

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

ĐỖ TRUNG ANH

NGHIÊN CỨU TỐI ƯU HÓA THÔNG LƯỢNG VÀ
ĐỘ TRỄ TRONG MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG
SỬ DỤNG KỸ THUẬT ĐỆM DỮ LIỆU

LUẬN ÁN TIẾN SĨ KỸ THUẬT

HÀ NỘI - 2021

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

ĐỖ TRUNG ANH

NGHIÊN CỨU TỐI ƯU HÓA THÔNG LƯỢNG VÀ
ĐỘ TRỄ TRONG MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG
SỬ DỤNG KỸ THUẬT ĐỆM DỮ LIỆU

Chuyên ngành: KỸ THUẬT VIỄN THÔNG
Mã số: 9.52.02.08

LUẬN ÁN TIẾN SĨ KỸ THUẬT

NGƯỜI HƯỚNG DẪN KHOA HỌC:
PGS. TS. ĐẶNG HOÀI BẮC

HÀ NỘI - 2021

MỤC LỤC

MỤC LỤC.....	i
Danh mục các từ viết tắt.....	1
Danh sách hình vẽ.....	iv
Danh sách bảng.....	vi
Danh mục ký hiệu toán học.....	vii
MỞ ĐẦU.....	1
Chương 1. TỔNG QUAN VỀ MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG.....	13
1.1. Mô hình mạng vô tuyến hướng nội dung.....	13
1.2. Các tham số hiệu năng mạng và ký hiệu toán học sử dụng trong luận án.....	19
1.2.1. Các tham số hiệu năng mạng.....	19
1.2.2. Ký hiệu toán học sử dụng trong luận án.....	19
1.3. Các công trình nghiên cứu khoa học liên quan.....	20
1.4. Nhận xét về công trình nghiên cứu của các tác giả khác và hướng nghiên cứu của luận án.....	23
1.4.1. Nhận xét về công trình nghiên cứu của các tác giả khác.....	23
1.4.2. Hướng nghiên cứu của luận án.....	25
1.5. Kết luận Chương 1.....	29

Chương 2. TỐI ƯU HÓA THÔNG LƯỢNG VÀ ĐỘ TRỄ CỦA MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG SỬ DỤNG MÔ HÌNH DÒNG CHẢY.....	30
2.1. Mô hình mạng vô tuyến hướng nội dung sử dụng mô hình dòng chảy	32
2.2. Đề xuất phương pháp định tuyến truyền tin	36
2.3. Thông lượng và độ trễ của mạng.....	41
2.4. Tối ưu hóa thông lượng và độ trễ	47
2.4.1. Xây dựng bài toán tối ưu hóa thông lượng và độ trễ mạng...	48
2.4.2. Giải bài toán tối ưu hóa thông lượng và độ trễ mạng.....	49
2.4.3. Nghiệm tối ưu hóa sử dụng phần mềm tính toán trên máy tính ..	61
2.4.4. Thông lượng và độ trễ tối ưu.....	63
2.5. Hiệu năng mạng vô tuyến hướng nội dung sử dụng phương pháp lưu trữ dữ liệu cơ bản.....	66
2.5.1. Xây dựng bài toán tối ưu hóa thông lượng và độ trễ.....	67
2.5.2. Giải bài toán tối ưu hóa và so sánh	68
2.6. So sánh và đánh giá	71
2.7. Kết luận Chương 2	73

Chương 3. TỐI ƯU HÓA THÔNG LƯỢNG VÀ ĐỘ TRỄ CỦA MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG SỬ DỤNG PHƯƠNG PHÁP PHÂN MẢNH TẬP DỮ LIỆU	76
3.1. Mô hình mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tập dữ liệu.....	78
3.2. Phương pháp thu nhận mảnh tin và đề xuất phương pháp định tuyến truyền tin	80
3.2.1. Phương pháp thu nhận mảnh tin.....	80
3.2.2. Phương pháp định tuyến truyền tin.....	81
3.3. Thông lượng và độ trễ.....	84
3.3.1. Tối ưu hóa thông lượng và độ trễ trường hợp sử dụng phương pháp thu nhận mảnh tin tuần tự.....	85
3.3.2. Tối ưu hóa thông lượng và độ trễ trường hợp sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên.....	90
3.3.3. Nghiệm tối ưu hóa sử dụng phần mềm tính toán trên máy tính ..	97
3.4. So sánh và đánh giá	99
3.5. Kết luận Chương 3	102
KẾT LUẬN.....	104
DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ.....	109
TÀI LIỆU THAM KHẢO	111

Danh mục các từ viết tắt

Từ viết tắt	Nghĩa Tiếng Anh	Nghĩa Tiếng Việt
5G	The fifth generation	Hệ thống thông tin di động thế hệ thứ năm
CoMP	Coordinated multipoint transmission	Truyền dẫn đa điểm liên kết
LTE	Long term evolution	Tiến hóa dài hạn
MIMO	Multiple input multiple output	Nhiều đầu vào nhiều đầu ra
MAC	Medium access control	Điều khiển truy nhập môi trường
PHY	Physical layer protocol	Giao thức lớp vật lý
RWMM	Random walk mobility model	Mô hình bước đi ngẫu nhiên

Danh sách hình vẽ

1	Mô hình mạng vô tuyến truyền thống.	2
2	Mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu.	4
1.1	Mô hình mạng vô tuyến hướng nội dung.	14
1.2	Mô hình bước đi ngẫu nhiên (RWMM).	16
2.1	Mạng vô tuyến hướng nội dung sử dụng mô hình dòng chảy.	33
2.2	Giai đoạn thứ nhất của quá trình định tuyến truyền dữ liệu.	37
2.3	Giai đoạn thứ hai của quá trình định tuyến truyền dữ liệu.	39
2.4	Phương pháp tối ưu lưu trữ dữ liệu tương ứng với sự biến thiên của tham số m	59
2.5	Các tập tệp dữ liệu $\{A_m^*\}_{m=1}^M$ và $\{B_m^*\}_{m=1}^M$ biến thiên theo tham số m	61
2.6	Tập nghiệm tối ưu $\{A_m^* + B_m^*\}_{m=1}^{200}$ tương ứng với tệp dữ liệu m	63
2.7	Các chế độ hoạt động của mạng theo mối quan hệ của các tham số α , δ , β , và γ	65
2.8	Các tập nghiệm tối ưu $\{A_m^*\}_{m=1}^{200}$ và $\{B_m^*\}_{m=1}^{200}$ tương ứng với sự biến thiên của tham số m	70
3.1	Mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu.	79

3.2	Phương pháp thu nhận mảnh tin của tệp dữ liệu m	82
3.3	Giai đoạn thứ nhất của quá trình định tuyến truyền tin.	83
3.4	Giai đoạn thứ hai của quá trình định tuyến truyền tin.	83
3.5	Nghiệm tối ưu số lượng bản sao lưu trữ của các tệp dữ liệu theo tham số m	95
3.6	Tập nghiệm tối ưu $\{X_m^*\}_{m=1}^{200}$ của các bài toán sử dụng phương pháp phân mảnh tệp dữ liệu tương ứng với tham số m	98

Danh sách bảng

2.1	Bảng các tham số mạng được sử dụng để tính toán trên máy tính	62
2.2	Bảng giá trị thông lượng và độ trễ mạng tối ưu	66
2.3	Bảng so sánh độ trễ mạng tối ưu	72
2.4	Bảng so sánh thông lượng mạng tối ưu	72
3.1	Bảng các tham số mạng được sử dụng để tính toán trên máy tính	98
3.2	Bảng giá trị thông lượng mạng tối ưu	101
3.3	Bảng giá trị độ trễ mạng tối ưu	101

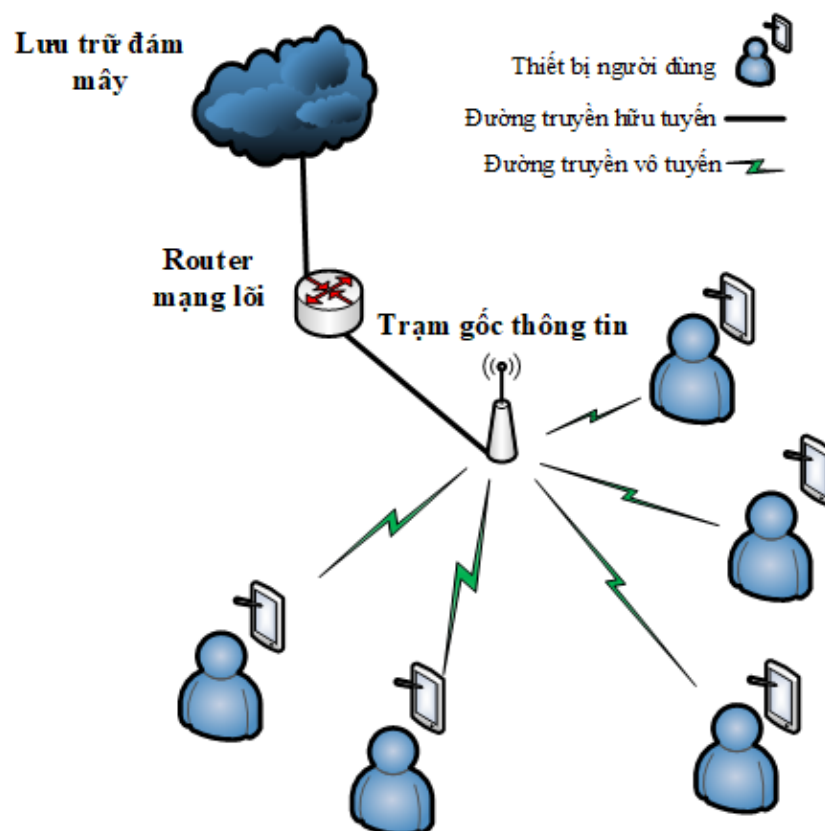
Danh mục ký hiệu toán học

Ký hiệu	Ý nghĩa
n	Số lượng thiết bị di động người dùng trong mạng
$f(n)$	Số lượng trạm gốc thông tin trong mạng
M	Số lượng tệp dữ liệu trong thư viện mạng
p_m	Xác suất yêu cầu tải tệp dữ liệu m
A_m	Số lượng bản sao của tệp dữ liệu m được lưu tại bộ nhớ đệm của các thiết bị di động người dùng
B_m	Số lượng bản sao của tệp dữ liệu m được lưu tại bộ nhớ đệm của các trạm gốc thông tin
X_m	Số lượng bản sao của tệp dữ liệu m được lưu tại bộ nhớ đệm của các nút mạng
K_n	Độ lớn bộ nhớ đệm của thiết bị di động người dùng
K_{BS}	Độ lớn bộ nhớ đệm của trạm gốc thông tin
K	Số lượng mảnh tin được phân mảnh của mỗi tệp dữ liệu
$\lambda(n)$	Thông lượng trung bình tại các nút mạng
$D(n)$	Độ trễ của mạng

MỞ ĐẦU

Thế hệ mạng vô tuyến 5G và các thế hệ mạng vô tuyến tiếp theo hứa hẹn khả năng hỗ trợ các kết nối nhanh với độ tin cậy cao, và đồng thời đáp ứng được mức độ gia tăng về lưu lượng dữ liệu người dùng trong tương lai. Tuy nhiên, các yêu cầu về các tài nguyên như năng lượng và băng thông truyền dữ liệu lại không thể tăng lên tỷ lệ thuận với sự phát triển của lưu lượng dữ liệu người dùng. Trong thời đại bùng nổ về thiết bị di động thông minh hiện nay, số lượng người dùng sử dụng điện thoại thông minh đang gia tăng nhanh chóng. Theo đó, lưu lượng dữ liệu sử dụng của người dùng internet từ các thiết bị di động cũng bùng nổ theo cấp số nhân. Theo dự báo của Cisco [10, 11], lưu lượng dữ liệu sử dụng của các thiết bị di động năm 2022 cao hơn gấp 7 lần so với năm 2017, đạt xấp xỉ 77,5 exabytes dữ liệu mỗi tháng cho đến năm 2022. Trong đó, các đoạn phim video là đối tượng dữ liệu chính do sự phát triển lớn mạnh của các dịch vụ video trực tuyến theo yêu cầu người dùng [52] từ các nhà cung cấp phổ biến như Youtube, Netflix, iTune, hay Amazon Prime.

Các nghiên cứu ở [10, 29, 52, 60] chỉ ra rằng, một phần rất lớn của dữ liệu mạng được trao đổi hàng ngày liên quan đến các lượt tải dữ liệu của các nội dung phổ biến trên mạng tại thời điểm đó. Ví dụ, 10 % các video được xem nhiều nhất chiếm hơn 80 % tổng số lượt xem trên Youtube [29], đặc biệt là



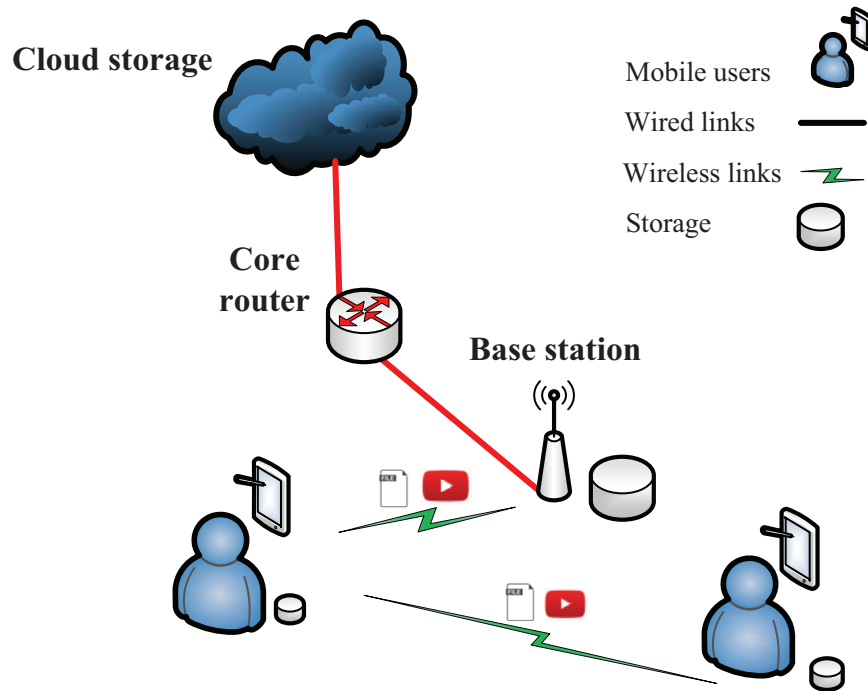
Hình 1: Mô hình mạng vô tuyến truyền thống.

các video thuộc top các video thịnh hành trên youtube (top trending). Chính vì vậy, các thiết bị đầu cuối trong các mạng truy nhập vô tuyến và các mạng truyền tải lõi của các nhà cung cấp dịch vụ mạng có xu hướng truyền đi cùng những dữ liệu giống nhau lặp đi lặp lại rất nhiều lần trong ngày.

Những yêu cầu đối với các dịch vụ truyền thông vô tuyến đã và đang dịch chuyển dần từ các dịch vụ hướng kết nối là các dịch vụ thoại truyền thống và tin nhắn văn bản sang các dịch vụ hướng nội dung, điển hình là các dịch vụ đa phương tiện, mạng xã hội và các ứng dụng di động. Với việc lưu lượng dữ liệu người dùng sử dụng tăng cao đáng kể trong thế hệ thông tin di động thứ 5 và các thế hệ tương lai, các điểm kết nối của đường truyền mạng lõi backhaul và các điểm truy nhập sẽ phải xử lý khối lượng lưu lượng dữ liệu

trao đổi rất lớn [35] (Hình 1). Trên thực tế, luôn luôn tồn tại khoảng cách rất lớn giữa mong muốn, nhu cầu sử dụng dịch vụ của người dùng với sự đáp ứng của các nhà cung cấp dịch vụ mạng. Người tiêu dùng có nhu cầu sử dụng lưu lượng dữ liệu rất lớn nhưng chi trả cho dịch vụ dữ liệu của các nhà mạng lại hạn chế. Trong khi đó, do giới hạn của các đường truyền mạng lõi và năng lực xử lý dữ liệu tại các nút mạng, các nhà mạng bị giới hạn về khả năng đáp ứng nhu cầu dữ liệu của người tiêu dùng nên luôn có những quy định chặt chẽ về mặt băng thông và các gói cước kèm kinh phí. Những bùng nổ về nhu cầu sử dụng dữ liệu của người tiêu dùng dẫn đến nhiều vấn đề đối với các nhà cung cấp dịch vụ mạng. Để có thể đáp ứng được nhu cầu và nhận được sự hài lòng về chất lượng dịch vụ của người tiêu dùng thì việc tìm kiếm các giải pháp về mặt kỹ thuật để giải quyết các vấn đề liên quan tới giới hạn của khả năng truyền tải của mạng là việc cấp thiết của các nhà cung cấp dịch vụ mạng. Điều này dẫn đến gánh nặng về tài chính đối với các nhà mạng khi yêu cầu đòi hỏi về việc nâng cấp đường truyền mạng lõi trở nên vô cùng rõ ràng.

