .

Để tìm giá trị cực đại \bar{Q}_F trong (3.35) bằng thuật toán di truyền, được trình bày trong Thuật giải 3.1, nghĩa là phải xác định: 1) video nào trong I video và phiên bản nào trong V_i phiên bản của video thứ i được lưu ở FBS nào trong J FBS; và 2) SU nào trong K SU chia sẻ tài nguyên kênh truyền cho cặp CU và NU nào trong N CU và M NU. Để làm được điều này, ta sẽ khởi tạo quần thể ban đầu gồm các cá thể, mỗi nhiễm sắc thể (được biểu diễn dưới dạng chuỗi nhị phân) của từng cá thể chính là mỗi ứng viên nghiệm cho bài toán tối ưu trên, mỗi ứng viên nghiệm này phải có chiều dài $N_B = J \times I \times V_i + K \times N \times M$ bit (bằng tổng số phần tử trong ma trận $U_{IV_i}^j$ và ma trận W_{NM}^k).

Quần thể ban đầu được khởi tạo bao gồm N_{PG} cá thể được chọn ngẫu nhiên từ tổng số 2^{N_b} cá thể. Tùy thuộc vào độ dài của nhiễm sắc thể mà ta lựa chọn số quần thể N_{PG} sao cho phù hợp. Hơn nữa, trong quá trình triển khai thuật giải di truyền, giá trị \bar{Q}_F (3.35) được tính trực tiếp thông qua các biến số gồm các giá trị 0 và 1 trong nhiễm sắc thể. Do đó, trong tính toán này sẽ bỏ qua bước chuyển đổi giá trị nhị phân sang giá trị thực mà thuật giải di truyền thông thường phải thực hiện. Thuật giải di truyền cho bài toán CRS được mô tả như sau:

Thuật giải 3.1. Thuật giải di truyền

- 1 **Input:** Tham số hệ thống được liệt kê ở Bảng 3-2
 $N_P = 20.000$; số cá thể trong quần thể
 $PREC = J * I * V_i + K * N * M$: chiều dài của nhiễm sắc thể.
 $P_G = 0,9$: khoảng cách giữa thế hệ
 $P_C = 0,9$: xác suất lai ghép
 $P_m = 10^{-6}$: xác suất đột biến
 Khởi tạo thế hệ đầu tiên gồm tập N_P cá thể ngẫu nhiên $\{X_z\}$, $z = 1, 2, \dots, N_P$
 $Gen = 0$: số thế hệ
 TC : điều kiện dừng vòng lặp
 - 2 **Output:** X^* và \bar{Q}_F^*
 - 3 **While** TC chưa thỏa **do**
 - 4 $Gen = Gen + 1$
 - 6 Đưa $\{X_z\}$ và $\bar{Q}_F(X_z)$ vào nhóm sẵn sàng lai ghép để xếp hạng.
 - 6 Chọn $N_{PG} \times P_G$ cá thể tốt nhất trong tập $\{X_z\}$ có $\bar{Q}_F(X_z)$ cao để chuẩn bị lai ghép cho ra thế hệ kế tiếp.
 - 7 Chọn các cặp cha mẹ theo phương pháp lai ghép 2 điểm (Double Point Crossover) để lai ghép ra thế hệ con kế tiếp. Việc lai ghép không áp dụng với tất cả các các cặp cha mẹ mà chỉ áp dụng với xác suất P_C .
 - 8 Thực hiện đột biến theo phương pháp bù nhị phân (Inversion Mutation) ở thế hệ con được tạo ra với xác suất đột biến là P_M .
 - 9 Đánh giá giá trị thích nghi (có được sau khi đã thay thế các giá trị chuỗi nhị phân của từng cá thể vào hàm thích nghi) cho các thế hệ con được tạo ra sẵn sàng cho các bước lai ghép tiếp theo.
 - 10 **End While**
 - 11 Chọn cá thể X^* tương ứng với giá trị cực đại \bar{Q}_F^* trong thế hệ cuối cùng.
-

Trong Thuật giải 3.1, TC là điều kiện dừng của thuật giải di truyền (vòng lặp While). Thuật giải sẽ dừng khi TC thỏa 1 trong 3 điều kiện sau: 1) Khi số lần lặp $Gen > 20$ và giá trị $Error = 0$ (hàm phạt F) và Giá trị cực đại \bar{Q}_F^* không đổi trong 10 thế hệ liên tiếp; 2) Khi số lần lặp $Gen > 150$ và $Error < 10^{-3}$ và Giá trị cực đại \bar{Q}_F^* không đổi trong 30 thế hệ liên tiếp; 3) Khi số lần lặp $Gen > 200$.

Ngoài ra, $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6$ và λ_7 được thiết lập ban đầu bằng 10^{-3} để kiểm tra kết quả của hàm phạt. Nếu kết quả của hàm phạt (3.34) là lớn và tất cả các phần tử trong hàm phạt khác 0, thì sẽ tăng giá trị các λ lên 10 lần. Giá trị của một λ nào đó sẽ không đổi nếu phần tử tương ứng của nó bằng không. Quá trình được lặp lại cho đến khi hàm phạt hội tụ về 0 hoặc thỏa điều kiện hội tụ cho trước.

3.5. Đánh giá hiệu suất cơ chế CRS và thuật giải di truyền

3.5.1. Thiết lập thông số hệ thống và cấu hình máy tính

Để đánh giá hiệu suất cơ chế CRS và hiệu quả thuật giải di truyền, mô hình hệ thống được triển khai với các tham số được liệt kê trong Bảng 3-2. Giả sử MBS bao phủ 1 vùng có bán kính cực đại là 1500m và khoảng cách giữa MBS và MU, FBS và MU, CU và SU, CU và NU được phân phối đều ngẫu nhiên trong các phạm vi lần lượt tương ứng là [500, 1500]m, [20, 100]m, [50, 150]m, [1, 20]m. Ngoài ra, 3 video được xem xét gồm: $V_1^{v_1}$ (Basketballpass), $V_2^{v_2}$ (Racehorses) và $V_3^{v_3}$ (Foreman). Các video này được phân tích để thu được các đường cong RD thực nghiệm của chúng bằng cách sử dụng phần mềm HM phiên bản 12.0 [103], từ đó ta có thể mã hóa mỗi video với 3 mức phân giải khác nhau ($V_i = 3$) cũng có được các giá trị $L_i^{v_i}$, $C_{th}^{v_i}$, γ_i và β_i như trong Bảng 3-2.

Cơ chế CRS được đánh giá bằng cách so sánh với 3 cơ chế khác bao gồm cơ chế chỉ lưu trữ (OCC – Only Caching), cơ chế chỉ chia sẻ tài nguyên phổ tần kênh truyền xuống (ORS – Only Downlink Resource Sharing) và cơ chế không lưu trữ cũng như không chia sẻ tài nguyên phổ tần kênh truyền xuống (NCS – No Caching Nor Downlink Resource Sharing). Một máy tính thông thường với các thông số chính

được mô tả trong Bảng 3-3 được sử dụng để triển khai tính toán, mô phỏng và đánh giá kết quả.

Bảng 3-2. Các tham số đầu vào cho bài toán CRS

Ký hiệu	Trị số
I, J, K, N, M	5 video, 5 FBS, 5 SU, 5 CU và 10 NU
V_i	3 mức phân giải của mỗi video
W, P_M^0, P_F^j, P_T^n	5MHz; 5W; 1W; 0,1W
η, γ_0, N_0	4; 10dB; 10^{-12} W
$\{a, b\}$	{0,5; 0,5}
r_n	1, nghĩa là CU thứ n còn trống 100% dung lượng để lưu video (nhằm tăng ưu thế cho truyền thông D2D).
$L_i^{v_i}$	[11867 23734 35600 198680 264906 351000 33382 66763 113496 1172340 1758510 2344680 160410 320820 453960] Kbit
$C_{th}^{v_i}$	[1000 2000 3000 3000 4000 5300 50000 100000 170000 10000 15000 20000 10000 20000 28300]Kbps
γ_i	[9806 76520 1644000]
β_i	[-0,9972 -1,1530 -1,0920]
δ, μ	1; 0,5
C^*	10 Gbps, nghĩa là mỗi MU được phục vụ với dung lượng trung bình 0,5 Gbps
N_P	20.000
N_B	$J \times I \times V_i + K \times N \times M$
P_G, P_M, P_C	0,9; 10^{-6} ; 0,9
$\lambda_1, \lambda_2, \lambda_3, \lambda_4$ $\lambda_5, \lambda_6, \lambda_7$	1; 1; 1 10^3 ; 10^3 ; 10^{25}

Bảng 3-3. Cấu hình máy tính

Thông số cấu hình máy tính	
Processor	Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz
Processor type	x64 Family 6 Model 14 Stepping 10, GenuineIntel
PHY processor packages	1
Processor cores	6
Logical processors	12
Total PHY memory	16.599.444KB
Operating System	Windows 10