Hiện nay, có rất nhiều phương pháp và kỹ thuật đã và đang được nghiên cứu để góp phần giảm lưu lượng truyền tải của mạng và giải quyết những vấn đề nêu trên. Tuy nhiên, bất chấp những nỗ lực của các nhà cung cấp dịch vụ mạng và các công ty sản xuất thiết bị mạng nhằm nâng cao băng thông đường truyền mạng nhờ việc áp dụng các kỹ thuật tinh vi ở cả phần lớp vật lý (PHY) của mạng và các lớp điều khiển truy nhập kênh (MAC) trong các hệ thống thông tin di động thế hệ mới LTE và LTE-Advanced, ví dụ như các kỹ thuật MIMO (massive multiple-input multiple-output), kết hợp sóng mang, và truyền dẫn đa điểm liên kết (CoMP - coordinated multipoint



Hình 2: Mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu.

transmission), hiệu năng tối ưu của phổ tần vô tuyến đã đang gần đạt đến mức giới hạn lý thuyết. Trong các kỹ thuật mới nổi khác liên quan đến vấn đề đáp ứng nhu cầu lớn của truyền tải dữ liệu của mạng đang được đề xuất, kỹ thuật đệm dữ liệu (caching), lưu trữ dữ liệu trong mạng cho phép truyền dữ liệu offloading [4, 19, 21, 24, 54, 59] đang là phương pháp nhận được nhiều sự quan tâm chú ý của các nhà khoa học với những ưu điểm hứa hẹn phù hợp với tương lai mạng vô tuyến (Hình 2).

Tại các điểm biên của mạng vô tuyến như các trạm gốc thông tin di động và các thiết bị đầu cuối người dùng luôn được trang bị sẵn sàng các mô đun lưu trữ dữ liệu (ổ cứng) [3]. Những mô đun lưu trữ này có thể được sử dụng để lưu trữ các dữ liệu phổ biến thường được người dùng quan tâm. Ngày nay, nhờ tiến bộ của ngành công nghiệp chế tạo, các mô đun lưu trữ dữ liệu là

các ổ đĩa cứng có giá ngày càng rẻ sau mỗi năm và mỗi thiết bị đầu cuối di động thông minh thường được trang bị dung lượng lưu trữ lên tới hàng trăm gigabytes. Đây là cơ sở cho việc áp dụng kỹ thuật đệm dữ liệu được thuận lợi hơn. Như vậy, đối với mạng viễn thông trong tương lai khi được áp dụng kỹ thuật đệm dữ liệu, các thuê bao người dùng có thể xem youtube hoặc các nội dung đa phương tiện với phần nguồn cung cấp dữ liệu đến từ các bộ nhớ chia sẻ của các thiết bị di động khác xung quanh, hoặc bộ nhớ chia sẻ của chính trạm gốc thông tin đang phục vụ cho thuê bao đó mà không cần thiết lập kênh truyền hữu tuyến của mạng lõi kết nối tới các máy chủ lưu trữ nội dung của các ứng dụng đang hoạt động nữa. Từ đó sẽ giúp giảm tải lưu lượng của mạng lõi, đáp ứng được nhu cầu về dữ liệu của người dùng, và duy trì sự ổn định của mạng, đảm bảo được chất lượng dịch vụ của các nhà cung cấp dịch vụ. Bên cạnh việc giảm bớt gánh nặng của việc nâng cấp đường truyền mạng lõi, việc sử dụng kỹ thuật đệm dữ liệu cũng là một cách hiệu quả giúp cho giảm độ trễ và nghẽn mạng khi các thuê bao di động có thể tải được các dữ liệu mong muốn từ các trạm gốc thông tin di động hoặc các thiết bị di động khác trong phạm vi gần một cách trực tiếp mà không cần phải thực hiện thông qua các kết nối với mạng lõi.

Ý tưởng về mạng hướng nội dung áp dụng kỹ thuật đệm dữ liệu đã được đưa ra và được nghiên cứu và phát triển rộng rãi trong mạng hữu tuyến, trong đó các thành phần của dữ liệu được định tuyến và truyền trực tiếp theo dạng gói, và các gói dữ liệu được tự động lưu trữ tại các bộ định tuyến trên đường truyền tin [2, 7, 25]. Theo đó, thiết kế đệm dữ liệu tại các bộ định tuyến, bao gồm cả việc thiết kế lưu trữ nội dung và cập nhật nội dung, là yếu tố quan trọng ảnh hưởng đến hiệu năng hệ thống là thông lượng và độ trễ.

Tiếp thu những kết quả đã đạt được từ mạng hữu tuyến, sử dụng kỹ thuật đệm dữ liệu để lưu trữ dữ liệu tại bộ nhớ lưu trữ chia sẻ của các điểm biên của mạng vô tuyến là các trạm gốc thông tin và thiết bị người dùng là ý tưởng mới có thể tận dụng được những bài học đã có và xem xét thêm những đặc điểm mới riêng đối với mạng vô tuyến. Đặc tính tự nhiên của đường truyền vô tuyến sẽ chắc chắn ảnh hưởng tới việc thiết kế đệm dữ liệu tại các thực thể mạng vô tuyến và quá trình truyền tin. Đây cũng là đặc điểm khác biệt so với mạng hữu tuyến cần được nghiên cứu đối với việc sử dụng kỹ thuật đệm dữ liệu cho mạng vô tuyến hướng nội dung. Phần lớn các nghiên cứu đưa ra hiệu năng tối ưu của mạng vô tuyến hướng nội dung trước đây đều đặt vấn đề sử dụng kỹ thuật đệm dữ liệu đối để chia sẻ dữ liệu giữa các bộ nhớ lưu trữ chia sẻ của các thiết bị người dùng với nhau. Nghiên cứu [36] đã đề cập sự tham gia đồng thời kỹ thuật đệm dữ liệu đối với thiết bị người dùng và trạm gốc thông tin, tuy nhiên lại áp dụng với dung lượng bộ nhớ lưu trữ tại trạm gốc thông tin là không giới hạn. Do vậy, khả năng tham gia và ảnh hưởng của dung lượng lưu trữ chia sẻ của các trạm gốc thông tin khi sử dụng kỹ thuật đệm dữ liệu vẫn chưa được đánh giá một cách đầy đủ. Ngoài ra, để đơn giản hóa mô hình tính toán, các nghiên cứu trước đây đều giả sử kích thước các tệp dữ liệu là lý tưởng, đủ nhỏ để có thể truyền đi hoàn toàn giữa các thực thể mạng với nhau trong khoảng thời gian của mỗi khe thời gian. Ý tưởng sử dụng phương pháp phân mảnh tệp dữ liệu [5] có thể được áp dụng để đánh giá sự ảnh hưởng của kích thước tệp dữ liệu đối với các tham số hiệu năng tối ưu của mạng là thông lượng và độ trễ, trong đó các tệp dữ liệu có kích thước lớn được phân mảnh thành các mảnh tin có kích thước tương đương nhau, đủ nhỏ để truyền đi hoàn toàn giữa các thực thể

mạng trong mỗi khe thời gian. Xuất phát từ các yếu tố thực tế là cần xem xét tới sự ảnh hưởng của các tham số kích thước bộ nhớ lưu trữ chia sẻ của các trạm gốc thông tin, kích thước tệp dữ liệu đối với hiệu năng mạng tối ưu trong mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu, luận án này lựa chọn hướng nghiên cứu **“Nghiên cứu tối ưu hóa thông lượng và độ trễ trong mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu”**.

Với cơ sở là những khảo sát các công trình nghiên cứu trước đó cùng với xem xét những yếu tố thực tế, luận án này được thực hiện nghiên cứu với hai đóng góp chính như sau:

- Thực hiện nghiên cứu trên mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu theo mô hình dòng chảy (fluid), từ đó đưa ra giải pháp tối ưu hóa thông lượng và độ trễ của mô hình mạng đề xuất.

Mô hình mạng được đưa ra để thực hiện nghiên cứu xuất phát từ thực tiễn và có những đặc điểm mới và chưa được tiến hành nghiên cứu trước đó. Cụ thể, thay vì không xem xét đến vai trò của các trạm gốc thông tin khi sử dụng kỹ thuật đệm dữ liệu, hoặc có xem xét nhưng giả sử rằng các trạm gốc thông tin có độ lớn của bộ nhớ lưu trữ chia sẻ không giới hạn, luận án này sẽ xem xét sử dụng kỹ thuật đệm dữ liệu áp dụng cho các bộ nhớ lưu trữ chia sẻ có giới hạn của cả trạm gốc thông tin di động và các thiết bị di động người dùng. Các tham số chỉ độ lớn có giới hạn của các bộ nhớ lưu trữ này sẽ là yếu tố quan trọng khi tính toán tối ưu và đánh giá hiệu năng của mạng là thông lượng và độ trễ trong mô hình đó.

- Thực hiện nghiên cứu trên mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu áp dụng kỹ thuật phân mảnh tệp dữ liệu, từ đó đưa ra giải pháp tối ưu hóa thông lượng và độ trễ của mô hình mạng đề xuất.

Mô hình mạng được đưa ra để thực hiện nghiên cứu xuất phát từ thực tiễn và có những đặc điểm mới và chưa được tiến hành nghiên cứu trước đó. Cụ thể, thay vì cho rằng mỗi tệp dữ liệu có độ lớn lý tưởng đủ nhỏ để có thể truyền đi hoàn toàn giữa các thực thể mạng với nhau trong mỗi khe thời gian như các nghiên cứu trước đó, luận án này giả sử kích thước các tệp dữ liệu là rất lớn và cần phải phân mảnh thành các mảnh tin có kích thước đủ nhỏ để truyền đi hoàn toàn giữa các thực thể mạng trong mỗi khe thời gian. Kích thước của tệp dữ liệu sẽ là yếu tố quan trọng khi tính toán tối ưu và đánh giá hiệu năng của mạng là thông lượng và độ trễ.

Mục tiêu chính mà luận án hướng tới là đề xuất mô hình nghiên cứu ý nghĩa thực tế và hiệu quả của việc sử dụng phương pháp đệm dữ liệu cho mạng vô tuyến hướng nội dung dựa trên việc tối ưu và đánh giá hai tham số hiệu năng mạng là thông lượng và độ trễ truyền tin đối với hai mô hình đó. Trong đó, các yếu tố mới của các mô hình mạng đề xuất là giá trị của biến số dung lượng lưu trữ chia sẻ có giới hạn tại các trạm gốc thông tin của mô hình dòng chảy và giá trị của biến số kích thước tệp dữ liệu của mô hình áp dụng kỹ thuật phân mảnh tệp dữ liệu sẽ được xem xét và đánh giá chi tiết. Để đạt được mục tiêu chính này, luận án phải mô hình hóa toán học được các mô hình mạng đề xuất và các yếu tố đặc tính của mạng vô tuyến hướng nội

dung, đề xuất phương pháp định tuyến truyền tin giữa các thực thể mạng, tìm ra công thức tính thông lượng và độ trễ của mạng, từ đó xây dựng và giải được bài toán tối ưu hóa để tìm ra phương pháp lưu trữ dữ liệu tại bộ nhớ của các thực thể mạng để tối ưu hai tham số hiệu năng mạng là thông lượng và độ trễ.

Đối tượng nghiên cứu của luận án là hai tham số hiệu năng mạng là thông lượng và độ trễ tối ưu của hai mô hình mạng vô tuyến hướng nội dung được đề xuất, trong đó các nút mạng được trang bị các bộ nhớ trong, có thể chia sẻ dung lượng để lưu trữ dữ liệu, phục vụ cho nhu cầu truyền nhận thông tin trong mạng. Hai tham số hiệu năng mạng sẽ được tính toán và phân tích dựa trên các tham số mạng, đặc biệt là các tham số mới trong hai mô hình mạng đề xuất là biến số dung lượng lưu trữ chia sẻ có giới hạn tại các trạm gốc thông tin của mô hình dòng chảy và giá trị của biến số kích thước tệp dữ liệu của mô hình mạng áp dụng kỹ thuật phân mảnh tệp dữ liệu. Các kết quả phân tích và tính toán sẽ được kiểm chứng lại bởi phần mềm tính toán Mathematica hoặc Matlab trên máy tính.

Phạm vi nghiên cứu: Với mục tiêu của luận án là đánh giá ý nghĩa thực tế và hiệu quả của việc sử dụng kỹ thuật đệm dữ liệu cho mạng vô tuyến hướng nội dung, luận án này sẽ đề xuất mới hai mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu. Cụ thể, hiệu năng mạng là thông lượng và độ trễ của mạng sẽ được phân tích, đánh giá và tối ưu dựa trên các tham số mạng cùng sự ảnh hưởng của các biến số mới theo các mô hình mạng được đề xuất là biến số dung lượng lưu trữ chia sẻ có giới hạn tại các trạm gốc thông tin của mô hình dòng chảy và giá trị của biến số kích thước tệp dữ liệu của mô hình mạng áp dụng kỹ thuật phân mảnh tệp dữ

liệu. Đối với mỗi mô hình mạng đề xuất, phương pháp truyền tin phù hợp và các bước phân tích để tìm ra công thức tính thông lượng và độ trễ của mạng cũng sẽ có sự phức tạp khác biệt so với các nghiên cứu trước đây. Dựa trên các yếu tố đó, luận án sẽ đưa ra giải pháp tối ưu hiệu năng mạng nhờ vào việc phân tích và tính toán tìm ra số lượng bản sao của mỗi tập dữ liệu được lưu tại các bộ nhớ lưu trữ chia sẻ của các thực thể mạng. Các giả thuyết về cách thức lưu trữ các bản sao này cũng như các mối quan hệ khác giữa các tham số mạng sẽ được trình bày chi tiết trong các phân tích. Các giá trị tối ưu được tìm ra sẽ được kiểm chứng lại bởi các phần mềm tính toán Mathematica và Matlab. Kết quả nhận được sẽ giúp đưa ra những nguyên tắc và cách thức sử dụng kỹ thuật đệm dữ liệu trong mạng vô tuyến hướng nội dung sao cho hiệu năng mạng đạt được là tốt nhất. Đây là tiền đề để việc áp dụng kỹ thuật đệm dữ liệu trong mạng vô tuyến hướng nội dung được nghiên cứu sâu hơn với các nghiên cứu thử nghiệm, mô phỏng và ứng dụng thực tế hơn trong tương lai.

Phương pháp nghiên cứu chính được sử dụng trong luận án này là phương pháp phân tích. Dựa trên việc thu thập và khảo sát các công trình nghiên cứu khoa học đã được đăng tải trên các tạp chí và hội nghị khoa học chuyên ngành uy tín, từ đó phân tích điểm mạnh và điểm hạn chế của các nghiên cứu trước đối với những thay đổi và đòi hỏi của thực tiễn để tìm ra những vấn đề chưa được giải quyết ở các bài toán trước đây và tiến hành nghiên cứu. Các vấn đề được đặt ra khi thực hiện các nghiên cứu tại luận án này sẽ được giải quyết nhờ tham khảo và học tập các kỹ thuật, phương pháp phân tích và công cụ từ các công trình khoa học có liên quan, phù hợp với các hướng nghiên cứu đề xuất. Các kết quả phân tích toán học luôn được

kiểm chứng bởi các phần mềm tính toán máy tính có độ tin cậy và chính xác cao.

Với các mục tiêu và nội dung nghiên cứu đã nêu ở trên cùng các công trình khoa học đã được công bố, các kết quả nghiên cứu của luận án sẽ được bố cục thành ba chương với các nội dung chính như sau:

- **Chương 1: Tổng quan về vấn đề nghiên cứu**

Chương này trình bày về mô hình, các phần tử và nguyên lý hoạt động của mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu. Nội dung chính của Chương sẽ tập trung khảo sát các nghiên cứu liên quan tới hiệu năng mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu để từ đó tìm ra các hạn chế của các nghiên cứu trước đây và đề xuất hướng nghiên cứu mới có tính thực tế, phạm vi nghiên cứu cũng như phương pháp tiếp cận của luận án.

- **Chương 2: Tối ưu hóa thông lượng và độ trễ của mạng vô tuyến hướng nội dung sử dụng mô hình dòng chảy**

Chương này thực hiện nghiên cứu trên mô hình mạng sử dụng mô hình dòng chảy. Dựa trên mô hình mạng này đề xuất giao thức truyền tin phù hợp giữa các thực thể mạng, xây dựng các công thức để tính toán các tham số hiệu năng mạng là thông lượng và độ trễ dựa trên các tham số mạng, từ đó đưa ra được các bài toán tối ưu hóa, phân tích và giải các bài toán tối ưu này để tìm ra số lượng các bản sao được lưu trữ trong mạng của các tệp dữ liệu phù hợp sao cho hiệu năng mạng được tối ưu.

Các đóng góp của luận án trong chương này đã được công bố trong 01 bài báo tại Hội nghị khoa học quốc tế hàng đầu của ngành Lý thuyết thông

tin ISIT 2016 [IC1] và 01 bài báo đăng trên tạp chí quốc tế ISI (IEEE Access) [IJ1]. Ngoài ra, một số đóng góp khác của luận án trong chương này cũng được công bố trong các hội nghị khoa học KICS Summer 2015 [IC2], KICS Winter 2016 [IC3], và KICS Summer 2016 [IC4] được tổ chức tại Hàn Quốc.

- **Chương 3: Tối ưu hóa thông lượng và độ trễ của mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu**

Tương tự như Chương 2, Chương này thực hiện nghiên cứu trên mô hình mạng sử dụng phương pháp phân mảnh tệp dữ liệu. Dựa trên mô hình mạng này, đề xuất giao thức truyền tin phù hợp giữa các thực thể mạng, xây dựng các công thức để tính toán các tham số hiệu năng mạng là thông lượng và độ trễ dựa trên các hệ số mạng, từ đó đưa ra được các bài toán tối ưu hóa, phân tích và giải các bài toán tối ưu này để tìm ra số lượng các bản sao được lưu trữ trong mạng của các tệp dữ liệu phù hợp sao cho hiệu năng mạng đạt được tối ưu.

Các đóng góp của luận án trong chương này đã được công bố trong 02 bài báo khoa học đăng trên Tạp chí Khoa học và Công nghệ - Đại học Đà Nẵng [DJ1] và Tạp chí Khoa học công nghệ và Thông tin - Học viện Công nghệ Bưu chính Viễn thông [DJ2], và 01 bài báo tại hội nghị khoa học quốc tế ATC 2017 [IC5].

- **Kết luận**

Trong phần này, luận án tóm tắt các kết quả nghiên cứu chính của luận án cùng với những bàn luận xung quanh đóng góp mới cả về ưu điểm và hạn chế. Từ đó đưa ra những gợi mở về các hướng nghiên cứu tiếp theo.

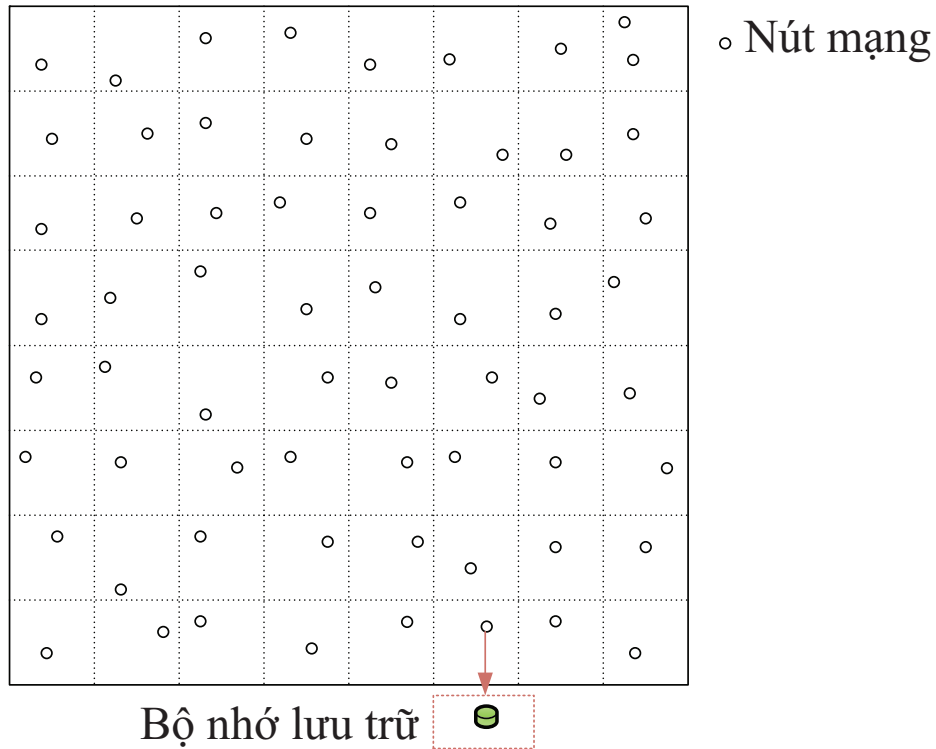
Chương 1

TỔNG QUAN VỀ MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG

***Giới thiệu chung:** Nội dung của Chương trình bày về mô hình và các thành phần mạng vô tuyến hướng nội dung được xem xét và nghiên cứu trong luận án. Trên cơ sở khảo sát các công trình nghiên cứu liên quan đến mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu đề xuất, luận án sẽ có những đánh giá và nhận xét để từ đó tìm ra các hạn chế của các nghiên cứu trước đây và đề xuất hướng nghiên cứu và tiếp cận của luận án.*

1.1. Mô hình mạng vô tuyến hướng nội dung

Kỹ thuật đệm dữ liệu, cho phép các nút mạng có thể lưu trữ dữ liệu tại các bộ nhớ lưu trữ của mình và chia sẻ cho các nút mạng khác trong toàn hệ thống cung cấp một phương pháp hỗ trợ truyền tải thông tin mới, đột phá, và hứa hẹn giúp làm giảm các sự cố nghẽn mạng tại các hệ thống truyền tin vô tuyến thế hệ mới. Các luồng dữ liệu thay vì được truyền tải qua các hạ tầng mạng thông tin truyền thống về thiết bị đầu cuối đích sẽ được truyền tải và chia sẻ bởi các thiết bị đầu cuối khác trong cùng hệ thống đang lưu trữ những dữ liệu này trong bộ nhớ chia sẻ của mình. Các thực thể mạng phổ biến có thể tham gia chia sẻ bộ nhớ lưu trữ của mình cho mục đích đệm dữ liệu điển hình là các điểm truy cập mạng, bộ phát wifi, các trạm gốc thông tin di động [16,36], thiết bị chuyển tiếp, hoặc các thiết bị di động người



Hình 1.1: Mô hình mạng vô tuyến hướng nội dung.

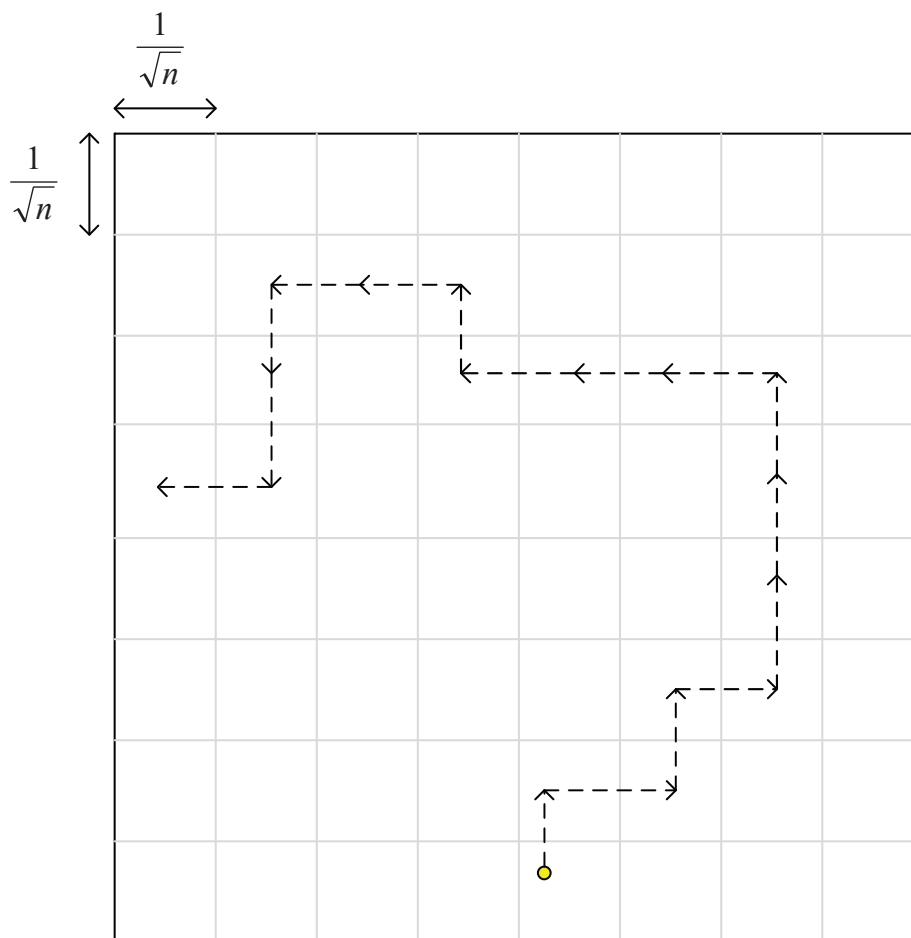
dùng [1, 33, 34, 43, 46]. Các hệ thống thực tế có thể áp dụng kỹ thuật đệm dữ liệu đang thu hút được nhiều quan tâm nghiên cứu là mạng thông tin di động với các thực thể là các trạm gốc thông tin di động và các thiết bị đầu cuối di động; mạng người dùng ngang hàng dạng mạng cảm biến; và mạng ad hoc. Việc sử dụng kỹ thuật đệm dữ liệu cho phần truy nhập vô tuyến hứa hẹn đem lại những hỗ trợ to lớn, giúp giảm nghẽn mạng tại các thời điểm cao điểm và nâng cao hiệu năng, duy trì sự hoạt động ổn định của hệ thống.

Các nghiên cứu của luận án này sẽ tập trung quan tâm nghiên cứu mô hình mạng vô tuyến di động hướng nội dung với sự tham gia của n thiết bị người dùng di động (xem hình 1.1). Đây cũng là mô hình mạng cơ bản được sử dụng tại các nghiên cứu trước [1, 15, 33, 34, 36, 43, 46] với n thiết bị di động người dùng được phân bố một cách độc lập và ngẫu nhiên trên mô hình mạng

Torus có kích thước đơn vị chuẩn (a unit-sized torus). Đây cũng là mô hình mạng được sử dụng phổ biến trong các nghiên cứu đối với những mạng có số lượng thực thể mạng vô cùng lớn (tiến đến vô cùng) do khả năng dễ áp dụng và mô hình hóa các tính chất và đặc điểm cơ bản của mạng để thực hiện các phân tích toán học. Trong mô hình mạng nghiên cứu, mỗi thiết bị di động và trạm gốc thông tin di động (nếu có) được giả sử rằng đều được trang bị các bộ nhớ lưu trữ có thể được sử dụng để lưu trữ một phần dữ liệu trong thư viện mạng, trong đó kích thước dữ liệu của thư viện mạng có thể chứa M tệp dữ liệu. Mỗi tệp dữ liệu được giả sử rằng có độ lớn tương đương nhau.

Trước hết, tính chất di động của các thiết bị người dùng được mô hình hóa theo mô hình bước đi ngẫu nhiên (random walks mobility model - RWMM) trên mô hình mạng Torus kích thước đơn vị chuẩn đã được mô tả tại [1, 23]. Mô hình bước đi ngẫu nhiên là mô hình có nhiều ưu điểm nhờ sự đơn giản và phù hợp với các bài toán thống kê, từ đó cho phép thực hiện các phân tích có tính chất phức tạp liên quan đến việc tính toán độ trễ, tham số hiệu năng mạng phụ thuộc đặc biệt vào các mô hình chuyển động. Chi tiết hơn, mạng Torus có diện tích đơn vị chuẩn sẽ được chia thành n tế bào nhỏ với diện tích $\frac{1}{n}$. Mỗi thiết bị di động sẽ thực hiện các bước đi đơn ngẫu nhiên một cách độc lập với độ dài bước đi là $\frac{1}{\sqrt{n}}$ ở trong mạng có $\sqrt{n} \times \sqrt{n}$ tế bào con sao cho mỗi thiết bị di động có thể di chuyển tới một trong 4 tế bào lân cận sau mỗi khe thời gian (xem hình 1.2).

Giả sử thời gian thực được phân chia thành các khe thời gian $t = 0, 1, 2, \dots$. Một cách độc lập, tất cả các nút mạng sẽ có những yêu cầu tải dữ liệu khác nhau ở mỗi khe thời gian t . Khi các yêu cầu tại mỗi thiết bị người dùng được tổng hợp lại với số lượng không giới hạn trong một khoảng thời



Hình 1.2: Mô hình bước đi ngẫu nhiên (RWMM).

gian dài, ta sẽ dễ dàng nhận được xác suất tải p_m của mỗi tệp dữ liệu $m \in \mathcal{M} \triangleq \{1, \dots, M\}$, trong đó M là số lượng tệp dữ liệu trong thư viện mạng. Khi đó, ta có thể nói rằng, tại mỗi khe thời gian, mỗi thiết bị người dùng sẽ có nhu cầu tải tệp dữ liệu m với xác suất p_m một cách độc lập và không bị ảnh hưởng bởi các thiết bị người dùng hoặc thực thể mạng khác. Trong luận án này, tôi giả sử xác suất yêu cầu tải của các tệp dữ liệu trong mạng tuân theo phân bố Zipf. Phân bố Zipf đã được chỉ ra rằng cũng phù hợp sử dụng trong các nghiên cứu về mô hình mạng vô tuyến di động [22, 30, 37, 44, 51]. Đây là phân bố thường được sử dụng phổ biến trong các nghiên cứu với nhiều loại dữ liệu thời gian thực một cách chính xác như

dữ liệu web, dữ liệu chia sẻ, dữ liệu người dùng tạo ra và các dữ liệu hình ảnh theo yêu cầu [9].

Chúng ta chuyển sang xem xét các thủ tục truyền dữ liệu trong mạng vô tuyến hướng nội dung. Đây là bước cho phép các tệp dữ liệu được yêu cầu tải có thể được truyền tới các thiết bị người dùng tương ứng trong mạng qua kênh truyền vô tuyến. Trong bước truyền dữ liệu này, mỗi thiết bị người dùng thực hiện tải một tệp dữ liệu mong muốn từ một trong các thiết bị người dùng khác hoặc trạm gốc thông tin có chứa tệp dữ liệu được yêu cầu trong bộ nhớ lưu trữ của nó. Thuật toán định tuyến không dây cơ bản có thể định tuyến đơn chặng hoặc đa chặng dựa vào các thuộc tính liên kết khác nhau và giao thức định tuyến. Định tuyến đơn chặng đơn giản hơn định tuyến đa chặng ở vấn đề cấu trúc và thực hiện với việc dữ liệu được truyền đi trực tiếp từ điểm nguồn tới điểm đích trong phạm vi truyền tải không dây. Đối với định tuyến đa chặng, khi truyền các gói dữ liệu từ một nguồn của nó đến điểm đích, các gói dữ liệu sẽ được chuyển tiếp qua một hoặc nhiều nút trung gian trước khi điểm đích nhận được gói dữ liệu. Trên cơ sở tham khảo từ nghiên cứu trước đó [1], chi tiết về thủ tục truyền dữ liệu và định tuyến truyền tin sẽ được mô tả cụ thể, phù hợp với từng mô hình mạng nghiên cứu tại Chương 2 và Chương 3. Luận án sử dụng giao thức mẫu *protocol model* được trình bày ở [38] để xác định điều kiện tải dữ liệu thành công. Ký hiệu $d(u, v)$ là khoảng cách Euclidean giữa các thiết bị người dùng u và v . Theo đó, bước truyền tệp dữ liệu từ thiết bị u tới thiết bị v được cho là thành công khi và chỉ khi có $d(u, v) \leq r$ với r là giới hạn khoảng cách truyền tin và không có thiết bị truyền phát thông tin nào hoạt động trong bán kính $(1 + \Delta)r$ từ thiết bị v , với $\Delta > 0$ là tham số truyền dữ liệu.