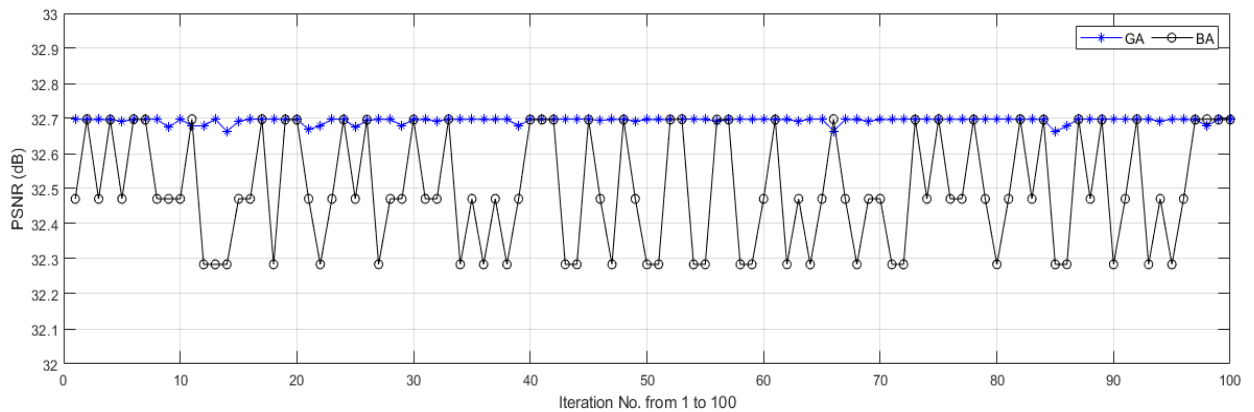
3.5.2. Đánh giá hiệu quả của thuật giải di truyền

Để đánh giá hiệu quả của thuật giải di truyền (GA), thuật giải đàn dơi nhĩ phân (BBA) và thuật toán vét cạn (EA) đồng thời được triển khai để làm cơ sở so sánh về độ chính xác và thời gian thực thi cũng như làm cơ sở để chọn lựa các thông số cho thuật giải di truyền khi giải bài toán CRS. BBA được trình bày trong Chương 1, các thông số ($L_i^0 = 0,25$; $E_i^0 = 0,1$; $f_{min} = 0$; $f_{max} = 2$; $\varepsilon = 0,9$; $\zeta = 0,9$) để triển khai BBA được tham khảo từ [61]. Tương tự như trong thuật toán GA, phương pháp hàm phạt cũng được áp dụng để triển khai thuật toán BBA. Mặc dù cùng một bộ dữ liệu của hệ thống, nhưng do sự không đồng nhất về các toán tử và các thông số được sử dụng trong GA và BBA, trong luận án này, GA và BBA được đánh giá dựa trên: 1) sự tương đương về thời gian thực thi, ví dụ: $N_{PG} = 20.000$ trong GA với thời gian thực thi trung bình là 1003,68 giây và $N_{PB} = 10.000$ trong BBA với thời gian thực thi trung bình là 957,12 giây sẽ được chọn để đánh giá GA và BBA, 2) độ chính xác so với EA và 3) độ ổn định. EA được triển khai tương tự như trong Chương 2. Trong trường hợp này, các tham số kích thước hệ thống gồm (J , N , M và K) được thiết lập với những giá trị thấp tương ứng là (3, 3, 5 và 3) để thuật giải vét cạn thực thi được. Bài toán CRS sẽ được giải bằng cách tìm giá trị cực đại của chất lượng trung bình (được đo bằng dB) theo hệ số mũ phản ánh độ nổi tiếng của video α [88].

Kết quả được trình bày trong Bảng 3-4 cho thấy, độ chính xác khi sử dụng GA có thể đạt từ 99.97 đến 100% so với EA. Trong trường hợp này, độ chính xác của GA gần như tuyệt đối vì kích thước hệ thống tương đối thấp, trong khi đó số lượng cá thể được chọn tương đối cao ($N_{PG} = 20.000$), nhưng vẫn đảm bảo thời gian thực thi trung bình của GA (1003,68 giây) thấp hơn nhiều so với EA (5664,45 giây). Ngoài ra, so với BBA trong giới hạn thời gian thực thi tương đương nhau, GA cho kết quả chính xác hơn. Một vấn đề quan trọng nữa đó là, khi thực thi giải thuật 100 lần (với $\alpha = 1$), GA cho kết quả ổn định hơn BBA như được mô tả trong Hình 3-3. Từ những kết quả này, GA được chọn để áp dụng cho việc giải bài toán CRS với kích thước hệ thống lớn hơn.

Bảng 3-4. So sánh độ chính xác giữa GA, BBA và EA

α	0	0,5	1,0	1,5	2,0	2,5	3,0
EA (dB)	29,0286	31,2176	32,6980	34,1178	36,0832	37,2044	38,0584
BBA (dB)	28,9100	31,2176	32,4708	33,6306	36,0832	36,5090	37,6689
Accuracy	99,59%	100,00%	99,31%	98,57%	100,00%	98,13%	98,98%
GA (dB)	29,0286	31,2176	32,6924	34,1178	36,0716	37,1935	38,0476
Accuracy	100,00%	100,00%	99,98%	100,00%	99,97%	99,97%	99,97%



Hình 3-3. Đánh giá độ ổn định của GA và BBA

Ngoài ra, để có cái nhìn về việc đánh đổi giữa thời gian thực thi và độ chính xác của thuật giải di truyền, tôi đã so sánh bằng cách thay đổi số lượng cá thể N_{PG}

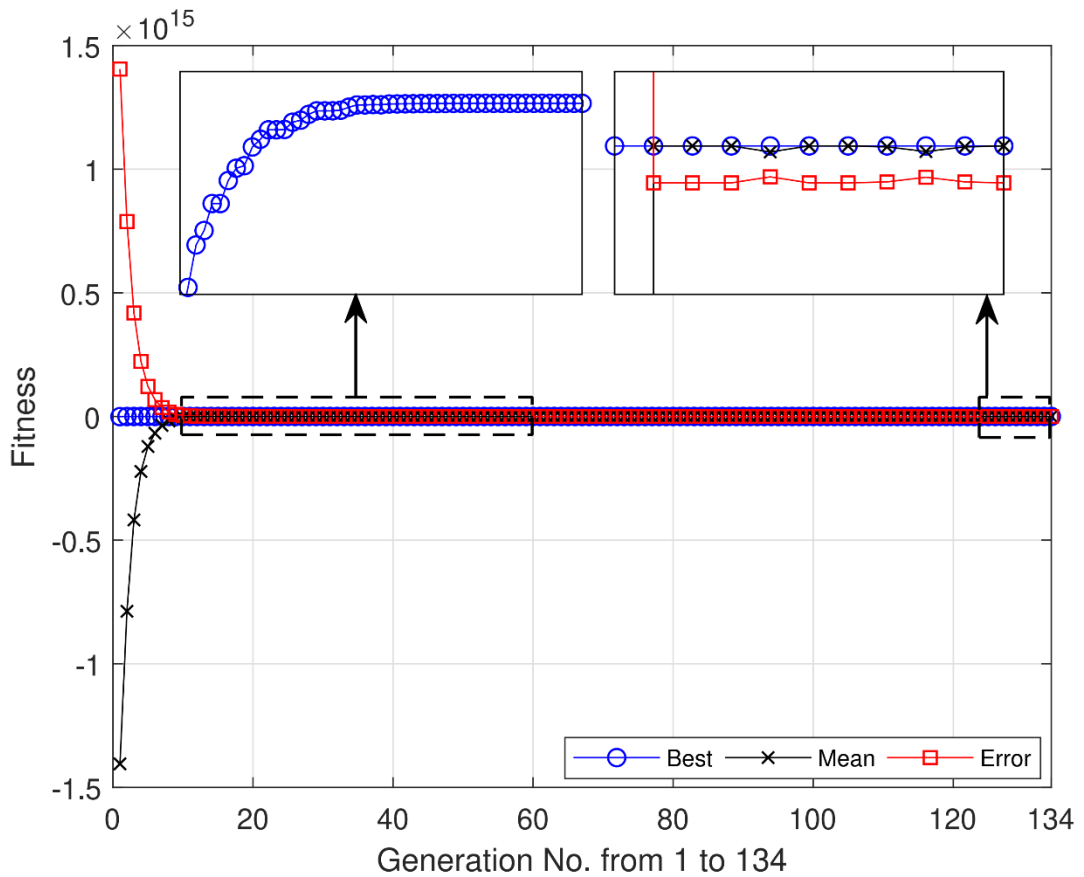
trong khi vẫn giữ $\alpha = 1$. Cách thức thực hiện so sánh như sau: với mỗi giá trị N_{PG} ($N_{PG} = 1.000; 5.000; 10.000; 15.000; 20.000$), thực thi 100 lần để tính PSNR và chọn kết quả của lần thực thi có độ chính xác thấp nhất. Kết quả từ Bảng 3-5 cho thấy, với mỗi số lượng cá thể N_{PG} khác nhau, ta có:

- Khi chấp nhận độ chính xác thấp là 97,34%, thì thời gian thực thi chỉ là 19.58 giây với số lượng cá thể $N_{PG} = 1.000$.
- Khi tăng số lượng cá thể lên, thì độ chính xác cũng tăng lên và thời gian thực thi cũng lâu hơn.