Các nghiên cứu của luận án này tiến hành dựa trên cơ chế mạng cổ điển, trong đó mỗi yêu cầu tải tin của người dùng mạng phục vụ bởi một kết nối dữ liệu riêng biệt do một máy chủ trung tâm điều khiển (một thực thể quan trọng của mạng phân tán hướng nội dung được đặt tại trung tâm của mạng). Cơ chế hoạt động này được E. Nygren và cộng sự tại [13] giới thiệu, trong đó điều khiển trung tâm có khả năng đáp ứng hàng trăm tỷ kết nối Internet mỗi ngày và phục vụ cho hoạt động của hàng ngàn ứng dụng. Cơ chế này hoạt động như sau: Khi người dùng có yêu cầu tải tin trên thiết bị di động (ví dụ: đăng nhập vào trang web, click vào đường dẫn của youtube,...), thông tin mapping của người dùng sẽ được phân hệ định vị trung tâm của hệ thống nhận ra và hướng tới tất cả các thực thể mạng có lưu trữ tệp dữ liệu yêu cầu đó. Từ đó, tùy thuộc vào các định luật của mạng và thuật toán định tuyến của mạng mà dữ liệu sẽ được xác định truyền tải từ thực thể nào và truyền về thiết bị của người dùng. Như vậy, trong luận án này, tương tự như kỹ thuật đệm dữ liệu được sử dụng trong mô hình mạng hữu tuyến, vấn đề của phương pháp định tuyến bao gồm các thông tin sơ đồ mạng và phân bố dữ liệu trong các thực thể mạng cùng vị trí của chúng đã được tích hợp và trong suốt đối với mỗi thiết bị người dùng. Nhờ đó, khi có nhu cầu tải dữ liệu, mỗi thiết bị đầu cuối người dùng đều biết những thực thể nào đang lưu giữ tệp dữ liệu mong muốn trong bộ nhớ chia sẻ của chúng và lựa chọn thực thể đích để gửi đi tin yêu cầu tải dữ liệu.

1.2. Các tham số hiệu năng mạng và ký hiệu toán học sử dụng trong luận án

1.2.1. Các tham số hiệu năng mạng

Ở đây, các tham số hiệu năng mạng là thông lượng và độ trễ. Trong các hệ thống thông tin, thông lượng là lượng thông tin hữu ích được truyền đi trên mạng trong một đơn vị thời gian và đây là chỉ số quan trọng để đánh giá hiệu năng mạng. Độ trễ thể hiện sự chậm trễ phát sinh trong xử lý dữ liệu của mạng máy tính. Độ trễ càng nhỏ thì tốc độ mạng càng nhanh và độ trễ càng ít và ngược lại, độ trễ càng cao thì tốc độ mạng càng chậm và thời gian trễ càng nhiều. Các tham số thông lượng và độ trễ của mạng được định nghĩa như sau:

Định nghĩa 1.2.1 (Thông lượng). *Thông lượng của mạng là giá trị trung bình của dung lượng dữ liệu mà thiết bị người dùng nhận được trong một khe thời gian.*

Định nghĩa 1.2.2 (Độ trễ). *Độ trễ là thời gian trung bình tính từ thời điểm thiết bị nguồn gửi đi bản tin đầu tiên yêu cầu tải thông tin cho đến khi thiết bị nguồn này nhận được đủ tệp dữ liệu mong muốn.*

Tùy thuộc vào các mô hình mạng đề xuất ở mỗi nghiên cứu khác nhau, công thức tính thông lượng và độ trễ tổng quát sẽ được xây dựng là hàm số biến thiên theo số lượng thiết bị đầu cuối người dùng trong mạng.

1.2.2. Ký hiệu toán học sử dụng trong luận án

Giả thiết số lượng thiết bị đầu cuối người dùng n vô cùng lớn ($+\infty$), luận án này sử dụng các ký hiệu toán học và tính toán thông lượng và độ trễ theo độ phức tạp thuật toán như sau:

i) $f(x) = O(g(x))$: tồn tại các hằng số C và c sao cho $f(x) \leq Cg(x)$ với $\forall x > c$;

ii) $f(x) = o(g(x))$: tương đương $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$;

iii) $f(x) = \Omega(g(x))$: tương đương $g(x) = O(f(x))$;

iv) $f(x) = \omega(g(x))$: tương đương $g(x) = o(f(x))$;

v) $f(x) = \Theta(g(x))$: tương đương $f(x) = O(g(x))$ và $f(x) = \Omega(g(x))$ [12].

Cụ thể hơn, $f(x) = O(g(x))$ có nghĩa là $g(x)$ biến thiên với tốc độ nhanh hơn hoặc bằng với $f(x)$. Tương tự, $f(x) = o(g(x))$ có nghĩa là $g(x)$ biến thiên với tốc độ nhanh hơn hẳn $f(x)$. Và cuối cùng, $f(x) = \Omega(g(x))$ có nghĩa là $f(x)$ và $g(x)$ có cùng tốc độ biến thiên tăng.

1.3. Các công trình nghiên cứu khoa học liên quan

Hiện nay, kỹ thuật đệm dữ liệu đã được một số nhóm nghiên cứu của các đơn vị trong nước quan tâm nghiên cứu như Đại học Tôn Đức Thắng, Đại học Quốc tế TP HCM, Học viện Công nghệ Bưu chính Viễn thông [41, 50]. Tuy nhiên, các nghiên cứu này hiện nay mới chỉ đang áp dụng trên mạng hữu tuyến hướng nội dung. Việc áp dụng kỹ thuật đệm dữ liệu cho mạng vô tuyến hướng nội dung vẫn chưa được quan tâm nghiên cứu.

Các nghiên cứu về hiệu năng mạng vô tuyến

Gupta và Kumar đã nghiên cứu thông lượng mạng vô tuyến [38] và chỉ ra thông lượng tại mỗi nút mạng nhận được là $\Theta\left(\frac{1}{\sqrt{n \log n}}\right)$ trong mạng vô tuyến ad hoc tĩnh sử dụng phương pháp định tuyến đa chặng, trong đó n là số lượng nút mạng trên toàn mạng. Cho tới nay đã có rất nhiều các nghiên cứu quan tâm đến việc sử dụng phương pháp truyền dẫn đa chặng [14, 31, 39, 55], trong đó đã chứng minh được rằng thông lượng tại mỗi nút mạng luôn có

giá trị biến thiên chậm hơn $\Theta(1)$. Bên cạnh sử dụng kỹ thuật truyền dẫn đa chặng, nhiều kỹ thuật khác cũng đã được nghiên cứu, áp dụng để cải thiện thông lượng trung bình tại mỗi nút mạng đạt tới giá trị $\Theta(1)$ như kỹ thuật hợp tác phân cấp [6], tính di động của thiết bị [1, 32], sử dụng anten có hướng [17, 18, 40] và hỗ trợ cơ sở [8, 56].

Các nghiên cứu về hiệu năng mạng vô tuyến hướng nội dung

Không sử dụng giả thuyết là các cặp truyền dẫn nguồn - đích (S-D) được thiết lập trước và không thay đổi trong mô hình mạng vô tuyến truyền thống, việc nghiên cứu kỹ thuật đệm dữ liệu trong mạng vô tuyến hướng nội dung có khá nhiều điểm thách thức giới nghiên cứu. Với việc các tệp dữ liệu được lưu trữ tại các bộ nhớ đệm của số lượng lớn các thiết bị di động người dùng trong mạng, việc tìm kiếm thiết bị di động gần nhất có lưu trữ tệp dữ liệu mong muốn để gửi tin nhắn yêu cầu truyền tin, lập kế hoạch và tìm đường truyền tin tối ưu nhất là những điểm tối quan trọng trong vấn đề cải thiện hiệu năng của mạng. Vấn đề này đã thu hút được nhiều quan tâm nghiên cứu [15, 33, 36, 43, 46, 58].

Giải pháp ứng dụng kỹ thuật đệm dữ liệu trong mạng ad hoc tính hướng nội dung cũng đã được nghiên cứu tại [43, 46]. Trong đó, cơ chế lưu trữ bộ nhớ đệm xác định và tập trung đã được trình bày tại [43], trong đó các bản sao của tệp dữ liệu được lưu trữ sẵn trong mạng dựa trên mức độ phổ biến của các tệp dữ liệu theo mô hình tập trung và xác định. Cơ chế lưu trữ bộ nhớ đệm phân tán và ngẫu nhiên áp dụng phương pháp truyền dẫn đa chặng cục bộ đã được giới thiệu tại [46], trong đó các tệp dữ liệu được gán một cách độc lập và ngẫu nhiên tại bộ nhớ đệm của tất cả các thiết bị người dùng

trong mạng. Các kết quả nghiên cứu đã chỉ ra rằng biến thiên thông lượng sử dụng phương pháp truyền dẫn đa chặng có ưu điểm vượt trội hơn so với phương pháp đơn chặng ở các nghiên cứu [33, 34]. Ngoài ra, với trường hợp hệ số phân bố Zipf lớn hơn giới hạn ngưỡng (> 1), xóa bỏ các tệp dữ liệu không phổ biến trong mạng có thể vừa giảm dung lượng thư viện mạng lưu trữ và đồng thời cải thiện được hiệu năng mạng.

Trong mạng vô tuyến di động, người ta tính hiệu năng của mạng là thông lượng và độ trễ theo mô hình di động reshuffling, trong đó, vị trí của mỗi thiết bị di động được thay đổi một cách độc lập theo kỹ thuật bước đi đi ngẫu nhiên (random walks) với độ dài của mỗi bước đi có thể thay đổi và cập nhật lại tại thời điểm bắt đầu của mỗi khe thời gian [36]. Alfano và cộng sự [15] đã chỉ ra là tăng mức độ di động của thiết bị người dùng sẽ dẫn tới sự suy giảm của hiệu năng mạng, trong đó cơ chế đặt các tệp dữ liệu tại các thiết bị di động được sử dụng tương tự như [43] là xác định. Ở [5], xem xét đối với hiệu năng mạng như trên đã được mở rộng sang trường hợp kích thước của mỗi tệp dữ liệu là rất lớn và do đó cần phân mảnh tệp dữ liệu thành những mảnh tin có kích cỡ nhỏ vừa đủ để có thể truyền đi hoàn toàn trong mỗi khe thời gian của mạng. Ở [58] sử dụng mô hình di động có tương quan, trong đó các thiết bị di động được nhóm lại thành các cụm khác nhau và sự di chuyển của các thiết bị nằm trong cùng một cụm có tính liên quan, ảnh hưởng với nhau, mối tương quan trong di chuyển của các thiết bị di động ảnh hưởng tới hiệu năng mạng.

Nghiên cứu mạng vô tuyến hướng nội dung cũng được mở rộng sang trường hợp hạ tầng mạng hỗn hợp sử dụng phương pháp truyền tin đa chặng tại [16, 36]. Tại nghiên cứu [36], bên cạnh các thiết bị người dùng vô tuyến, giả

sử các trạm gốc thông tin được phân bố đều trong mạng có bộ nhớ chia sẻ lưu trữ toàn bộ các tệp dữ liệu của thư viện mạng thông qua kết nối liên tục với đường truyền mạng lõi để đáp ứng nhu cầu dữ liệu của các thiết bị trong mạng. Kết quả nghiên cứu chỉ ra rằng, khi tỷ lệ số lượng các trạm gốc thông tin trong mạng lớn hơn thiết bị di động và số lượng tệp dữ liệu trong mạng, lúc đó mạng hỗn hợp vô tuyến hướng nội dung có hiệu năng mạng lớn hơn so với trường hợp không sử dụng trạm gốc thông tin di động cho kỹ thuật đệm dữ liệu.

Các nhà khoa học còn chú ý đến mạng vô tuyến hướng nội dung có sử dụng mã hóa tệp dữ liệu [26–28] trong đó các thiết bị phát tin có thể đồng thời đáp ứng nhiều yêu cầu tải tin trong mạng sử dụng phương pháp truyền tin đa chặng mã hóa. Với cách này, có thể tối ưu hóa bước lưu trữ dữ liệu tại các nút mạng sao cho kỹ thuật truyền tin đa chặng được phát huy hiệu quả nhất, đáp ứng được đồng thời nhiều yêu cầu tải tin nhất ở bước truyền tin.

1.4. Nhận xét về công trình nghiên cứu của các tác giả khác và hướng nghiên cứu của luận án

1.4.1. Nhận xét về công trình nghiên cứu của các tác giả khác

Các nghiên cứu trước đây chưa đề cập đến việc sử dụng kỹ thuật đệm dữ liệu với các trạm gốc thông tin hoặc giả sử rằng các trạm gốc thông tin được phân bố đều trong mạng được có bộ nhớ chia sẻ lưu trữ toàn bộ các tệp dữ liệu của thư viện mạng thông qua kết nối liên tục với đường truyền mạng lõi để đáp ứng nhu cầu dữ liệu của các thiết bị trong mạng [36]. Trên thực tế, các trạm gốc thông tin di động thường cũng được trang bị các bộ nhớ lưu trữ lớn để phục vụ cho việc lưu trữ thông tin và hoạt động của hệ thống. Các bộ

nhớ này thậm chí có dung lượng lớn hơn nhiều bộ nhớ của các thiết bị đầu cuối và cũng có thể sử dụng để lưu trữ các dữ liệu chia sẻ cho người dùng sử dụng kỹ thuật đệm dữ liệu. Với vùng phủ sóng lớn có thể đáp ứng cho nhiều thuê bao cùng lúc, trong những trường hợp nghẽn mạng hoặc đường truyền mạng lõi có vấn đề, việc áp dụng kỹ thuật đệm dữ liệu sử dụng bộ nhớ lưu trữ chia sẻ của các trạm gốc thông tin di động cùng với các bộ nhớ lưu trữ chia sẻ của người dùng có thể góp phần đảm bảo việc đáp ứng nhu cầu thông tin của các thiết bị người dùng trong mạng, duy trì sự ổn định của mạng lưới. Như vậy, các nghiên cứu trước đó chưa thực hiện với trường hợp đồng thời bộ nhớ lưu trữ chia sẻ của trạm gốc thông tin và thiết bị đầu cuối di động cùng được sử dụng kỹ thuật đệm dữ liệu, cung cấp nội dung cho toàn mạng. Giải quyết được bài toán này là điều cần thiết và sẽ được thực hiện nghiên cứu trong luận án này.

Ngoài ra, trong các nghiên cứu trước đây [15, 33, 36, 43, 46, 58], mô hình mạng nghiên cứu sử dụng là mô hình dòng chảy, trong đó thời gian hoạt động của mạng được phân chia thành các khe thời gian và mỗi tệp dữ liệu có kích thước đủ nhỏ sao cho có thể được truyền đi hoàn toàn giữa các nút mạng lân cận với nhau trong khoảng thời gian là mỗi khe thời gian. Tuy nhiên, trong thực tế khi các nút mạng di động với tốc độ đủ lớn, tương quan độ lớn của mỗi khe thời gian và độ lớn của mỗi tệp dữ liệu bị thay đổi khiến cho mỗi tệp dữ liệu khó có thể truyền đi hoàn toàn trong khoảng thời gian là mỗi khe thời gian. Do đó, đề xuất tại [5] với mỗi tệp dữ liệu có thể được chia nhỏ thành nhiều mảnh tin nhỏ có kích thước phù hợp để truyền đi hoàn toàn trong khoảng thời gian của mỗi khe thời gian là phù hợp hơn với thực tế mạng vô tuyến và cần có những nghiên cứu sâu hơn để đánh giá thêm.

Nghiên cứu trên mô hình mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu như trên đây là chưa được xem xét trên các công trình khoa học trước đó và sẽ được thực hiện trong luận án này.

Trong các nghiên cứu về việc sử dụng kỹ thuật đệm dữ liệu cho mạng vô tuyến hướng nội dung, mỗi giả thiết mới trên mô hình mạng nghiên cứu sẽ phát sinh thêm một tham số mạng phụ thuộc và khiến cho độ phức tạp của các lời giải thay đổi, từ đó có thể phát triển thêm nhiều phân tích tối ưu và kết quả ở nhiều góc độ khác nhau, góp phần xây dựng bức tranh về mô hình mạng vô tuyến sử dụng kỹ thuật đệm dữ liệu được đầy đủ hơn.

1.4.2. Hướng nghiên cứu của luận án

Nghiên cứu [45] đã chỉ ra rằng dung lượng bộ nhớ lưu trữ chia sẻ sử dụng cho kỹ thuật caching là yếu tố quan trọng ảnh hưởng tới hiệu năng mạng. Do đó, ngoài ra, thay vì không xem xét đến vai trò của các trạm gốc thông tin khi sử dụng kỹ thuật đệm dữ liệu như những nghiên cứu trước đây, phát triển mô hình mạng hỗn hợp tại [36] khi sử dụng kỹ thuật đệm dữ liệu với các trạm gốc thông tin có độ lớn của bộ nhớ lưu trữ chia sẻ không giới hạn, luận án này sẽ xem xét sử dụng kỹ thuật đệm dữ liệu áp dụng cho các bộ nhớ lưu trữ chia sẻ có giới hạn của cả trạm gốc thông tin di động và các thiết bị di động người dùng. Với kênh truyền tin và cơ chế truyền tin được xác định, kích thước tệp dữ liệu được truyền đi tại mỗi khe thời gian cũng là tham số ảnh hưởng đến khả năng truyền tin thành công của mạng. Các tham số chỉ độ lớn có giới hạn của các bộ nhớ lưu trữ có giới hạn tại các trạm gốc thông tin và thiết bị người dùng sẽ là yếu tố khác biệt so với các nghiên cứu trước đây. Để thuận lợi cho việc phân tích, mô hình *dòng chảy* ở [1] sẽ được

sử dụng để nghiên cứu. Chi tiết hơn, kích cỡ của mỗi tệp dữ liệu được giả sử là vô cùng nhỏ sao cho thời gian cần thiết để truyền một tệp dữ liệu giữa một thiết bị người dùng và thiết bị người dùng lân cận hoặc trạm gốc thông tin trong cùng cụm tế bào truyền tin nhỏ hơn nhiều khoảng thời gian của một khe thời gian. Nhờ đó, dữ liệu được gửi đi từ một thiết bị trong một khe thời gian có thể tương đương với nhiều tệp dữ liệu và do đó tất cả các tệp dữ liệu của thiết bị sẽ đảm bảo được truyền đi hết trong một khe thời gian. Tuy nhiên, thiết bị nhận được mỗi tệp dữ liệu trong khe thời gian này không được phép truyền đi tiếp cho đến khe thời gian tiếp theo. Vì lẽ đó, xây dựng phương pháp định tuyến truyền dữ liệu phù hợp với mô hình mạng đề xuất, tính toán hiệu năng mạng là thông lượng và độ trễ, xây dựng và giải bài toán tối ưu để tìm ra số lượng các bản ghi tối ưu của mỗi tệp dữ liệu được phân bố trong mạng sao cho hiệu năng mạng nhận được tốt nhất là những vấn đề đặt ra và cần được giải quyết trong luận án này.

Ngoài ra, thay vì cho rằng mỗi tệp dữ liệu có độ lớn lý tưởng đủ nhỏ để có thể truyền đi hoàn toàn giữa các thực thể mạng với nhau trong mỗi khe thời gian như các nghiên cứu trước đó, luận án này giả sử kích thước các tệp dữ liệu là rất lớn và không thể truyền đi hoàn toàn giữa các thực thể mạng trong mỗi khe thời gian. Kích thước của tệp dữ liệu sẽ là tham số quan trọng khi tính toán tối ưu và đánh giá hiệu năng của mạng là thông lượng và độ trễ. Để quá trình truyền nhận dữ liệu trong mạng được đảm bảo, thay vì sử dụng mô hình dòng chảy, hướng nghiên cứu này sẽ sử dụng kỹ thuật phân mảnh tệp dữ liệu (*segmentation*) trong đó mỗi tệp dữ liệu được phân mảnh thành nhiều mảnh tin khác nhau và có kích thước bằng nhau sao cho mỗi mảnh tin có thể được truyền đi hoàn toàn trong một khe thời gian từ thiết

bị nguồn tới một trong những thiết bị mạng lân cận. Nhờ đó, các tệp dữ liệu sẽ đảm bảo được truyền đi hết bởi thiết bị đó trong một khe thời gian. Tuy nhiên, mỗi mảnh tin nhận được bởi thiết bị trong khe thời gian này không được phép truyền đi tiếp cho đến khe thời gian tiếp theo. Thiết bị người dùng đích được cho là nhận được hoàn toàn tệp dữ liệu khi và chỉ khi nhận được đầy đủ số lượng mảnh tin cần thiết để tái tạo hoặc phục hồi lại tệp dữ liệu ban đầu. Xây dựng phương pháp định tuyến truyền dữ liệu phù hợp với mô hình mạng đề xuất, tính toán thông lượng và độ trễ, xây dựng và giải bài toán tối ưu để tìm ra số lượng các bản ghi tối ưu của mỗi tệp dữ liệu được phân bố trong mạng sao cho hiệu năng mạng nhận được là tốt nhất là những vấn đề đặt ra và cần được giải quyết trong luận án này.

Như vậy, để góp phần đưa ra góc nhìn đầy đủ hơn về việc sử dụng kỹ thuật đệm dữ liệu trong mạng vô tuyến hướng nội dung, hướng nghiên cứu được đề xuất trong luận án này là:

- Thực hiện nghiên cứu trên mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu theo mô hình dòng chảy, trong đó kỹ thuật đệm dữ liệu áp dụng cho các trạm gốc thông tin di động và các thiết bị người dùng. Khi đó, dung lượng lưu trữ chia sẻ của các trạm gốc thông tin giả sử lớn hơn dung lượng lưu trữ chia sẻ của các thiết bị người dùng, và đây cũng là một tham số ảnh hưởng đến hiệu năng của mạng. Từ mô hình mạng đề xuất, luận án này sẽ đưa ra giải pháp tối ưu hóa thông lượng và độ trễ của mạng phù hợp. Với kết quả này làm cho thông lượng mạng tăng lên, độ trễ mạng giảm xuống so với giải pháp cũ.
- Thực hiện nghiên cứu trên mô hình mạng vô tuyến hướng nội dung sử

dụng kỹ thuật đệm dữ liệu áp dụng kỹ thuật phân mảnh tệp dữ liệu. Trong đó, thay vì sử dụng mô hình dòng chảy như các nghiên cứu trước với giả thuyết là kích thước của các tệp dữ liệu lưu trữ rất nhỏ và có thể truyền đi hoàn toàn giữa các thực thể mạng với nhau trong mỗi khe thời gian, kích thước các tệp dữ liệu giả sử rất lớn và cần phải phân mảnh thành những mảnh tin có kích thước đủ nhỏ để truyền đi hoàn toàn giữa các thực thể mạng trong mỗi khe thời gian. Khi đó, kích thước của các tệp dữ liệu sẽ được xem xét như là một tham số mạng quan trọng và ảnh hưởng đến hiệu năng của mạng. Từ mô hình mạng đề xuất, luận án sẽ trình bày giải pháp tối ưu hóa thông lượng và độ trễ của mạng tương ứng. Áp dụng giải pháp này làm tăng các loại hình dịch vụ dữ liệu mà chất lượng mạng vẫn đảm bảo.

Đối với mỗi hướng nghiên cứu, do các giả thuyết khác nhau nên các tham số mạng được xem xét khác nhau và có những ảnh hưởng nhất định đối với hiệu năng của mạng là thông lượng và độ trễ. Dựa trên mô hình mạng được đưa ra và các tham số mạng ở mỗi hướng nghiên cứu, luận án sẽ phân tích và tính toán đưa ra bài toán tối ưu đối với thông lượng và độ trễ của mạng, từ đó có những so sánh và nhận định về mức độ ảnh hưởng và sự thay đổi của hiệu năng mạng ở mỗi trường hợp. Các phân tích và kết quả tính toán sẽ được kiểm tra lại bởi các kết quả được giải bằng các chương trình toán học trên máy tính như Matlab hoặc Mathematica để đảm bảo tính đúng đắn của các kết quả nghiên cứu.

1.5. Kết luận Chương 1

Nội dung Chương 1 đã trình bày khái quát về mô hình, nguyên lý hoạt động, các tham số mạng và các yếu tố ảnh hưởng lên các tham số hiệu năng mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu. Các công trình nghiên cứu liên quan đến mạng vô tuyến hướng nội dung nói chung và hiệu năng mạng sử dụng kỹ thuật đệm dữ liệu nói riêng cũng đã được khảo sát, phân tích, đánh giá trong chương này. Qua đó, các hạn chế của các nghiên cứu trước đây về mô hình mạng vô tuyến hướng nội dung, các tham số và yếu tố ảnh hưởng tới hiệu năng mạng đã được chỉ ra. Trên cơ sở là các yếu tố thực tế của mạng vô tuyến hướng nội dung hiện nay, luận án đã đưa ra hai hướng nghiên cứu chính khi xem xét những yếu tố thực tế của mạng vô tuyến sử dụng kỹ thuật đệm dữ liệu là sự hiện diện của các trạm gốc thông tin đối với mô hình mạng hỗn hợp và ứng dụng kỹ thuật phân mảnh tệp dữ liệu khi kích thước mỗi tệp dữ liệu đủ lớn để không thể truyền đi hoàn toàn trong mỗi khe thời gian.

Chương 2

TỐI ƯU HÓA THÔNG LƯỢNG VÀ ĐỘ TRỄ CỦA MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG SỬ DỤNG MÔ HÌNH DÒNG CHẢY

Giới thiệu chung: Từ ý tưởng của nghiên cứu [36] với mô hình mạng vô tuyến hỗn hợp sử dụng kỹ thuật đệm dữ liệu có các trạm gốc thông tin được phân bố đều trong mạng với bộ nhớ lưu trữ toàn bộ các tệp dữ liệu của thư viện mạng thông qua kết nối liên tục với đường truyền mạng lõi để đáp ứng nhu cầu dữ liệu của các thiết bị trong mạng, nội dung nghiên cứu tại Chương này sẽ thực hiện với các trạm gốc sử dụng bộ nhớ lưu trữ có giới hạn áp dụng kỹ thuật đệm dữ liệu. Thực tế chỉ ra rằng, khi kết nối đường truyền mạng lõi bị nghẽn, trạm gốc thông tin không thể nhận dữ liệu thông tin từ thư viện mạng, tức là các trạm gốc thông tin trong mô hình mạng tại [36] không thể thực hiện kỹ thuật đệm dữ liệu. Mô hình mạng, trong đó các trạm gốc thông tin sử dụng bộ nhớ ngoài có giới hạn của mình để lưu trữ dữ liệu và chia sẻ cho các thiết bị người dùng rõ ràng có ý nghĩa thực tế và cần được nghiên cứu thêm. Việc thêm tham số mạng là dung lượng bộ nhớ lưu trữ chia sẻ thông tin của các trạm gốc sẽ dẫn đến việc tối ưu hóa thông lượng và độ trễ mạng trở nên phức tạp và thách thức hơn.

Ở chương này, mô hình nghiên cứu mạng vô tuyến hỗn hợp hướng nội dung được đề xuất, trong đó mỗi thiết bị di động người dùng di chuyển trong mạng theo mô hình bước đi ngẫu nhiên (RWMM) và đưa ra yêu cầu tải tệp dữ liệu

nằm trong thư viện mạng một cách độc lập và ngẫu nhiên. Đồng thời, các trạm gốc thông tin di động được đặt cố định tại các vị trí phân bố đều trong mạng. Để đáp ứng các yêu cầu tải tệp dữ liệu của các thiết bị người dùng, các thiết bị người dùng và các trạm gốc thông tin đều được trang bị các bộ nhớ ngoài có khả năng lưu trữ các tệp dữ liệu để chia sẻ cho các thiết bị người dùng khác trong mạng. Dựa trên mô hình mạng đề xuất, chương này sẽ đưa ra giải pháp tối ưu hóa thông lượng và độ trễ của mạng được tóm tắt lại với trình tự như sau:

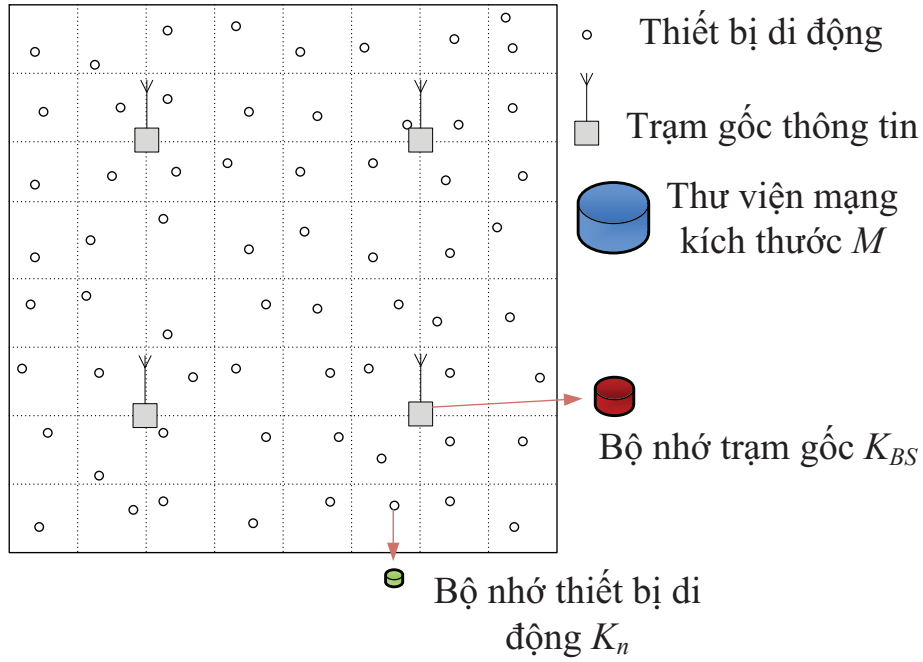
- Đề xuất mô hình mạng trong đó cả các trạm gốc thông tin di động và thiết bị di động người dùng đều có khả năng lưu trữ các tệp dữ liệu dữ liệu trong mạng với các dung lượng lưu trữ khác nhau.
- Từ mô hình mạng đề xuất sẽ đưa ra phương pháp định tuyến truyền tin trong mạng, cách thức tính toán thông lượng và độ trễ mạng.
- Để nhận được thông lượng và độ trễ tối ưu, số lượng bản sao của mỗi tệp dữ liệu trong thư viện mạng và phương pháp lưu trữ các bản sao này tại bộ nhớ ngoài của thiết bị di động người dùng và trạm gốc thông tin di động cần được tối ưu hóa. Kết quả phân tích thu được cho thấy khi tổng dung lượng lưu trữ tại các trạm gốc thông tin di động lớn hơn tổng dung lượng lưu trữ tại các thiết bị di động, các tệp dữ liệu phổ biến nhất trong mạng sẽ được ưu tiên lưu trữ tại bộ nhớ ngoài của các thiết bị người dùng trong khi các tệp dữ liệu ít phổ biến trong mạng sẽ được lưu trữ tại bộ nhớ của các trạm gốc thông tin di động để phục vụ yêu cầu tải tin của người dùng.
- Các kết quả phân tích và tính toán sẽ được kiểm tra lại bằng tính toán

bởi chương trình toán học trên máy tính Mathematica. Kết quả nhận được cho thấy rằng các kết quả phân tích là chính xác và phù hợp.

- Để so sánh mức tối ưu hóa thông lượng và độ trễ nhận được trên mô hình mạng đề xuất, sẽ trình bày thêm phương pháp lưu trữ cơ bản trong đó số lượng bản sao của tệp dữ liệu tại các thiết bị di động và trạm gốc thông tin di động được tối ưu một cách độc lập với nhau.