Bảng 3-5. Thực thi GA theo số N_{PG} với kết quả xấu nhất

Metrics	$N_{PG} (\alpha = 1)$					EA
	1.000	5.000	10.000	15.000	20.000	
Time (s)	19,58	142,50	317,42	556,01	795,63	5.664,45
PSNR (dB)	30,8283	32,2547	32,4474	32,6486	32,6631	32,6980
Accuracy	97,34%	98,64%	99,23%	99,85%	99,89%	100%

Tuy nhiên, khi càng tăng số lượng cá thể N_{PG} lên thì độ chính xác sẽ tăng lên rất ít nhưng thời gian thực thi tăng lên rất nhiều. Ở Bảng 3-5 thì khi so sánh giữa $N_{PG} = 1.000$ và $N_{PG} = 10.000$, thì độ chính xác từ 97,34% lên 99,23% (tăng 1.89%), nhưng khi tăng tiếp số lượng cá thể từ $N_{PG} = 10.000$ lên $N_{PG} = 20.000$, thì độ chính xác chỉ tăng từ 99,23% lên 99,89% (tăng chỉ 0,66%). Như vậy, ở $N_{PG} = 10.000$ thì thuật giải di truyền cho kết quả chấp nhận được với mức độ chính xác là 99,23% và thời gian thực thi là 317,42 giây so với 5.664,45 giây của thuật giải vét cạn. Dựa trên kết quả từ Bảng 3-5, tôi chọn $N_{PG} = 20.000$ để thực thi thuật giải di truyền với các tham số kích thước hệ thống (J , N , M và K) lớn hơn như trong Bảng 3-2 để đảm bảo độ chính xác tốt nhất với thời gian thực thi chấp nhận được.



Hình 3-4. Mức độ hội tụ của thuật giải di truyền

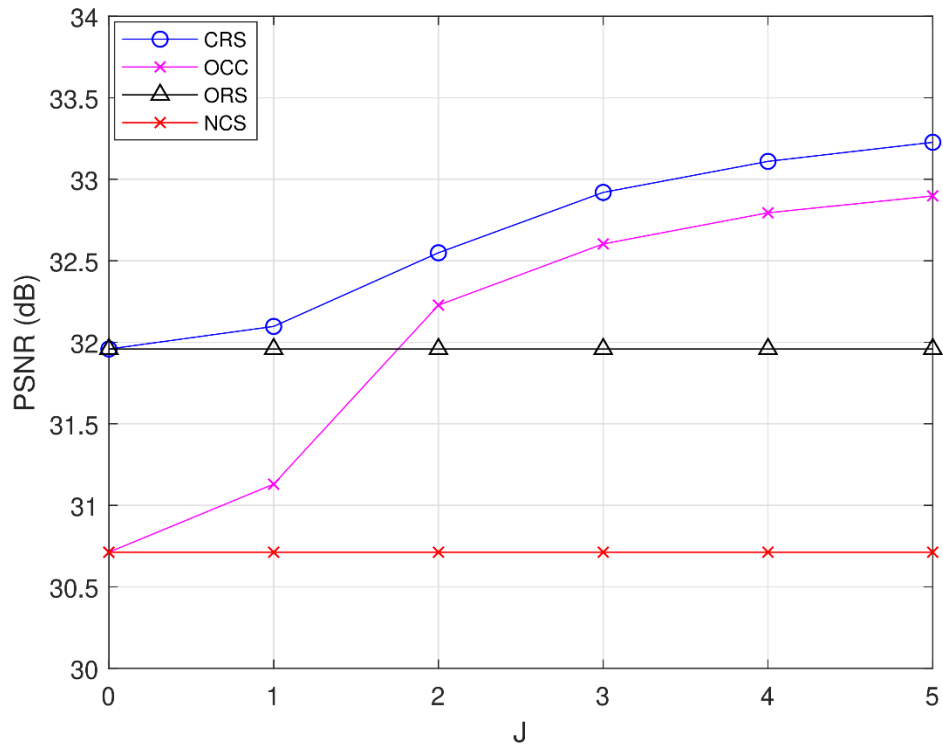
Để đánh giá kết quả thực hiện thuật giải di truyền, thuật giải di truyền được thực thi với các tiêu chí hội tụ (hay tiêu chí để dừng thuật giải) được trình bày trong Thuật giải 3-1. Kết quả thực thi theo số thế hệ được mô tả trong Hình 3-4, trong đó, “Best” là giá trị $\bar{Q}_F(X_Z)$ cao nhất có được từ cá thể tốt nhất trong từng thế hệ, “Mean” là giá trị trung bình của $\bar{Q}_F(X_Z)$ được tính cho tất cả các cá thể trong từng thế hệ, và “Error” chính là giá trị của hàm phạt F trong (3.34) và (3.35). Kết quả cho thấy, thuật giải di truyền bắt đầu hội tụ sau khoảng 20 thế hệ và đến thế hệ thứ 125 thì đáp ứng được tiêu chí hội tụ. Khi đó, “Error” tiến về 0 ($F = 0$) đồng nghĩa với các ràng buộc đã thỏa; giá trị “Best” không đổi và trùng với “Mean” đồng nghĩa với việc các cá thể đều tốt như nhau ở những thế hệ cuối cùng. Như vậy, từ các kết quả được phân tích trong Bảng 3-4, Bảng 3-5. Hình 3-3, và Hình 3-4, thuật giải di truyền cho thấy tính linh động và khả thi để giải bài toán tối ưu CRS trong phạm vi lớn hơn của 5G UDN.

3.5.3. *Đánh giá hiệu suất cơ chế CRS*

Cơ chế CRS được đánh giá dựa trên tiêu chí về chất lượng video thu được (tính bằng dB) theo các thông số đầu vào của hệ thống như: số lượng FBS, số lượng SU, độ lệch về sự nổi tiếng giữa các video và các ràng buộc. Trong mỗi đánh giá, cơ chế SCS được so sánh với 3 cơ chế OCC (Only Caching), ORS (Only Resource Sharing) và NCS (None Caching nor Sharing) cụ thể như sau:

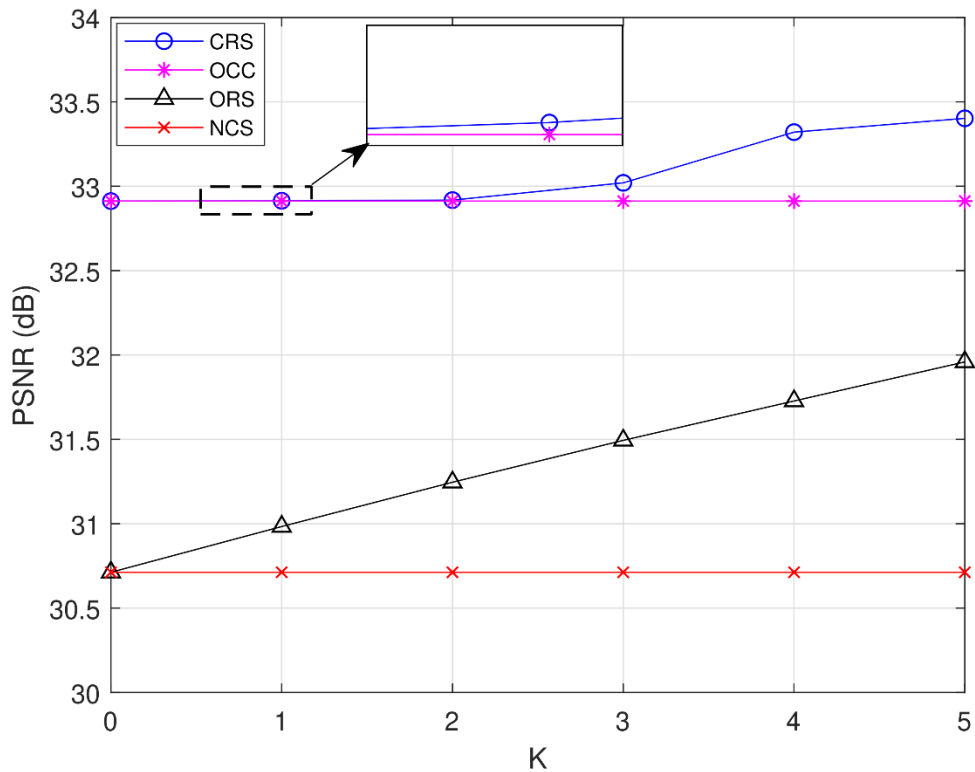
- Chỉ có lưu trữ (OCC – Only Caching): Đây là trường hợp hệ thống chỉ thực hiện tối ưu lưu trữ mà không thực hiện tối ưu chia sẻ tài nguyên kênh truyền xuống.
- Chỉ có chia sẻ phổ tần kênh truyền xuống (ORS – Only Resource Sharing): Ngược lại với cơ chế OCC, đây là trường hợp hệ thống chỉ thực hiện tối ưu chia sẻ tài nguyên kênh truyền xuống mà không thực hiện tối ưu lưu trữ.
- Không có lưu trữ cũng không có chia sẻ (NCS – None Caching nor Sharing): Đây là trường hợp hệ thống không thực hiện tối ưu lưu trữ cũng như tối ưu chia sẻ tài nguyên kênh truyền xuống. Trong trường hợp này, MU chỉ được phục vụ bởi MBS.