2.1. Mô hình mạng vô tuyến hướng nội dung sử dụng mô hình dòng chảy

Trong chương này, mô hình mạng nghiên cứu là mạng vô tuyến hướng nội dung trong đó bao gồm n thiết bị di động và xem xét trường hợp mô hình mạng chứa $f(n)$ trạm gốc thông tin di động (hoặc các thiết bị hỗ trợ truyền dẫn được đặt tại các vị trí cố định trong mạng) (xem hình 2.1). Do trong thực tế, mỗi trạm gốc thông tin được xây dựng để phục vụ nhiều thiết bị người dùng. Do đó, số lượng các trạm gốc thông tin giả sử rằng tỷ lệ với số lượng thiết bị người dùng $f(n) = \Theta(n^\delta)$ và hệ số δ sẽ có giá trị $0 \leq \delta < 1$. Đây cũng là mô hình mạng chung phổ biến được sử dụng tại các nghiên cứu trước [1, 15, 33, 34, 36, 43, 46]. Giả sử rằng n thiết bị di động người dùng được phân bố đều và ngẫu nhiên trên mô hình mạng có kích thước đơn vị chuẩn và $f(n)$ trạm gốc thông tin được đặt vị trí một cách đồng đều trên mạng. Theo đó, mạng có thể được chia thành $f(n)$ cụm tế bào truyền tin femto cells có kích thước tương đương $b(n) = \Theta\left(\frac{1}{f(n)}\right)$ với mỗi cụm có một trạm gốc thông tin được đặt ở vị trí giữa. Tính chất di động của các thiết bị được mô hình hóa theo mô hình bước đi ngẫu nhiên RWMM đã được mô tả tại [1, 23]. Trong mô hình mạng nghiên cứu, mỗi thiết bị di động và trạm gốc thông tin



Hình 2.1: Mạng vô tuyến hướng nội dung sử dụng mô hình dòng chảy.

di động được giả sử rằng đều được trang bị các bộ nhớ lưu trữ có thể được sử dụng để lưu trữ một phần dữ liệu trong thư viện mạng, trong đó kích thước dữ liệu của thư viện mạng có thể chứa M tệp dữ liệu. Luận án này nghiên cứu mô hình mạng cỡ lớn với số lượng thiết bị người dùng n tiến tới vô cùng. Trong khi đó, mỗi thư viện mạng đều có kích thước có giới hạn. Do đó, giả sử rằng $M = \Theta(n^\gamma)$ và hệ số $0 < \gamma < 1$. Mỗi tệp dữ liệu được chứa trong thư viện mạng được giả sử rằng có cùng kích cỡ. Thêm vào đó, mô hình mạng đề xuất trong chương này sẽ xem xét sử dụng bộ nhớ lưu trữ có giới hạn của các thiết bị người dùng K_n và trạm gốc thông tin K_{BS} để áp dụng kỹ thuật đệm dữ liệu. Giả sử rằng mỗi thiết bị người dùng chỉ lưu trữ được một vài tệp dữ liệu $K_n = \Theta(1)$. Dung lượng lưu trữ chia sẻ của các trạm gốc thông tin là có giới hạn và không thể chứa hết số lượng tệp dữ liệu của thư viện mạng, do đó ta giả sử rằng $K_{BS} = \Theta(n^\beta)$ và hệ số $0 < \beta < \gamma$. Để ảnh hưởng và

lợi ích của việc sử dụng bộ nhớ lưu trữ của các trạm gốc chia sẻ dữ liệu với các thiết bị người dùng được thể hiện rõ sự khác biệt so với các nghiên cứu trước, nghiên cứu tại chương này chỉ tập trung trường hợp tổng dung lượng bộ nhớ lưu trữ chia sẻ của các trạm gốc thông tin lớn hơn tổng dung lượng bộ nhớ lưu trữ chia sẻ của các thiết bị người dùng trong mạng, tương đương với $f(n)K_{BS} \geq nK_n$ và $\beta + \gamma \geq 1$.

Giả sử rằng các các thiết bị di động sẽ thực hiện các yêu cầu tải dữ liệu một cách độc lập theo phân bố Zipf. Phân bố Zipf phù hợp sử dụng trong các nghiên cứu trong mô hình mạng vô tuyến di động [22, 30, 37, 44, 51]. Theo đó, xác suất yêu cầu tải tin của tệp dữ liệu được thể hiện bởi $m \in \mathcal{M} \triangleq \{1, \dots, M\}$ như sau¹:

$$p_m = \frac{m^{-\alpha}}{H_\alpha(M)}, \quad (2.1)$$

trong đó $\alpha > 0$ là hệ số Zipf và $H_\alpha(M) = \sum_{i=1}^M i^{-\alpha}$ là hệ số chuẩn hóa được cho bởi hàm Riemann zeta và được tính như sau

$$H_\alpha(M) = \begin{cases} \Theta(1) & \text{với } \alpha > 1 \\ \Theta(\log M) & \text{với } \alpha = 1 \\ \Theta(M^{1-\alpha}) & \text{với } \alpha < 1. \end{cases} \quad (2.2)$$

Trong mạng vô tuyến hướng nội dung, vấn đề mấu chốt khi sử dụng kỹ thuật đệm dữ liệu là cần xem xét hai bước chính là bước phân bố các tệp dữ liệu vào bộ nhớ chia sẻ của các thực thể mạng và bước truyền dữ liệu trong mạng. Cụ thể, vấn đề bao gồm bài toán lưu trữ dữ liệu tối ưu tại bộ nhớ lưu trữ của các trạm gốc thông tin, thiết bị di động và bài toán tìm đường truyền tin hiệu quả nhất để đáp ứng các yêu cầu về thông tin và dữ liệu trong mạng.

¹Số mã hóa của các tệp dữ liệu m được đặt theo thứ tự giảm dần của xác suất yêu cầu tải của M tệp dữ liệu trong mạng.

Trong bước lưu trữ đệm dữ liệu, các tập dữ liệu từ thư viện của mạng được lựa chọn để lưu trữ tại các bộ nhớ của n thiết bị người dùng và $f(n)$ trạm gốc thông tin. Mỗi tập dữ liệu $m \in \mathcal{M}$ trong thư viện mạng có thể có một hoặc nhiều bản sao được phân bố ngẫu nhiên trong mạng nghiên cứu. Đặt A_m và B_m lần lượt là ký hiệu của số lượng bản sao của tập dữ liệu m được lưu trữ tại các thiết bị người dùng và trạm gốc thông tin tương ứng. Hai tham số này sẽ được tối ưu để tìm ra thông lượng và độ trễ tối ưu nhất của mạng. Để bước lưu trữ đệm dữ liệu được thực hiện, $\{A_m\}_{m=1}^M$ và $\{B_m\}_{m=1}^M$ phải thỏa mãn các điều kiện sau:

$$\begin{cases} \sum_{m=1}^M A_m \leq nK_n, \\ \sum_{m=1}^M B_m \leq f(n)K_{BS}. \end{cases} \quad (2.3)$$

Giới hạn ràng buộc:

$$\begin{cases} A_m \leq n, \\ B_m \leq f(n), \\ A_m + B_m \geq 1 \end{cases} \quad (2.4)$$

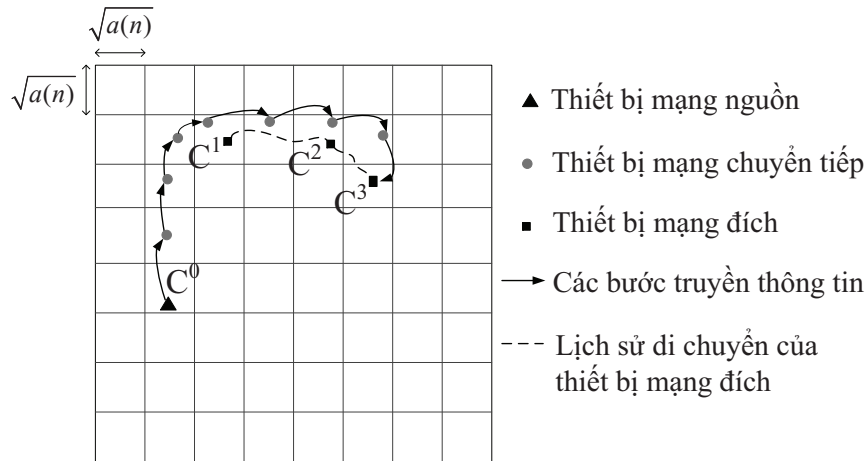
với mọi $m \in \mathcal{M}$. Chú ý rằng giới hạn cuối cùng ở (2.4) được sử dụng để tránh trường hợp một tập dữ liệu bất kỳ nào đó trong thư viện mạng không được lưu trữ tại bất kỳ thiết bị di động hay trạm gốc thông tin nào. Tương tự như các nghiên cứu [15, 43], nghiên cứu này áp dụng cơ chế lưu trữ đệm dữ liệu ngẫu nhiên, trong đó, các tập bản sao tập dữ liệu thỏa mãn (2.3) và (2.4) được phân bố đều và ngẫu nhiên tại các bộ nhớ lưu trữ của n thiết bị người dùng và $f(n)$ trạm gốc thông tin.

Nghiên cứu này sẽ được thực hiện bằng cách tham khảo và sử dụng mô hình *dòng chảy* được nghiên cứu ở [1]. Trong mô hình này, kích cỡ của mỗi

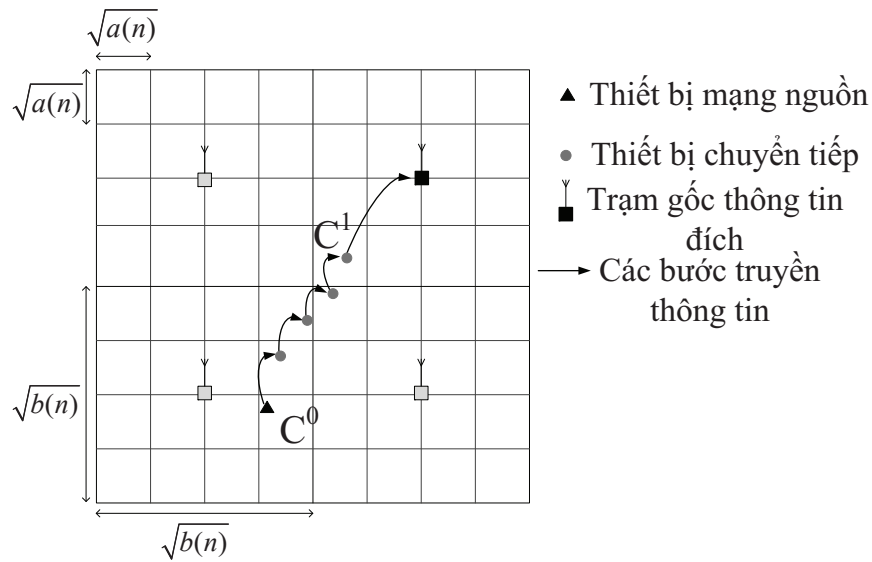
tệp dữ liệu được giả sử là vô cùng nhỏ sao cho thời gian cần thiết để truyền một tệp dữ liệu giữa một thiết bị người dùng và thiết bị người dùng lân cận hoặc trạm gốc thông tin trong cùng cụm tế bào truyền tin nhỏ hơn nhiều khoảng thời gian của một khe thời gian. Nhờ đó, dữ liệu được gửi đi từ một thiết bị trong một khe thời gian có thể tương đương với nhiều tệp dữ liệu và do đó tất cả các tệp dữ liệu của thiết bị sẽ đảm bảo được truyền đi hết bởi thiết bị đó trong khoảng thời gian đơn vị là một khe thời gian. Tuy nhiên, mỗi tệp dữ liệu nhận được bởi thiết bị trong khe thời gian này không được phép truyền đi tiếp cho đến khi bắt đầu khe thời gian tiếp theo. Trong nghiên cứu tại chương tiếp theo, mô hình truyền tin dòng chảy sẽ được mở rộng áp dụng kỹ thuật phân mảnh tệp dữ liệu sao cho mỗi tệp dữ liệu được chia ra thành nhiều mảnh tin nhỏ có kích thước nhỏ vừa đủ để có thể truyền đi hết trong mỗi khe thời gian.

2.2. Đề xuất phương pháp định tuyến truyền tin

Trong phần này đưa ra phương pháp định tuyến để truyền tải các tệp dữ liệu tới các thiết bị đầu cuối đích đang yêu cầu tải tệp dữ liệu đó. Do tính chất di động của các thuê bao, phương pháp định tuyến sẽ được xây dựng dựa trên cơ chế định tuyến đa chặng lân cận gần nhất [1] và điều chỉnh lại cho phù hợp với mô hình mạng vô tuyến hướng nội dung trong nghiên cứu này. Với kỹ thuật truyền tin đa chặng, mô hình mạng nghiên cứu dạng Torus có kích thước đơn vị chuẩn (a unit-sized torus) được chia thành $a(n)^{-1}$ ô tế bào vuông nhỏ có kích thước tương đương nhau, trong đó $a(n) = \Omega\left(\frac{\log n}{n}\right)$ và $a(n) = O(1)$. [38] chỉ ra rằng, mỗi ô tế bào này có xác suất rất cao luôn chứa ít nhất một thiết bị người dùng di động. Ở đây sử dụng cơ chế định



(a) Người dùng tới người dùng



(b) Người dùng tới trạm gốc thông tin

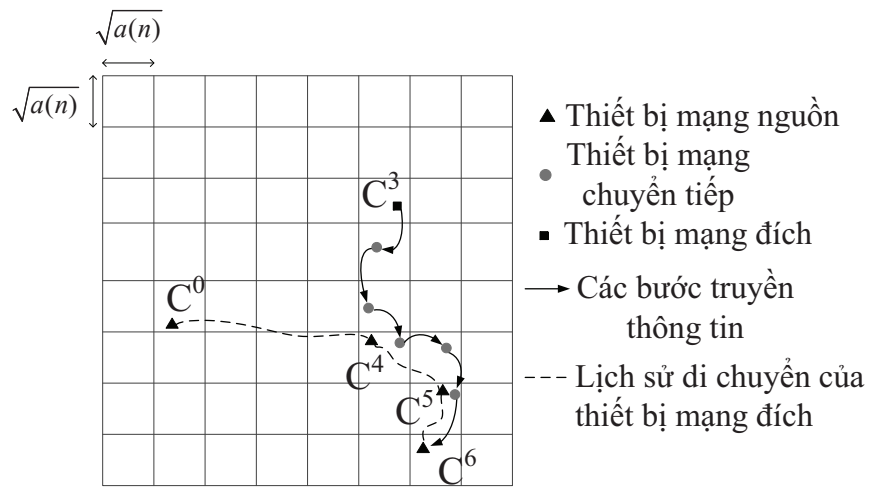
Hình 2.2: Giai đoạn thứ nhất của quá trình định tuyến truyền dữ liệu.

tuyến đa chặng dành cho bước truyền tin dựa trên các ô tế bào và vùng phủ sóng của trạm gốc thông tin di động có kích cỡ $a(n)$ và $b(n)$ tương ứng. Mỗi ô tế bào được kích hoạt hoạt động lại đều sau mỗi $1 + c$ khe thời gian để đảm bảo việc xung đột trong truyền tin, trong đó $c > 0$ là số tự nhiên độc lập với n . Tương tự, mỗi vùng phủ sóng của trạm gốc thông tin di động được kích hoạt hoạt động đều sau mỗi $1 + C$ khe thời gian. Việc thực hiện thủ tục định tuyến này không xét tới yếu tố hàng đợi của các yêu cầu ở các ô tế bào.

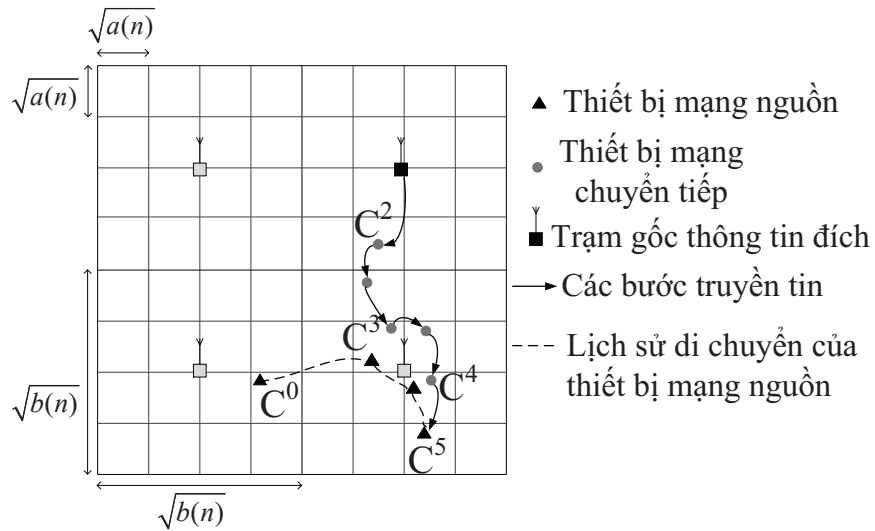
Thiết bị di động người dùng có yêu cầu tải tệp dữ liệu đầu tiên sẽ tìm thực thể có khoảng cách gần nhất trong A_m thiết bị và B_m trạm gốc đang lưu trữ tệp dữ liệu mong muốn (theo khoảng cách Euclidean). Sau đó, tin nhắn yêu cầu được tải dữ liệu sẽ được gửi tới thực thể mạng đó thông qua các ô tế bào lân cận theo phương pháp truyền tin đa chặng, tương đương với bước đầu tiên của quá trình truyền tin. Tương tự, tệp dữ liệu mong muốn sẽ được gửi tới thiết bị đích có yêu cầu tải dữ liệu theo phương pháp truyền tin đa chặng theo chiều ngược lại. Lưu ý rằng các thiết bị đích yêu cầu tải dữ liệu cũng đang di chuyển trong mạng theo mô hình bước đi ngẫu nhiên RWMM. Mỗi khe thời gian sẽ được chia thành hai khe thời gian nhỏ hơn, trong đó bước thứ nhất và thứ hai của thủ tục truyền tin sẽ được thực hiện tương ứng với khoảng thời gian của hai khe thời gian nhỏ này. Trong trường hợp thiết bị đầu cuối yêu cầu tải dữ liệu có vị trí đứng ở trong phạm vi định tuyến truyền tin cơ bản của thực thể mạng lưu trữ tệp dữ liệu, yêu cầu tải dữ liệu sẽ được đáp ứng ngay nhờ phương pháp truyền tin đơn chặng chỉ trong khoảng thời gian của một khe thời gian tiêu chuẩn. Chi tiết thủ tục truyền tin được mô tả như sau:

Step 1) Bước yêu cầu tải dữ liệu

- (a) Nếu thực thể gần nhất lưu trữ tệp dữ liệu mong muốn là thiết bị người dùng đầu cuối di động, bản tin yêu cầu tải dữ liệu sẽ di chuyển tìm tới thiết bị này theo thủ tục như sau. Như thể hiện ở Hình vẽ 2.2(a), xuất phát từ ô tế bào C^0 , bản tin yêu cầu tải tin sẽ được truyền theo phương thức đa chặng tới các ô tế bào lân cận tới ô tế bào C^1 đang chứa thiết bị đích, với khoảng cách mỗi bước truyền tin được tính bởi $\Theta\left(\sqrt{a(n)}\right)$.



(a) Người dùng tới người dùng



(b) Trạm gốc thông tin tới người dùng

Hình 2.3: Giai đoạn thứ hai của quá trình định tuyến truyền dữ liệu.

Khi tin nhắn yêu cầu tải dữ liệu được truyền tới ô tế bào C^1 , thiết bị đích chứa tệp dữ liệu mong muốn đã di chuyển tới vị trí khác C^2 do đặc tính di động của thiết bị theo mô hình bước đi ngẫu nhiên RWMM. Do đó, tin nhắn yêu cầu tải dữ liệu phải tiếp tục di chuyển từ ô tế bào C^1 tới ô tế bào C^2 . Quy trình này sẽ tiếp tục cho tới khi bản tin yêu cầu tải dữ liệu và thiết bị đích cùng gặp nhau tại ô tế bào C^3 .

(b) Như mô tả ở Hình vẽ 2.2(b), nếu thực thể gần nhất lưu trữ tệp dữ liệu

yêu cầu là trạm gốc thông tin, bản tin yêu cầu tải dữ liệu sẽ được truyền theo phương pháp truyền tin đa chặng tới các ô tế bào lân cận thẳng về hướng của trạm gốc thông tin đích, với khoảng cách bước truyền tin là $\Theta\left(\sqrt{a(n)}\right)$. Khi tin nhắn yêu cầu tải tin rời khỏi thiết bị người dùng di động ở ô tế bào C^1 thuộc phạm vi vùng phủ hoạt động của trạm gốc thông tin đích, thiết bị chuyển tiếp sẽ gửi tin nhắn yêu cầu tải tin tới trạm gốc thông tin ngay lập tức theo phương pháp đơn chặng trong thời gian của một khe thời gian chuẩn, trong đó chiều dài bước truyền tin sau cùng được cho bởi $\Theta\left(\sqrt{b(n)}\right)$. Chiều dài bước truyền tin này sẽ giúp cho mạng đạt hiệu năng tốt nhất có thể kỳ vọng (vấn đề này sẽ được phân tích kỹ sau).

Step 2) Bước truyền tin

- (a) Khi thực thể mạng đích chứa tệp dữ liệu nhận được yêu cầu tải tin, thiết bị di động người dùng yêu cầu tải dữ liệu đã di chuyển tới vị trí khác so với ban đầu C^4 do tính chất di động theo mô hình bước đi ngẫu nhiên RWMM. Như thể hiện tại Hình vẽ 2.3(a), tệp dữ liệu được yêu cầu tải sẽ được truyền tới thiết bị đích bằng cách di chuyển theo thiết bị đích này theo cách thức tương tự như tin nhắn yêu cầu tải dữ liệu phải đi trong bước truyền tin thứ nhất.
- (b) Khi thực thể mạng đích là trạm gốc thông tin nhận được tin nhắn yêu cầu tải tin, thiết bị di động yêu cầu tải dữ liệu đã di chuyển tới vị trí khác C^3 . Như thể hiện ở Hình vẽ 2.3(b), tệp dữ liệu được yêu cầu tải sẽ được truyền đi từ trạm gốc thông tin này qua một thiết bị chuyển tiếp ở ô tế bào C^2 có vị trí thẳng tới hướng của thiết bị di động yêu cầu tải

Algorithm 1 Thủ tục truyền tin

- 1: Bước 1. Bước thứ nhất (bước yêu cầu truyền tin)
 - 2: Bước 1-1. Truyền tin theo dạng Người dùng tới người dùng
 - 3: **if** Thực thể gần nhất chứa tệp dữ liệu là trạm gốc thông tin **then**
 - 4: Bước 1-2. Truyền tin theo dạng Người dùng tới trạm gốc
 - 5: **end if**
 - 6: Bước 2: Bước thứ hai (bước truyền tin)
 - 7: **if** Thực thể gần nhất chứa tệp dữ liệu là trạm gốc thông tin **then**
 - 8: Bước 2-1. Truyền tin theo dạng Trạm gốc tới người dùng
 - 9: **end if**
 - 10: Bước 2-2. Truyền tin theo dạng Người dùng tới người dùng
-

dữ liệu chỉ trong khoảng thời gian 01 khe thời gian chuẩn. Sau đó, thiết bị chuyển tiếp sẽ di chuyển theo thiết bị người dùng đích theo phương pháp truyền tin đa chặng cho tới khi bắt kịp tới ô tế bào có chứa thiết bị người dùng đích ở đó.

Thủ tục chung của phương pháp định tuyến truyền tin đề xuất được mô tả ngắn gọn tại Thuật toán 1.

2.3. Thông lượng và độ trễ của mạng

Ở phần này sẽ đưa ra công thức thể hiện mối quan hệ và tính cân bằng giữa hai thông số quan trọng thể hiện hiệu năng mạng là thông lượng và độ trễ trong mạng vô tuyến hỗn hợp hướng nội dung sử dụng phương pháp định tuyến truyền tin được đề xuất ở phần trước.

Các tham số thông lượng và độ trễ của mô hình mạng đề xuất trong Chương 2 được định nghĩa lại như sau:

Thông lượng: Ở đây, tham số thông lượng của mạng là giá trị trung bình của dung lượng các tệp dữ liệu mà thiết bị người dùng nhận được trong một khe thời gian.

Độ trễ: Tham số độ trễ của mạng là thời gian trung bình tính từ thời điểm tin nhắn yêu cầu tải tin gửi đến thiết bị đích chứa tệp dữ liệu cho đến

khi thiết bị người dùng yêu cầu tải tin nhận được tệp dữ liệu mong muốn.

Như đã đề cập trong phần nội dung của phần trước về phương pháp truyền dẫn dữ liệu, phương pháp truyền dẫn đa chặng được sử dụng để tìm đường gần nhất giữa thiết bị người dùng yêu cầu tải dữ liệu và thực thể mạng gần nhất lưu trữ tệp dữ liệu mong muốn trong bộ nhớ của mình, trong đó, khoảng cách giữa hai thiết bị này là nhân tố quan trọng và được xác định bởi tham số là số lượng bản ghi của tệp dữ liệu $m \in \mathcal{M}$, $A_m + B_m$. Theo giả thuyết, số lượng bản ghi của mỗi tệp dữ liệu trong mạng được phân bố đều và độc lập tại bước đệm dữ liệu tại các thực thể mạng, khoảng cách Euclidean trung bình từ thiết bị người dùng yêu cầu tải dữ liệu tới thực thể đích chứa tải tin gần nhất đã được chứng minh là biến thiên theo căn bậc hai của số lượng bản ghi của tệp dữ liệu trong mạng [1, 36]. Áp dụng luận giải tương tự cho mô hình mạng đã đưa ra, mệnh đề sau đây được thiết lập thể hiện mối quan hệ và tính cân bằng giữa thông lượng và độ trễ của mạng.

Bổ đề 2.3.1. [36] *Khi một thiết bị mạng yêu cầu tải tệp dữ liệu $m \in \mathcal{M}$, khoảng cách trung bình ban đầu giữa thiết bị mạng yêu cầu tải dữ liệu tới thực thể mạng gần nhất lưu trữ tệp dữ liệu m tại bộ nhớ đệm được tính theo công thức $\Theta\left(\frac{1}{\sqrt{A_m+B_m}}\right)$, với A_m và B_m là số lượng bản ghi của tệp dữ liệu m được lưu trữ tại bộ nhớ đệm của các thiết bị và trạm gốc thông tin tương ứng.*

Chứng minh. Chi tiết để chứng minh mệnh đề trên đã được luận giải tại tài liệu [36, Lemma 3]. Với $X_m = A_m + B_m$, phần chứng minh này có thể được trích lại như sau.

Do các bản ghi của tệp dữ liệu m được phân bố độc lập và ngẫu nhiên trong toàn mạng nên xác suất để không có thiết bị hay trạm gốc nào chứa

bản ghi của tập dữ liệu m nằm trong khoảng cách τ của thiết bị có nhu cầu tải tập tin là $Pr(d \geq \tau) = (1 - \pi\tau^2)^{X_m}$ với $0 \leq \tau \leq 1/\sqrt{\pi}$. Do đó, khoảng cách trung bình từ thiết bị có nhu cầu tải tin tới thiết bị hoặc trạm gốc gần nhất chứa tập tin mong muốn là

$$E[d] = \int_0^\infty Pr(d \geq \tau) d\tau = \int_0^{1/\sqrt{\pi}} (1 - \pi\tau^2)^{X_m} d\tau.$$

Ta thực hiện đổi biến với $\sqrt{\pi}\tau = \cos\theta$ và áp dụng tích phân từng phần, ta có

$$E[d] = \frac{1}{\sqrt{\pi}} \int_0^{\pi/2} (\sin\theta)^{2X_m+1} d\theta \quad (2.5)$$

$$= \frac{1}{\sqrt{\pi}} \frac{2X_m}{2X_m+1} \cdot \frac{2X_m-2}{2X_m-1} \cdots \frac{2}{3} \int_0^{\pi/2} \sin\theta d\theta \quad (2.6)$$

$$= \frac{1}{\sqrt{\pi}} \frac{2X_m}{2X_m+1} \cdot \frac{2X_m-2}{2X_m-1} \cdots \frac{2}{3} \quad (2.7)$$

$$= \Theta\left(\frac{1}{\sqrt{X_m}}\right). \quad (2.8)$$

Với (2.6) được phân tích từ công thức

$$\int \sin^n x dx = -\frac{1}{n} \sin^{n-1} x \cos x + \frac{n-1}{n} \int \sin^{n-2} x dx. \quad (2.9)$$

và (2.8) được chứng minh như sau:

Trước hết ta định nghĩa

$$g(n) = \frac{n-1}{n} \cdot \frac{n-3}{n-2} \cdots \frac{2}{3}. \quad (2.10)$$

Ta cần chứng minh $g(2X_m+1) = \Theta(1/\sqrt{X_m})$, hoặc tương đương, $g(n)$ biến thiên theo $1/\sqrt{n}$ với n là số lẻ. Đặt n_1 và n_2 là các số lẻ và $n_1 > n_2$. Ta có thể viết như sau

$$g(n_1) = \frac{n_1-1}{n_1} \cdot \frac{n_1-3}{n_1-2} \cdots \frac{n_2+1}{n_2+3} g(n_2). \quad (2.11)$$

Thay đổi lại thứ tự và chia cả hai vế cho $g(n_2)$, ta có

$$\frac{g(n_1)}{g(n_2)} = \frac{n_2 + 1}{n_1} \left(\frac{n_1 - 1}{n_1 - 2} \frac{n_1 - 3}{n_1 - 4} \dots \frac{n_2 + 3}{n_2 + 2} \right) \quad (2.12)$$

$$= \frac{n_2 + 1}{n_1} \cdot \frac{g(n_2 + 1)}{g(n_1 - 1)}. \quad (2.13)$$

Từ đây ta có

$$\frac{g(n_1)}{g(n_2)} \cdot \frac{g(n_1 - 1)}{g(n_2 + 1)} = \frac{n_2 + 1}{n_1}. \quad (2.14)$$

Ta thấy rằng, $g(n)$ là hàm không biến thiên tăng lên theo sự biến thiên của n . Do đó,

$$\frac{g(n_1 - 1)}{g(n_2 + 1)} \geq \frac{g(n_1)}{g(n_2)}. \quad (2.15)$$

Từ đây,

$$\left(\frac{g(n_1)}{g(n_2)} \right)^2 \leq \frac{g(n_1)}{g(n_2)} \cdot \frac{g(n_1 - 1)}{g(n_2 + 1)} = \frac{n_2 + 1}{n_1}. \quad (2.16)$$

và tương tự

$$()^2 \geq \frac{g(n_1)}{g(n_2)} \cdot \frac{g(n_1 - 1)}{g(n_2 + 1)} = \frac{n_2 + 1}{n_1}. \quad (2.17)$$

Kết hợp (2.16) và (2.17), ta có điều kiện biên (2.8) được chứng minh với

$$\frac{n_2}{n_1 + 1} \leq \left(\frac{g(n_1)}{g(n_2)} \right)^2 \leq \frac{n_2 + 1}{n_1}.$$

Mệnh đề được chứng minh khi thay $n_2 = 3$. □

Sử dụng Bổ đề 2.3.1, luận án này thiết lập Định lý sau đây.

Định lý 2.3.1. *Giả sử mạng vô tuyến hỗn hợp hướng nội dung sử dụng phương pháp truyền tin được trình bày tại Chương 1, mức cân bằng thông lượng và độ trễ của mạng được tính bởi công thức như sau*

$$\lambda(n) = \Theta \left(\frac{D(n)}{n \left(\sum_{m=1}^M \frac{p_m}{\sqrt{A_m + B_m}} \right)^2} \right) \quad (2.18)$$

trong đó, biên trên của $\lambda(n)$ là

$$\lambda(n) = O\left(\frac{1}{\sum_{m=1}^M p_m \sqrt{\frac{n \log n}{A_m + B_m}}}\right)$$

và p_m là xác suất yêu cầu tải tin của tập dữ liệu $m \in \mathcal{M}$.

Chứng minh. Mức cân bằng thông lượng và độ trễ được tính bởi công thức tính thông lượng theo một phương trình của độ trễ, trong đó sự thay đổi giá trị của độ trễ cũng làm thay đổi giá trị của thông lượng. Trước hết, khoảng cách ban đầu của một cặp nguồn đích truyền tin ngẫu nhiên trong mô hình mạng hướng nội dung đề xuất cần được tính toán. Chiều dài định tuyến của toàn bộ quãng đường cần đi của bản tin yêu cầu tải dữ liệu được quyết định bởi khoảng cách ban đầu của cặp truyền tin nguồn đích, trong đó khoảng cách này được cho bởi Bổ đề 2.3.1 như sau $\Theta\left(\frac{1}{\sqrt{A_m + B_m}}\right)$. Từ đó, ta có tổng số bước của quãng đường đi định tuyến của bản tin yêu cầu tải dữ liệu và tập dữ liệu mong muốn biến thiên theo $\Theta\left(\frac{1}{\sqrt{a(n)(A_m + B_m)}}\right)$, trong đó, $a(n) = \Omega\left(\frac{\log n}{n}\right)$ và giới hạn biên trên $a(n) = O(1)$ là kích thước ô tế bào. Độ trễ mạng $D(n)$ được xác định từ thời điểm bản tin yêu cầu tải dữ liệu rời đi từ thiết bị yêu cầu tải dữ liệu cho tới khi thiết bị này nhận được tập dữ liệu mong muốn. Thêm vào đó, độ trễ luôn tỷ lệ thuận với chiều dài quãng đường định tuyến và mỗi bước truyền thông tin được tính là mất quãng thời gian tương đương với một khe thời gian đơn vị. Từ đó, độ trễ mạng có thể được tính theo công thức sau:

$$D(n) = \Theta\left(\sum_{m=1}^M \frac{p_m}{\sqrt{a(n)(A_m + B_m)}}\right), \quad (2.19)$$

trong đó p_m là xác suất thiết bị người dùng bất kỳ yêu cầu tải tập dữ liệu $m \in \mathcal{M}$.