Đầu tiên, CRS được đánh giá so với 3 trường hợp khác là: OCC, ORS và NCS khi thay đổi số lượng FBS Hình 3-5. Nếu không có FBS để lưu trữ video ($J=0$), thì cơ chế OCC trở thành cơ chế NCS và cả 2 đều cho chất lượng thấp nhất vì không có lưu trữ cũng không có chia sẻ. Đồng thời, cơ chế CRS không có lưu trữ thì trở thành cơ chế ORS, nhưng vẫn cho chất lượng cao hơn so với OCC và NCS vì CRS có được sự chia sẻ tài nguyên phổ tần kênh truyền xuống. Khi tăng số lượng FBS, thì chất lượng của ORS và NCS không tăng lên, trong khi CRS và OCC tăng lên theo số lượng của FBS. Tuy nhiên, CRS và OCC sẽ có khuynh hướng bão hòa khi tiếp tục tăng số lượng FBS. Điều này cho thấy rằng, với một lượng người dùng nhất định, ta không cần thiết phải lưu tất cả các video trong nhiều FBS. Nói một cách khác, số lượng FBS để lưu trữ các video phải được tính toán cẩn thận để đảm bảo dung lượng lưu trữ hợp lý cũng như để giảm mức độ phức tạp trong tính toán.



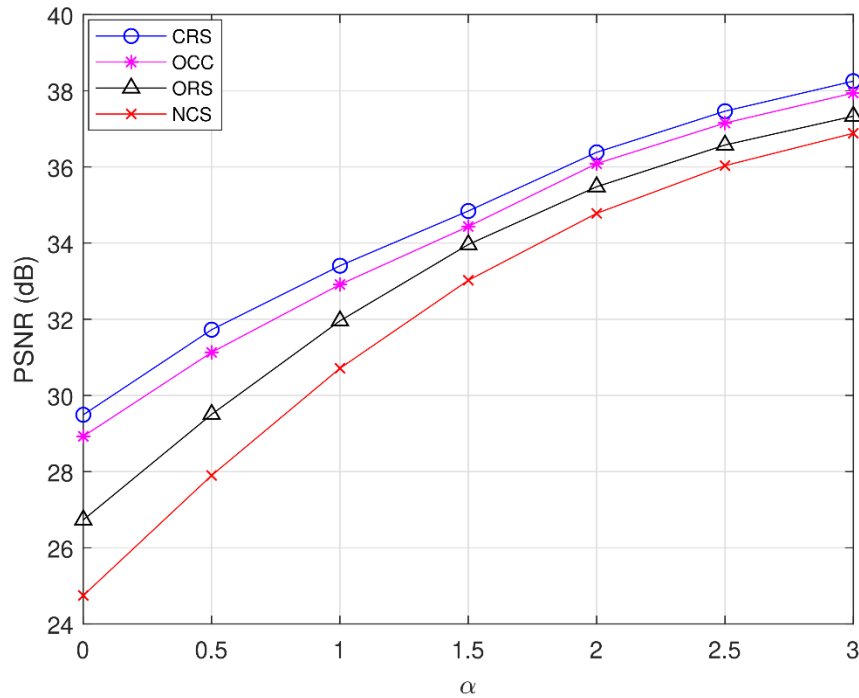
Hình 3-5. Chất lượng hệ thống (HT) của CRS, OCC, ORS, NCS theo số lượng FBS

Tiếp theo, cơ chế CRS được so sánh với các cơ chế OCC, ORS và NCS theo số lượng SU đồng ý chia sẻ tài nguyên (Hình 3-6). Tương tự như với kết quả đánh giá theo số lượng FBS, nếu không có SU nào chia sẻ tài nguyên kênh truyền, thì cơ chế CRS trở thành OCC và cơ chế ORS trở thành NCS. Khi tăng số lượng SU lên đồng nghĩa với việc mở rộng chọn lựa chia sẻ tài nguyên kênh truyền thì chất lượng của CRS và ORS cũng tăng lên theo, trong khi OCC và NCS không thay đổi. Hai cơ chế CRS và ORS sẽ bảo hòa khi tiếp tục tăng số lượng SU vì khi đó các SU được thêm vào không thể tốt hơn các SU trước đó trong quá trình tìm kiếm bộ 3 (SU, CU, NU) tốt nhất.



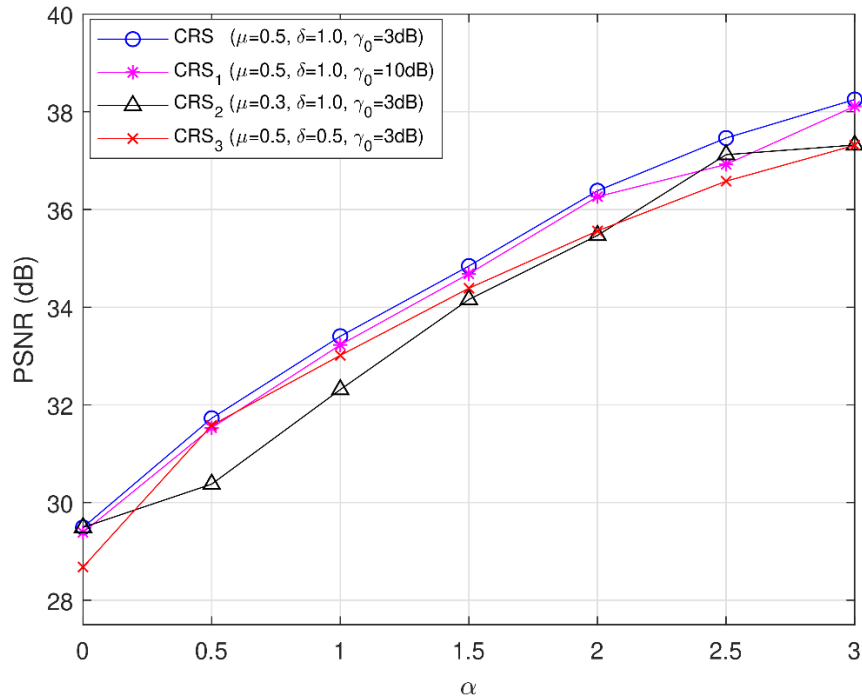
Hình 3-6. Chất lượng HT của CRS, OCC, ORS, NCS theo số lượng SU

Hình 3-7 mô tả kết quả đánh giá hiệu suất giữa CRS, OCC, ORS và NCS theo hệ số độ lệch nổi tiếng của video α . Kết quả chỉ ra rằng hệ thống đạt hiệu suất cao hơn nếu nó tập trung vào việc phục vụ MU các video có mức độ phổ biến cao, tức là giá trị của α cao hơn. Và kết quả đánh giá cũng cho thấy rằng, cơ chế đề xuất CRS vượt trội hơn so với cơ chế OCC, ORS và đương nhiên tốt hơn nhiều so với cơ chế NCS. Và trường hợp OCC tốt hơn trường hợp ORS vì nó cung cấp nhiều khả năng lưu trữ và truyền qua các kênh tốt hơn từ FBS. Trường hợp xấu nhất là NCS vì nó không có sự hỗ trợ của cơ chế lưu trữ và chia sẻ tài nguyên phổ tần kênh truyền xuống.



Hình 3-7. Chất lượng HT của CRS, OCC, ORS, NCS theo α

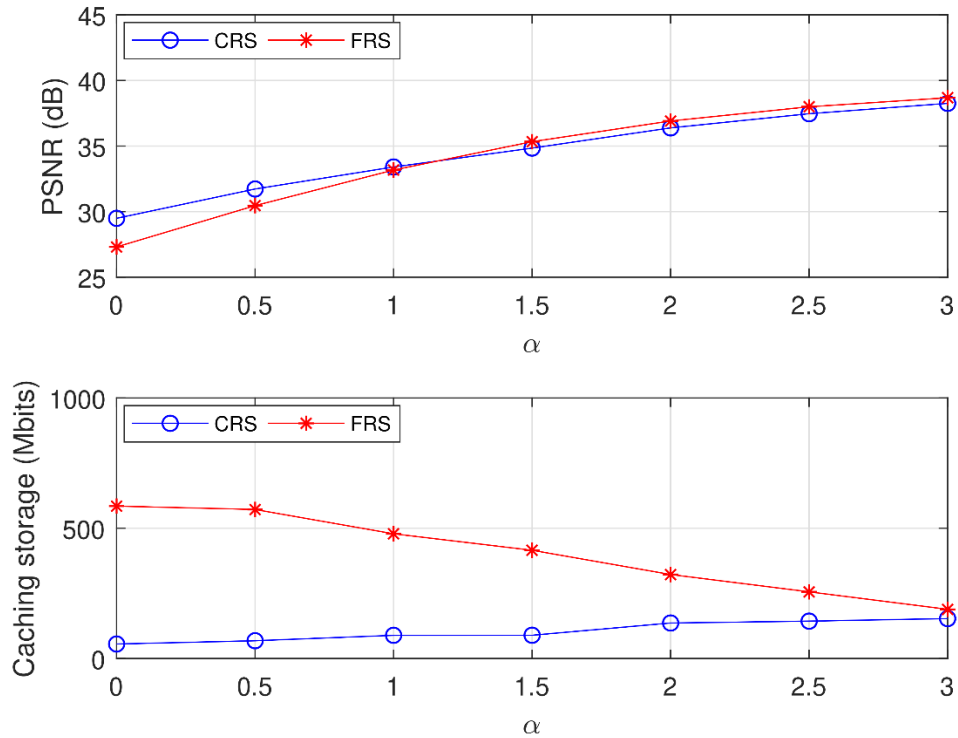
Tiếp theo, ảnh hưởng của các ràng buộc đến hiệu suất của cơ chế CRS được đánh giá trong Hình 3-8. Để làm như vậy, dung lượng yêu cầu bởi MU (δ) được giảm từ 1 xuống 0.5, ràng buộc về dung lượng lưu trữ của FBS (μ) được giảm từ 0.5 xuống 0.3 và tăng tỉ số tín hiệu trên can nhiễu và nhiễu trắng của SU (γ_0) từ 3dB lên 5dB. Kết quả cho thấy, nếu ràng buộc dung lượng δ được yêu cầu bởi MU giảm (từ 1 xuống 0.5), tức là MU chỉ yêu cầu các video có mức phân giải thấp, các kênh truyền cung cấp dung lượng cao sẽ không được chọn để truyền phát nhằm tiết kiệm tài nguyên băng thông của hệ thống. Điều đó có nghĩa là hệ thống chỉ cần phục vụ ở mức chất lượng thấp vì đã đáp ứng đủ yêu cầu của MU. Trong trường hợp giảm ràng buộc về dung lượng lưu trữ của FBS μ (từ 0.5 xuống 0.3), hệ thống có ít khả năng để lưu trữ video và do đó sẽ hạn chế cung cấp kênh truyền chất lượng cao từ FBS đến MU dẫn đến chất lượng video được thu cũng giảm. Ngoài ra, khi tăng tỉ số tín hiệu trên can nhiễu và nhiễu trắng γ_0 (từ 3dB lên 5dB) để đảm bảo QoS cao hơn cho SU, thì số lượng cặp truyền thông D2D giữa CU và NU sẽ giảm để giảm tác động can nhiễu lên SU. Điều này cũng làm giảm hiệu suất hệ thống vì hệ thống không thể khai thác hiệu quả của việc truyền thông D2D để truyền phát video trong khoảng cách gần.



Hình 3-8. Chất lượng HT của CRS với các tham số HT khác nhau theo α

Cuối cùng, để kiểm chứng hiệu suất của cơ chế CRS về chất lượng phát lại (playback) video đạt được và mức độ tiêu tốn dung lượng lưu trữ video tại các FBS, tôi so sánh CRS với phương pháp có lưu trữ và chia sẻ tài nguyên kênh truyền nhưng chỉ lưu trữ với những phiên bản video có độ phân giải cao nhất (FRS – Full rate caching and Resource Sharing). Tức là phương pháp FRS không có cho phép lựa chọn video với các mức phân giải phù hợp để lưu trữ như trong cơ chế CRS. Kết quả trong Hình 3-9 cho thấy, khi $\alpha \leq 1$ thì CRS tốt hơn nhưng FRS sẽ tốt hơn CRS khi $\alpha > 1$. Điều này có thể giải thích là, khi α có giá trị thấp, hệ thống sẽ phục vụ MU với nhiều video có độ phổ biến không lệch nhau quá cao. Trong trường hợp này, việc chọn lựa video có mức phân giải phù hợp để lưu trữ và phục vụ người dùng sẽ làm tăng hiệu quả hệ thống (thỏa mãn người dùng hơn với dung lượng lưu trữ thấp hơn). Ngược lại, khi α có giá trị cao hơn, hệ thống tập trung vào việc phục vụ các video có độ phổ biến cao, nghĩa là sẽ tập trung phục vụ một số ít video hơn. Trong trường hợp này, trong khi hiệu suất của CRS bị hạn chế do bị ràng buộc về dung lượng bộ nhớ lưu trữ và các ràng buộc thông lượng cần thiết, thì các ràng buộc này hoàn toàn không được xét đến trong FRS. Khi xem xét mức tiêu thụ dung lượng bộ nhớ lưu trữ, FRS

rõ ràng yêu cầu sử dụng dung lượng bộ nhớ lưu trữ cao hơn CRS. Và có thể thấy rằng, dung lượng bộ nhớ cần cho cả CRS và FRS để lưu trữ video sẽ hội tụ về một giá trị nếu phiên bản video trong CRS và phiên video trong FRS là như nhau.



Hình 3-9. So sánh kết quả của CRS và FRS

3.6. Kết luận chương 3

Chương 3 đã trình bày cơ chế lưu trữ đa phân giải và chia sẻ tài nguyên tối ưu (CRS – Multi-Resolution Caching and Resource Sharing Optimization) theo nhu cầu người dùng nhằm nâng cao chất lượng truyền video trong 5G UDN. Trong chương này, các hạn chế của Chương 2 đã được cải tiến như: 1) QoS hệ thống được đánh giá tường minh hơn, 2) sử dụng hiệu quả tài nguyên dung lượng lưu trữ tại FBS và tài nguyên phổ tần của SU hơn và 3) thuật giải di truyền được áp dụng để giải bài toán CRS một cách hiệu quả.

KẾT LUẬN

1. Kết quả đạt được

Luận án tập trung vào việc nghiên cứu và đề xuất cơ chế lưu trữ và chia sẻ tài nguyên tối ưu cho các ứng dụng truyền video trong 5G UDN nhằm nâng cao chất lượng lượng dịch vụ (QoS) và sử dụng tài nguyên hiệu quả. Kết quả đạt được cụ thể như sau:

- 1- Đề xuất cơ chế lưu trữ và truyền video cộng tác đa tầng có quan tâm đến hành vi và mối quan hệ xã hội của người dùng (SCS) thông qua việc tận dụng hiệu quả tài nguyên lưu trữ và tài nguyên phổ tần sẵn có của các phần tử trong 5G UDN. Kết quả này được công bố trên tạp chí Springer Mobile Networks & Applications [C1] và được trình bày tại hội nghị Heterogeneous Networking for Quality, Reliability, Security and Robustness (Qshine 2018) [C4].
- 2- Đề xuất cơ chế lưu trữ đa phân giải và truyền video cộng tác có quan tâm đến nhu cầu người dùng (CRS) thông qua việc tận dụng hiệu quả tài nguyên lưu trữ và tài nguyên phổ tần sẵn có của các phần tử trong 5G UDN. Đặc biệt, CRS cải tiến cơ chế SCS bằng cách 1) thay đổi thông số đánh giá QoS từ dung lượng phân phối thành chất lượng video thu được, 2) tận dụng hiệu quả hơn tài nguyên lưu trữ và phổ tần và 3) đề xuất thuật giải di truyền để giải bài toán tối ưu lưu trữ và truyền video một cách hiệu quả. Kết quả này được công bố trên tạp chí IEEE Communications Letters [C2] và được trình bày tại hội nghị Recent Advances on Signal Processing, Telecommunications & Computing (SigTelCom 2020) [C3].

Các công bố liên quan đến luận án là những kết quả nghiên cứu quan trọng, cụ thể là đề xuất mô hình lưu trữ đa tầng kết hợp với chia sẻ tài nguyên tối ưu theo định hướng người dùng nhằm nâng cao hiệu quả truyền video trong 5G UDN. Đóng góp này mang giá trị tham khảo hữu ích cho cộng đồng nghiên cứu thuộc lĩnh vực lưu trữ và chia sẻ tài nguyên cho các ứng dụng truyền video trong 5G UDN.

2. Định hướng nghiên cứu

Nội dung nghiên cứu trong Chương 2 và Chương 3 có sự nối tiếp nhau. Nghiên cứu trong Chương 3 có kế thừa và bổ sung thêm cho các nghiên cứu đã thực hiện trong Chương 2. Tuy nhiên, luận án vẫn còn những hạn chế cần được phát triển trong tương lai liên quan đến mô hình, tiêu chí đánh giá và thuật giải, cụ thể như sau:

- Cải tiến mô hình hệ thống: Một hệ thống hoàn thiện hơn sẽ là hệ thống được kết hợp bởi kết quả của Chương 2 và Chương 3. Nghĩa là phải 1) xem xét đến các video với các mức phân giải khác nhau để lưu trữ hiệu quả hơn tại FBS, 2) tận dụng tài nguyên phổ tần hiệu quả hơn cho truyền thông D2D và 3) xem xét đồng thời đến cư xử, mối quan hệ xã hội và nhu cầu người dùng cũng như yếu tố di chuyển của người dùng. Ngoài ra, một tầng lưu trữ nữa cần được xem xét đến đó là lưu trữ tại các thiết bị bay không người lái (UAV – Unmanned Aerial Vehicle).
- Về tiêu chí đánh giá: Bộ thông số đánh giá chất lượng trải nghiệm người dùng (QoE – Quality of Experience) với những tiêu chí khắt khe hơn cần được xây dựng. Bộ tiêu chí này không đơn thuần là dung lượng phân phối hay chất lượng phát lại video, mà còn là tỷ lệ truy xuất thành công video, tính liên tục cũng như sự ổn định chất lượng trong quá trình phát lại.
- Về thuật giải tối ưu: Với tính phức tạp về hệ thống được mở rộng hoàn thiện hơn và bộ thông số QoE đánh giá hiệu suất hệ thống phức tạp hơn, thuật giải GA cần được nghiên cứu để có những cải tiến hơn về độ chính xác và thời gian thực thi. Ngoài ra, cũng cần tìm hiểu thêm các thuật giải phù hợp khác theo xu thế hiện nay như máy học, học sâu,... để giải quyết bài toán tối ưu và so sánh hiệu quả với thuật giải GA đã áp dụng nhằm chọn lựa thuật giải tốt nhất để triển khai hệ thống.

CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ CỦA NGHIÊN CỨU SINH

1. Công trình đã công bố của luận án

- [C1] **Minh-Phung Bui**, Nguyen-Son Vo, Sang Quang Nguyen, and Quang-Nhat Tran, "Social-Aware Caching and Resource Sharing Maximized Video Delivery Capacity in 5G Ultra-Dense Networks," *Springer Mobile Networks & Applications*, pp. 1-13, July 2019;
- [C2] Nguyen-Son Vo, **Minh-Phung Bui**, Phuc Quang Truong, Cheng Yin, and Antonino Masaracchia, "Multi-tier Caching and Resource Sharing for Video Streaming in 5G Ultra-dense Networks," *IEEE Communications Letters*, vol. 24, no. 7, pp. 1500-1504, July 2020;
- [C3] **Minh-Phung Bui**, Nguyen-Son Vo, Tien-Vu Truong, Thanh-Hieu Nguyen, Nam Van Nguyen and Cheng Yin, "Genetic Algorithms for Multi-tier Caching and Resource Sharing Optimized Video Streaming in 5G Ultra-dense Networks," in Proc. *International Conference on Recent Advances on Signal Processing, Telecommunications & computing (SigTelCom'20)*, Ha Noi, Vietnam, Aug. 2020, pp. 66-71.

2. Công trình đã công bố khác của nghiên cứu sinh

- [C4] **Minh-Phung Bui**, Nguyen-Son Vo, Tien-Thanh Nguyen, Quang-Nhat Tran, and Anh-Tuan Tran, "Social-aware Caching and Resource Sharing Optimization for Video Delivering in 5G Networks," in Proc. EAI International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (**Qshine'18**), Ho Chi Minh City, Vietnam, Dec. 2018, pp. 73-86;
- [C5] Quang-Nhat Tran, Nguyen-Son Vo, **Minh-Phung Bui**, Van-Ca Phan, and Zeeshan Kaleem, "Resource Sharing and Segment Allocation Optimized Video Streaming over Multi-hop Multi-path in Dense D2D 5G Networks," in Proc. *EAI International Conference on Industrial Networks and Intelligent Systems (INISCOM '19)*, Ho Chi Minh City, Vietnam, Aug. 2019, pp. 26-39.

- [C6] Thanh-Minh Phan, Nguyen-Son Vo, **Minh-Phung Bui**, Xuan-Kien Dang, and Dac-Binh Ha, "Downlink Resource Sharing and Caching Helper Selection Control Maximized Multicast Video Delivery Capacity in Dense D2D 5G Networks," *Journal of Science and Technology*, vol. 18, no. 4.2, pp. 12-20, May 2020;
- [C7] Nguyen-Son Vo, Thanh-Minh Phan, **Minh-Phung Bui**, Xuan-Kien Dang, Nguyen Trung Viet, and Cheng Yin , "Social-Aware Spectrum Sharing and Caching Helper Selection Strategy Optimized Multicast Video Streaming in Dense D2D 5G Networks", *IEEE Systems Journal*, pp. 1-12, Jun. 2020.
- [C8] Quang-Nhat Tran, Nguyen-Son Vo, Thanh-Minh Phan, **Minh-Phung Bui**, Minh-Nghia Nguyen, Ayse Kortun, " Downlink Resource Allocation Maximized Video Delivery Capacity over Multi-hop Multi-path in Dense D2D 5G Networks," in Proc. *International Conference on Recent Advances on Signal Processing, Telecommunications & computing (SigTelCom'20)*, Ha Noi, Vietnam, Aug. 2020, pp. 72-76.
- [C9] Thanh-Minh Phan, Nguyen-Son Vo, **Minh-Phung Bui**, Quang-Nhat Tran, Hien M. Nguyen, and Antonino Masaracchia, "Downlink Resource Sharing and Multi-tier Caching Selection Maximized Multicast Video Delivery Capacity in 5G Ultra-dense Networks," in Proc. of *EAI International Conference on Industrial Networks and Intelligent Systems (INISCOM '20)*, Ha Noi, Vietnam, Aug. 2020, pp. 19-31.

TÀI LIỆU THAM KHẢO

- [1]. Hammi, B., R. Khatoun, S. Zeadally, A. Fayad, and L. Khoukhi, *IoT technologies for smart cities*. IET Networks, 2017. **7**(1): p. 1-13.
- [2]. Gupta, A. and R.K. Jha, *A survey of 5G network: Architecture and emerging technologies*. IEEE access, 2015. **3**: p. 1206-1232.
- [3]. Petrov, V., D. Solomitckii, A. Samuylov, M.A. Lema, M. Gapeyenko, D. Moltchanov, S. Andreev, V. Naumov, K. Samouylov, and M. Dohler, *Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments*. IEEE Journal on Selected Areas in Communications, 2017. **35**(9): p. 2038-2055.
- [4]. Kela, P., M. Costa, J. Turkka, K. Leppänen, and R. Jäntti, *Flexible backhauling with massive MIMO for ultra-dense networks*. IEEE Access, 2016. **4**: p. 9625-9634.
- [5]. Gao, Z., L. Dai, D. Mi, Z. Wang, M.A. Imran, and M.Z. Shakir, *MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network*. IEEE Wireless Communications, 2015. **22**(5): p. 13-21.
- [6]. Yang, C., J. Li, P. Semasinghe, E. Hossain, S.M. Perlaza, and Z. Han, *Distributed interference and energy-aware power control for ultra-dense D2D networks: A mean field game*. IEEE Transactions on Wireless Communications, 2016. **16**(2): p. 1205-1217.
- [7]. Zhang, H., S. Huang, C. Jiang, K. Long, V.C. Leung, and H.V. Poor, *Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations*. IEEE Journal on Selected Areas in Communications, 2017. **35**(9): p. 1936-1947.
- [8]. Yang, C., J. Li, Q. Ni, A. Anpalagan, and M. Guizani, *Interference-aware energy efficiency maximization in 5G ultra-dense networks*. IEEE Transactions on Communications, 2016. **65**(2): p. 728-739.
- [9]. Liu, Y., G.Y. Li, and W. Han, *D2D enabled cooperation in massive MIMO systems with cascaded precoding*. IEEE Wireless Communications Letters, 2017. **6**(2): p. 238-241.
- [10]. Xu, C., L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, *Efficiency resource allocation for device-to-device underlay communication systems: A*

- reverse iterative combinatorial auction based approach*. IEEE Journal on Selected Areas in Communications, 2013. **31**(9): p. 348-358.
- [11]. Chou, P.A., H.J. Wang, and V.N. Padmanabhan, *Layered multiple description coding*, in *Proc. Packet video workshop*. 2007, Microsoft Corporation, Redmond, WA (US): U.S. p. 4.
- [12]. Lee, N., X. Lin, J.G. Andrews, and R.W. Heath, *Power control for D2D underlaid cellular networks: Modeling, algorithms, and analysis*. IEEE Journal on Selected Areas in Communications, 2014. **33**(1): p. 1-13.
- [13]. Nguyen, T.T.M. and T.T.B. Dong. *An adaptive cache consistency strategy in a disconnected mobile wireless network*. in *2011 IEEE International Conference on Computer Science and Automation Engineering*. 2011. IEEE.
- [14]. Pham, T.-M. *Analysis of ISP caching in information-centric networks*. in *The 2015 IEEE RIVF International Conference on Computing & Communication Technologies-Research, Innovation, and Vision for Future (RIVF)*. 2015. Can Tho, Vietnam: IEEE.
- [15]. Song, X., Y. Geng, X. Meng, J. Liu, W. Lei, and Y. Wen, *Cache-enabled device to device networks with contention-based multimedia delivery*. IEEE Access, 2017. **5**: p. 3228-3239.
- [16]. Wen, J., K. Huang, S. Yang, and V.O. Li, *Cache-enabled heterogeneous cellular networks: Optimal tier-level content placement*. IEEE Transactions on Wireless Communications, 2017. **16**(9): p. 5939-5952.
- [17]. Liao, J., K.-K. Wong, Y. Zhang, Z. Zheng, and K. Yang, *Coding, multicast, and cooperation for cache-enabled heterogeneous small cell networks*. IEEE Transactions on Wireless Communications, 2017. **16**(10): p. 6838-6853.
- [18]. Li, X., X. Wang, K. Li, Z. Han, and V.C. Leung, *Collaborative multi-tier caching in heterogeneous networks: Modeling, analysis, and design*. IEEE Transactions on Wireless Communications, 2017. **16**(10): p. 6926-6939.
- [19]. Le, T.H.T., N.H. Tran, P.L. Vo, Z. Han, M. Bennis, and C.S. Hong. *Contract-based cache partitioning and pricing mechanism in wireless network slicing*. in *GLOBECOM 2017-2017 IEEE Global Communications Conf*. 2017. Singapore: IEEE.
- [20]. Xu, C., J. Feng, Z. Zhou, J. Wu, and C. Perera, *Cross-Layer Optimization for Cooperative Content Distribution in Multihop Device-to-Device Networks*. IEEE Internet of Things Journal, 2017. **6**(1): p. 278-287.

- [21]. Nguyen, T.T. and T.T. Dong. *An Efficient Cache Invalidation Strategy in Mobile Information Systems*. in *IEEE-RIVF 2010 Int. Conf. on Computing and Commun. Tech.* 2010. Hanoi, Vietnam: IEEE.
- [22]. Minh, T.N.T. and T.D.T. Bich. *An efficient cache replacement strategy in mobile cooperative caching*. in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on*. 2011. Wuhan, China: IEEE.
- [23]. Minh, T.N.T. and T.D.T. Bich. *An efficient model for cooperative caching in mobile information systems*. in *Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on*. 2011. IEEE.
- [24]. Vu, T.X., S. Chatzinotas, B. Ottersten, and T.Q. Duong, *Energy minimization for cache-assisted content delivery networks with wireless backhaul*. *IEEE Wireless Communications Letters*, 2017. **7**(3): p. 332-335.
- [25]. Lin, P., Q. Song, Y. Yu, and A. Jamalipour, *Extensive cooperative caching in D2D integrated cellular networks*. *IEEE Communications Letters*, 2017. **21**(9): p. 2101-2104.
- [26]. Golrezaei, N., A.F. Molisch, A.G. Dimakis, and G. Caire, *Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution*. *IEEE Communications Magazine*, 2013. **51**(4): p. 142-149.
- [27]. Shanmugam, K., N. Golrezaei, A.G. Dimakis, A.F. Molisch, and G. Caire, *Femtocaching: Wireless content delivery through distributed caching helpers*. *IEEE Transactions on Information Theory*, 2013. **59**(12): p. 8402-8413.
- [28]. Chen, M., Y. Hao, L. Hu, K. Huang, and V.K. Lau, *Green and mobility-aware caching in 5G networks*. *IEEE Transactions on Wireless Communications*, 2017. **16**(12): p. 8347-8361.
- [29]. Ndikumana, A., K. Thar, T.M. Ho, N.H. Tran, P.L. Vo, D. Niyato, and C.S. Hong. *In-network caching for paid contents in content centric networking*. in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. 2017. IEEE.
- [30]. Ndikumana, A., N.H. Tran, T.M. Ho, D. Niyato, Z. Han, and C.S. Hong, *Joint incentive mechanism for paid content caching and price based cache replacement policy in named data networking*. *IEEE Access*, 2018. **6**: p. 33702-33717.

- [31]. Tran, T.D. and L.B. Le, *Joint resource allocation and content caching in virtualized content-centric wireless networks*. IEEE Access, 2018. **6**: p. 11329-11341.
- [32]. Ozfatura, E. and D. Gündüz, *Mobility and popularity-aware coded small-cell caching*. IEEE Communications Letters, 2018. **22**(2): p. 288-291.
- [33]. Wang, R., J. Zhang, S. Song, and K.B. Letaief, *Mobility-aware caching in D2D networks*. IEEE Transactions on Wireless Communications, 2017. **16**(8): p. 5001-5015.
- [34]. Shnaiwer, Y.N., S. Sorour, N. Aboutorab, P. Sadeghi, and T.Y. Al-Naffouri. *Network-coded content delivery in femtocaching-assisted cellular networks*. in *2015 IEEE Global Communications Conference (GLOBECOM)*. 2015. IEEE.
- [35]. Duy, B.N.T., Q.A. Nguyen, P.L. Vo, and T.-A. Le. *Optimal content placement for adaptive bit-rate streaming in cache networks*. in *2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)*. 2015. IEEE.
- [36]. Jiang, W., G. Feng, and S. Qin, *Optimal cooperative content caching and delivery policy for heterogeneous cellular networks*. IEEE Transactions on Mobile Computing, 2017. **16**(5): p. 1382 - 1393.
- [37]. Liao, J., K.-K. Wong, M.R. Khandaker, and Z. Zheng, *Optimizing cache placement for heterogeneous small cell networks*. IEEE Communications Letters, 2016. **21**(1): p. 120-123.
- [38]. Le, T.-A., N.D. Thai, and P.L. Vo. *The performance of caching strategies in content centric networking*. in *2017 international conference on information networking (ICOIN)*. 2017. IEEE.
- [39]. Han, W., A. Liu, and V.K. Lau, *PHY-caching in 5G wireless networks: Design and analysis*. IEEE Communications Magazine, 2016. **54**(8): p. 30-36.
- [40]. Qiao, J., Y. He, and X.S. Shen, *Proactive Caching for Mobile Video Streaming in Millimeter Wave 5G Networks*. IEEE Trans. Wireless Communications, 2016. **15**(10): p. 7187-7198.
- [41]. Baştuğ, E., M. Bennis, and M. Debbah, *Proactive caching in 5G small cell networks*, in *Towards 5G: Applications, Requirements and Candidate Technologies*. 2017, Wiley Telecom. p. 78-98.

- [42]. Chen, Z., N. Pappas, and M. Kountouris, *Probabilistic caching in wireless D2D networks: Cache hit optimal versus throughput optimal*. IEEE Communications Letters, 2017. **21**(3): p. 584-587.
- [43]. Chen, Y., M. Ding, J. Li, Z. Lin, G. Mao, and L. Hanzo, *Probabilistic small-cell caching: Performance analysis and optimization*. IEEE Transactions on Vehicular Technology, 2017. **66**(5): p. 4341-4354.
- [44]. Vo, P.L., L. Van Nguyen, T.-A. Le, and D.N.M. Dang. *A QoE-based caching algorithm for HTTP adaptive streaming contents in radio access networks*. in *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*. 2016. IEEE.
- [45]. Park, G.S., W. Kim, S.H. Jeong, and H. Song, *Smart base station-assisted partial-flow device-to-device offloading system for video streaming services*. IEEE Transactions on Mobile Computing, 2017. **16**(9): p. 2639-2655.
- [46]. Zhang, J., X. Zhang, Z. Yan, Y. Li, W. Wang, and Y. Zhang. *Social-aware cache information processing for 5G ultra-dense networks*. in *Wireless Communications & Signal Processing (WCSP), 2016 8th International Conference on*. 2016. IEEE.
- [47]. Zhu, K., W. Zhi, X. Chen, and L. Zhang, *Socially Motivated Data Caching in Ultra-Dense Small Cell Networks*. IEEE Network, 2017. **31**(4): p. 42-48.
- [48]. Gregori, M., J. Gómez-Vilardebó, J. Matamoros, and D. Gündüz, *Wireless content caching for small cell and D2D networks*. IEEE Journal on Selected Areas in Communications, 2016. **34**(5): p. 1222-1234.
- [49]. Networking, C.V. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*. 2017;
- [50]. Nguyen, M.T., *An energy-efficient framework for multimedia data routing in Internet of Things (IoTs)*. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 2019. **6**(19).
- [51]. Systems, C. *Cisco Annual Internet Report (2018–2023) White Paper*. 2020 [cited 2020];
- [52]. Chopra, G., R.K. Jha, and S. Jain, *A survey on ultra-dense network and emerging technologies: Security challenges and possible solutions*. Journal of Network and Computer Applications, 2017. **95**: p. 54-78.

- [53]. Hossain, E., L.B. Le, and D. Niyato, *Radio resource management in multi-tier cellular wireless networks*. 2013: John Wiley & Sons.
- [54]. Vo, N.-S., T.Q. Duong, M. Guizani, and A. Kortun, *5G Optimized Caching and Downlink Resource Sharing for Smart Cities*. IEEE Access, 2018. **6**: p. 31457 - 31468.
- [55]. Tran, Q.-N., N.-S. Vo, M.-P. Bui, V.-C. Phan, Z. Kaleem, and T.Q. Duong. *Resource Sharing and Segment Allocation Optimized Video Streaming over Multi-hop Multi-path in Dense D2D 5G Networks*. in *International Conference on Industrial Networks and Intelligent Systems*. 2019. Springer.
- [56]. Bui, M.-P., N.-S. Vo, S.Q. Nguyen, and Q.-N. Tran, *Social-Aware Caching and Resource Sharing Maximized Video Delivery Capacity in 5G Ultra-Dense Networks*. Mobile Networks and Applications, 2019: p. 1-13.
- [57]. Bui, M.-P., N.-S. Vo, T.-T. Nguyen, Q.-N. Tran, and A.-T. Tran. *Social-aware caching and resource sharing optimization for video delivering in 5G networks*. in *International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*. 2018. Springer.
- [58]. Tan, Y., Y. Yuan, T. Yang, Y. Xu, and B. Hu. *Femtocaching in wireless video networks: Distributed framework based on exact potential game*. in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*. 2016. IEEE.
- [59]. Quer, G., I. Pappalardo, B.D. Rao, and M. Zorzi, *Proactive caching strategies in heterogeneous networks with device-to-device communications*. IEEE Transactions on Wireless Communications, 2018. **17**(8): p. 5270-5281.
- [60]. Vasuki, A., *BAT Algorithms*, in *Nature-Inspired Optimization Algorithms*. 2020, CRC Press. p. 167-179.
- [61]. Mirjalili, S., S.M. Mirjalili, and X.-S. Yang, *Binary bat algorithm*. Neural Computing and Applications, 2014. **25**(3): p. 663-681.
- [62]. Chipperfield, A. and P. Fleming. *The MATLAB genetic algorithm toolbox*. in *IEE Colloquium on Applied Control Techniques Using MATLAB*. 1995. London, UK.
- [63]. Xiang, W., C. Zhu, C.K. Siew, Y. Xu, and M. Liu, *Forward error correction-based 2-D layered multiple description coding for error-resilient H. 264 SVC video transmission*. IEEE transactions on circuits and systems for video technology, 2009. **19**(12): p. 1730-1738.

- [64]. Du, X., N.-S. Vo, W. Cheng, T.Q. Duong, and L. Shu, *Joint replication density and rate allocation optimization for VoD systems over wireless mesh networks*. IEEE transactions on circuits and systems for video technology, 2013. **23**(7): p. 1260-1273.
- [65]. Oladejo, S.O. and O.E. Falowo, *Latency-Aware Dynamic Resource Allocation Scheme for Multi-Tier 5G Network: A Network Slicing-Multitenancy Scenario*. IEEE Access, 2020. **8**: p. 74834-74852.
- [66]. Vo, N.-S., T.Q. Duong, H.D. Tuan, and A. Kortun, *Optimal video streaming in dense 5g networks with d2d communications*. IEEE Access, 2018. **6**: p. 209-223.
- [67]. Vo, N.-S., T.Q. Duong, and M. Guizani. *QoE-oriented resource efficiency for 5G two-tier cellular networks: A femtocaching framework*. in *Global Communications Conference (GLOBECOM), 2016 IEEE*. 2016. IEEE.
- [68]. Huynh, D.-T., X. Wang, T.Q. Duong, N.-S. Vo, and M. Chen, *Social-aware energy efficiency optimization for device-to-device communications in 5G networks*. Computer Communications, 2018. **120**: p. 102-111.
- [69]. Liu, J., N. Kato, J. Ma, and N. Kadowaki, *Device-to-device communication in LTE-advanced networks: A survey*. IEEE Communications Surveys & Tutorials, 2014. **17**(4): p. 1923-1940.
- [70]. Lin, S.-H., K.-Y. Chen, J.-C. Kao, and Y.-F. Hsiao. *Fast spectrum reuse and power control for device-to-device communication*. in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*. 2017. IEEE.
- [71]. Hussain, F., M.Y. Hassan, M.S. Hossen, and S. Choudhury, *System capacity maximization with efficient resource allocation algorithms in D2D communication*. IEEE Access, 2018. **6**: p. 32409-32424.
- [72]. Ali, F., S. Jangsher, and F.A. Bhatti. *Resource sharing for D2D communication in multi small cell networks*. in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. 2017. IEEE.
- [73]. Wu, D., Q. Liu, H. Wang, Q. Yang, and R. Wang, *Cache less for more: Exploiting cooperative video caching and delivery in D2D communications*. IEEE Transactions on Multimedia, 2018. **21**(7): p. 1788-1798.

- [74]. Benevenuto, F., T. Rodrigues, M. Cha, and V. Almeida. *Characterizing user behavior in online social networks*. in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*. 2009. Illinois, Chicago, USA: ACM.
- [75]. Cha, M., H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. *I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system*. in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. 2007. ACM.
- [76]. Ma, C., M. Ding, H. Chen, Z. Lin, G. Mao, Y.-C. Liang, and B. Vucetic, *Socially aware caching strategy in device-to-device communication networks*. *IEEE Transactions on Vehicular Technology*, 2018. **67**(5): p. 4615-4629.
- [77]. Ma, C., M. Ding, H. Chen, Z. Lin, G. Mao, and X. Li. *Socially aware distributed caching in device-to-device communication networks*. in *2016 IEEE Globecom Workshops (GC Wkshps)*. 2016. IEEE.
- [78]. Chandrasekhar, V. and J.G. Andrews, *Spectrum allocation in tiered cellular networks*. *IEEE Transactions on Communications*, 2009. **57**(10): p. 3059-3068.
- [79]. Cheung, W.C., T.Q. Quek, and M. Kountouris, *Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks*. *IEEE Journal on Selected Areas in Communications*, 2012. **30**(3): p. 561-574.
- [80]. Bai, B., L. Wang, Z. Han, W. Chen, and T. Svensson, *Caching based socially-aware D2D communications in wireless content delivery networks: A hypergraph framework*. *IEEE Wireless Communications*, 2016. **23**(4): p. 74-81.
- [81]. Zhang, X., Y. Li, Y. Zhang, J. Zhang, H. Li, S. Wang, and D. Wang, *Information caching strategy for cyber social computing based wireless networks*. *IEEE Transactions on Emerging Topics in Computing*, 2017. **5**(3): p. 391-402.
- [82]. Zhang, Y., E. Pan, L. Song, W. Saad, Z. Dawy, and Z. Han, *Social network aware device-to-device communication in wireless networks*. *IEEE Transactions on Wireless Communications*, 2015. **14**(1): p. 177-190.
- [83]. Tegos, S.A., P.D. Diamantoulakis, A.S. Lioumpas, P.G. Sarigiannidis, and G.K. Karagiannidis, *Slotted ALOHA with NOMA for the next generation IoT*. *IEEE Transactions on Communications*, 2020. **68**(10): p. 6289-6301.

- [84]. Pandey, S., K. Shandilya, and S. Agarwal. *Prioritized S-ALOHA for URLLC*. in *2020 International Wireless Communications and Mobile Computing (IWCMC)*. 2020. IEEE.
- [85]. Balevi, E., F.T. Al Rabee, and R.D. Gitlin. *ALOHA-NOMA for massive machine-to-machine IoT communication*. in *2018 IEEE International Conference on Communications (ICC)*. 2018. IEEE.
- [86]. Yang, L., S. Song, and K.B. Letaief, *Optimal overlay cognitive spectrum access with F-ALOHA in macro-femto heterogeneous networks*. *IEEE Transactions on Wireless Communications*, 2015. **15**(2): p. 1323-1335.
- [87]. Elkourdi, M., A. Mazin, and R.D. Gitlin. *Slotted Aloha-NOMA with MIMO Beamforming for Massive M2M Communication in IoT Networks*. in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. 2018. IEEE.
- [88]. Breslau, L., P. Cao, L. Fan, G. Phillips, and S. Shenker. *Web caching and Zipf-like distributions: Evidence and implications*. in *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*. 1999. IEEE.
- [89]. Mehrotra, S., *On the implementation of a primal-dual interior point method*. *SIAM Journal on optimization*, 1992. **2**(4): p. 575-601.
- [90]. Zhang, Y., *Solving large-scale linear programs by interior-point methods under the MATLAB environment*. *Optimization Methods and Software*, 1998. **10**(1): p. 1-31.
- [91]. Lin, P., K.S. Khan, Q. Song, and A. Jamalipour, *Caching in Heterogeneous Ultradense 5G Networks: A Comprehensive Cooperation Approach*. *IEEE Vehicular Technology Magazine*, 2019. **14**(2): p. 22-32.
- [92]. Goian, H.S., O.Y. Al-Jarrah, S. Muhaidat, Y. Al-Hammadi, P. Yoo, and M. Dianati, *Popularity-based video caching techniques for cache-enabled networks: a survey*. *IEEE Access*, 2019. **7**: p. 27699-27719.
- [93]. Nguyen, H.T., H.D. Tuan, T.Q. Duong, H.V. Poor, and W.-J. Hwang, *Collaborative multicast beamforming for content delivery by cache-enabled ultra dense networks*. *IEEE Transactions on Communications*, 2019. **67**(5): p. 3396-3406.
- [94]. Nguyen, T., D. Nguyen, and V. Nguyen, *Quality of Service Provisioning for D2D Users in Heterogeneous Networks*. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, 2019. **6**(21).

- [95]. Vo, N.-S., M.-P. Bui, P.Q. Truong, C. Yin, and A. Masaracchia, *Multi-tier Caching and Resource Sharing for Video Streaming in 5G Ultra-dense Networks*. IEEE Communications Letters, 2020. **24**(7): p. 1500-1504.
- [96]. Bhardwaj, A. and S. Agnihotri, *Energy-and spectral-efficiency trade-off for D2D-multicasts in underlay cellular networks*. IEEE Wireless Communications Letters, 2018. **7**(4): p. 546-549.
- [97]. Ha, D.-B. and S.Q. Nguyen, *Outage performance of energy harvesting DF relaying NOMA networks*. Mobile Networks and Applications, 2018. **23**(6): p. 1572-1585.
- [98]. Sullivan, G.J., J.-R. Ohm, W.-J. Han, and T. Wiegand, *Overview of the high efficiency video coding (HEVC) standard*. IEEE Transactions on circuits and systems for video technology, 2012. **22**(12): p. 1649-1668.
- [99]. Chan, C. and C. Tsai, *Collision-Free Speed Alteration Strategy for Human Safety in Human-Robot Coexistence Environments*. IEEE Access, 2020. **8**: p. 80120-80133.
- [100]. Lin, Z., M. Lin, Y. Huang, T. De Cola, and W.-P. Zhu, *Robust multi-objective beamforming for integrated satellite and high altitude platform network with imperfect channel state information*. IEEE Transactions on Signal Processing, 2019. **67**(24): p. 6384-6396.
- [101]. Ilyas, M.A., G. Abbas, T. Alquthami, M. Awais, and M.B. Rasheed, *Multi-Objective Optimal Power Flow With Integration of Renewable Energy Sources Using Fuzzy Membership Function*. IEEE Access, 2020. **8**: p. 143185-143200.
- [102]. Liu, Q., Y. Xu, and Y. Zhou, *A class of exact penalty functions and penalty algorithms for nonsmooth constrained optimization problems*. Journal of Global Optimization, 2019. **76**(4): p. 745-768.
- [103]. *HM Reference Software Version 12.0*.