Tương tự như [1], số lượng tệp dữ liệu trung bình đi qua một ô tế bào bất kỳ ở mỗi khe thời gian được tính bởi công thức $O\left(n \sum_{m=1}^M p_m \sqrt{\frac{a(n)}{A_m+B_m}}\right)$. Từ đó, thông lượng tại mỗi thiết bị người dùng trung bình được tính bởi

$$\lambda(n) = \Theta\left(\frac{1}{n \sum_{m=1}^M p_m \sqrt{\frac{a(n)}{A_m+B_m}}}\right), \quad (2.20)$$

thông lượng đạt được tối ưu khi $a(n) = \Theta\left(\frac{\log n}{n}\right)$. Do đó, sử dụng (2.19) và (2.20) sẽ dẫn tới kết quả (2.18). Định lý đã được chứng minh. \square

Định lý 2.3.1 chỉ ra rằng mức cân bằng thông lượng và độ trễ của mạng bị ảnh hưởng bởi tổng số bản ghi của tệp dữ liệu m trong mạng, $A_m + B_m$. Từ các giới hạn tại (2.3) và (2.4), tối ưu quá trình lưu trữ các tệp dữ liệu tại bước đệm dữ liệu $\{A_m\}_{m=1}^M$ và $\{B_m\}_{m=1}^M$ sẽ dẫn đến việc nâng cao hiệu năng mạng thông qua mức cân bằng thông lượng và độ trễ. Trong phần tiếp theo của luận án sẽ giới thiệu phương pháp lưu trữ dữ liệu tối ưu tại bộ nhớ của các thực thể mạng sao cho đạt được mức thông lượng và độ trễ tối ưu nhất.

Chú ý 2.3.1. Từ kết quả ở trên, nếu các tệp dữ liệu được truyền đi tại bước truyền tin được thực hiện theo phương pháp đa chặng, mức cân bằng giữa thông lượng và độ trễ của mạng đối với mạng vô tuyến di động và mạng vô tuyến cố định [36] hướng nội dung là tương đương với nhau. Có nghĩa là, hiệu năng của mạng mức cân bằng thông lượng và độ trễ trong mạng vô tuyến hướng nội dung là không thay đổi dù thiết bị người dùng có tính di động hay không khi phương pháp định tuyến truyền tin đa chặng được sử dụng trong bước truyền tin.

Phân tích tương tự như [36] khi $\alpha \geq 3/2$, mức cân bằng thông lượng và độ trễ mạng tối ưu tại Định lý 2.3.1 được tính bởi $\lambda(n) = \Theta(D(n))$. Do đó,

trong phần còn lại của nghiên cứu này chỉ tập trung phân tích trường hợp $\alpha < 3/2$ trong bài toán tối ưu hóa vấn đề lưu trữ dữ liệu trong bước đệm dữ liệu.

2.4. Tối ưu hóa thông lượng và độ trễ

Trong phần này, mức tối ưu thông lượng và độ trễ của mạng vô tuyến hỗn hợp di động hướng nội dung sử dụng kỹ thuật đệm dữ liệu sẽ được tính toán dựa trên việc lựa chọn tối ưu các tham số về số lượng bản ghi $\{A_m\}_{m=1}^M$ và $\{B_m\}_{m=1}^M$. Bài toán đề xuất tối ưu hóa trước hết sẽ được giới thiệu để tối ưu mức cân bằng thông lượng và độ trễ mạng. Sau đó, phương pháp lưu trữ dữ liệu tại bước đệm dữ liệu sẽ được đề xuất nhờ việc tìm ra số lượng bản ghi tối ưu của mỗi tệp dữ liệu tại các thiết bị di động người dùng và các trạm gốc thông tin, từ đó có thể tìm ra mức cân bằng tối ưu của các tham số hiệu năng mạng được đề cập. Cuối cùng, phần phân tích sẽ được kiểm tra và đánh giá tính đúng đắn nhờ việc thực hiện sử dụng các phần mềm tính toán trên máy tính.

2.4.1. Xây dựng bài toán tối ưu hóa thông lượng và độ trễ mạng

Từ Định lý 2.3.1 và từ các mức giới hạn (2.3) và (2.4), bài toán tối ưu hóa được xây dựng nên như sau:

$$\max_{\{A_m\}_{m=1}^M, \{B_m\}_{m=1}^M} \lambda(n) \quad (2.21a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M A_m \leq nK_n, \quad (2.21b)$$

$$\sum_{m=1}^M B_m \leq f(n)K_{BS}, \quad (2.21c)$$

$$A_m \leq n \text{ trong đó } m \in \mathcal{M}, \quad (2.21d)$$

$$B_m \leq f(n) \text{ trong đó } m \in \mathcal{M}, \quad (2.21e)$$

$$A_m + B_m \geq 1 \text{ trong đó } m \in \mathcal{M}. \quad (2.21f)$$

Ta nhận thấy tối ưu hóa thông lượng $\lambda(n)$ và độ trễ mạng $D(n)$ tương đương với tối thiểu hóa $n \left(\sum_{m=1}^M \frac{p_m}{\sqrt{A_m + B_m}} \right)$ ở công thức (2.18), bài toán ở (2.21) có thể được viết lại như sau

$$\min_{\{A_m\}_{m=1}^M, \{B_m\}_{m=1}^M} \sum_{m=1}^M \frac{p_m}{\sqrt{A_m + B_m}} \quad (2.22a)$$

$$\text{với các điều kiện: } (2.21b)-(2.21f). \quad (2.22b)$$

Tham khảo [42], tương tự như những công trình nghiên cứu trước đây [1, 15, 36, 43, 46], đặt $f_m = \sum_{m=1}^M \frac{p_m}{\sqrt{A_m + B_m}}$, ta có khả vi lần thứ nhất và lần thứ hai của f_m theo các tập biến $\{A_m\}_{m=1}^M$ và $\{B_m\}_{m=1}^M$ như sau:

$$f'_m = - \sum_{m=1}^M p_m (A_m + B_m)^{-\frac{3}{2}}$$

và

$$f''_m = \sum_{m=1}^M 3p_m (A_m + B_m)^{-\frac{5}{2}}.$$

Ta thấy rằng, khả vi hai lần của hàm tối ưu f_m luôn dương. Do đó, ta có hàm tối ưu là hàm lồi. Từ đó, phương pháp Lagrange có thể được sử dụng để giải bài toán trên (2.22). Do nghiên cứu thực hiện dựa trên các phép toán về luật số lớn và xấp xỉ khi giá trị $n \rightarrow \infty$ nên ta có thể giả sử các tập biến $\{A_m\}_{m=1}^M$ và $\{B_m\}_{m=1}^M$ là các số thực dương. Từ đó ta thấy rằng, bài toán (2.22) có nghiệm tối ưu toàn cục và sẽ được giải ở phần nội dung tiếp theo.

2.4.2. Giải bài toán tối ưu hóa thông lượng và độ trễ mạng

Ở đây, kỹ thuật giải tách biến có thể áp dụng đối với các tập biến $\{A_m\}_{m=1}^M$ và $\{B_m\}_{m=1}^M$, nhờ đó có thể đơn giản hóa lời giải mà vẫn đảm bảo được kết quả sau cùng. Tổng các bản ghi $m \in \mathcal{M}$, $A_m + B_m$ có thể được diễn tả theo $\Theta(A_m)$ hoặc $\Theta(B_m)$ tùy thuộc vào độ lớn tương quan giữa A_m và B_m . Theo đó, giả sử tập nghiệm tối ưu là $\{A_m^*\}_{m=1}^M$ và $\{B_m^*\}_{m=1}^M$, định nghĩa \mathcal{M}_1 và \mathcal{M}_2 là tập các tập dữ liệu sao cho $A_m^* + B_m^* = \Theta(A_m^*)$ và $A_m^* + B_m^* = \Omega(f(n))$ tương ứng (chi tiết ở Định lý 2). Ngoài ra \mathcal{M}_3 là tập các tập dữ liệu sao cho $A_m^* + B_m^* = \Theta(B_m^*)$. Trong đó, $m \in \mathcal{M}_3$, $A_m = O(B_m) = O(f(n))$. Từ đây, nghiệm của bài toán tối ưu ở (2.22) sẽ được tìm ra bằng cách giải hai bài toán tối ưu sau:

$$\min_{\{A_m\}_{m \in \mathcal{M}_1}} \sum_{m \in \mathcal{M}_1} \frac{p_m}{\sqrt{A_m}} \quad (2.23a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M A_m \leq nK_n, \quad (2.23b)$$

$$A_m \leq n \text{ trong đó } m \in \mathcal{M}_1 \quad (2.23c)$$

và

$$\min_{\{B_m\}_{m \in \mathcal{M}_3}} \sum_{m \in \mathcal{M}_3} \frac{p_m}{\sqrt{B_m}} \quad (2.24a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M B_m \leq f(n)K_{BS}, \quad (2.24b)$$

$$B_m \leq f(n) \text{ trong đó } m \in \mathcal{M}_3. \quad (2.24c)$$

Hàm Lagrange tương ứng với (2.23) được tính như sau

$$\begin{aligned} & \mathcal{L}_1(\{A_m\}_{m \in \mathcal{M}_1}, \lambda, \{w_m\}_{m \in \mathcal{M}_1}) \\ &= \sum_{m \in \mathcal{M}_1} \frac{p_m}{\sqrt{A_m}} + \lambda \left(\sum_{m=1}^M A_m - nK_n \right) + \sum_{m \in \mathcal{M}_1} w_m (A_m - n), \end{aligned} \quad (2.25)$$

trong đó $w_m, \lambda \in \mathbb{R}$. Các điều kiện Karush–Kuhn–Tucker (KKT) đối với (2.23) được tính như sau

$$\frac{\partial \mathcal{L}_1(\{A_m^*\}_{m \in \mathcal{M}_1}, \lambda^*, \{w_m^*\}_{m \in \mathcal{M}_1})}{\partial A_m^*} = 0 \quad (2.26)$$

$$\lambda^* \geq 0$$

$$w_m^* \geq 0$$

$$w_m^* (A_m^* - n) = 0 \quad (2.27)$$

$$\lambda^* \left(\sum_{m=1}^M A_m^* - nK_n \right) = 0 \quad (2.28)$$

với $m \in \mathcal{M}_1$. Tương tự, hàm Lagrange tương ứng với (2.24) là

$$\begin{aligned} & \mathcal{L}_2(\{B_m\}_{m \in \mathcal{M}_3}, \mu, \{\nu_m\}_{m \in \mathcal{M}_3}) \\ &= \sum_{m \in \mathcal{M}_3} \frac{p_m}{\sqrt{B_m}} + \mu \left(\sum_{m=1}^M B_m - f(n)K_{FAP} \right) + \sum_{m \in \mathcal{M}_3} \nu_m (B_m - f(n)), \end{aligned} \quad (2.29)$$

trong đó $\nu_m, \mu \in \mathbb{R}$. Với $\forall m \in \mathcal{M}_3$, các điều kiện KKT dành cho (2.24) được tính như sau

$$\begin{aligned} \frac{\partial \mathcal{L}_2(\{B_m^*\}_{m \in \mathcal{M}_3}, \mu^*, \{\nu_m^*\}_{m \in \mathcal{M}_3})}{\partial B_m^*} &= 0 \\ \mu^* &\geq 0 \\ \nu_m^* &\geq 0 \\ \nu_m^*(B_m^* - f(n)) &= 0 \\ \mu^* \left(\sum_{m=1}^M B_m^* - f(n)K_{BS} \right) &= 0. \end{aligned} \quad (2.30)$$

Trước khi giải hai bài toán tối ưu trên, chúng tôi giới thiệu bổ đề sau.

Bổ đề 2.4.1. *Giả sử mạng vô tuyến hỗn hợp hướng nội dung sử dụng phương pháp truyền tin được đề xuất tại Chương 2 và $\alpha < 3/2$, giá trị nghiệm của (2.23), ký hiệu bởi A_m^* , có thuộc tính không tăng theo sự biến thiên tăng của tham số $m \in \mathcal{M}_1$ và các giá trị nghiệm của (2.24), ký hiệu bởi B_m^* , có thuộc tính không tăng theo sự biến thiên tăng của tham số $m \in \mathcal{M}_3$.*

Chứng minh. Trước hết, A_m^* sẽ được chứng minh rằng có thuộc tính không tăng theo sự biến thiên tăng của tham số $m \in \mathcal{M}_1$ như sau. Từ (2.26), ta có

$$-\frac{p_m}{2\sqrt{A_m^{*3}}} + \lambda^* + w_m^* = 0 \text{ trong đó } m \in \mathcal{M}_1. \quad (2.31)$$

$\mathcal{D}_1 \subset \mathcal{M}_1$ ký hiệu cho tập các tập dữ liệu sao cho $A_m^* = n$ và m_0 ký hiệu cho giá trị nhỏ nhất của tham số chỉ thị tập dữ liệu m thỏa mãn điều kiện $A_{m_0}^* < n$. Xét trường hợp của tập dữ liệu bất kỳ $k \in \mathcal{D}_1$, sử dụng công thức (2.31) và điều kiện thực tế là $w_{m_0}^* = 0$, ta có $\lambda^* = \frac{p_k}{2\sqrt{A_k^{*3}}} - w_k^* = \frac{p_{m_0}}{2\sqrt{A_{m_0}^{*3}}} > 0$. Do $A_k^* = n$, $A_{m_0}^* < n$, và $w_k^* \geq 0$, ta nhận được kết quả $p_k > p_{m_0}$, từ đó ta có $k < m_0$ do đặc tính của phân bố Zipf ($p_m = \frac{m^{-\alpha}}{H_\alpha(M)}$). Có nghĩa rằng, ta có

$\mathcal{D}_1 = \{1, 2, \dots, m_0 - 1\}$. Thêm vào đó, với $m \in \mathcal{M}_1 \setminus \mathcal{D}_1$, sử dụng (2.27) và (2.31), ta có $A_m^* = \frac{p_m^{\frac{2}{3}}}{(2\lambda^*)^{\frac{2}{3}}}$, giảm dần khi m biến thiên tăng lên. Do đó, A_m^* có thuộc tính không tăng theo sự biến thiên tăng của tham số $m \in \mathcal{M}_1$.

Ta chuyển sang phân tích thuộc tính của B_m^* khi $m \in \mathcal{M}_3$ biến thiên. $\mathcal{D}_2 \subset \mathcal{M}_3$ ký hiệu tập các tập dữ liệu thỏa mãn điều kiện $B_m^* = f(n)$ và \tilde{m}_0 ký hiệu số nhỏ nhất thỏa mãn $B_{\tilde{m}_0}^* < f(n)$. Bằng cách áp dụng các điều kiện KKT cho (2.24) và sử dụng phương pháp tiếp cận tương tự như trên, ta có $\mathcal{D}_2 = \{1, 2, \dots, \tilde{m}_0 - 1\}$ và $B_m^* = \frac{p_m^{\frac{2}{3}}}{(2\mu^*)^{\frac{2}{3}}}$ với $m \in \mathcal{M}_3 \setminus \mathcal{D}_2$, giảm dần khi m tăng dần. Do đó, B_m^* có thuộc tính không tăng theo sự biến thiên tăng của tham số $m \in \mathcal{M}_3$. \square

Từ bổ đề trên, các Định lý quan trọng sau đây được hình thành, thể hiện các nghiệm tối ưu của số lượng các bản ghi của các tập dữ liệu $m \in \mathcal{M}$ được lưu trữ tại bộ nhớ của các thiết bị người dùng và các trạm gốc thông tin.

Định lý 2.4.1. *Giả sử mạng vô tuyến hỗn hợp hướng nội dung sử dụng phương pháp truyền tin được đề xuất tại Chương 2, với $\alpha < 3/2$, nếu $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$, nghiệm của (2.22) là*

$$A_m^* + B_m^* = \Theta \left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})} \right).$$

trong trường hợp $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$, ta có

$$A_m^* + B_m^* = \begin{cases} \Theta \left(m^{-\frac{2\alpha}{3}} n^{\delta+(1-\delta)\frac{2\alpha}{3}} \right) & \text{trong đó } m \in \mathcal{M}_1, \\ \Theta(n^\delta) & \text{trong đó } m \in \mathcal{M}_2 \setminus \mathcal{M}_1, \\ \Theta \left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})} \right) & \text{trong đó } m \in \mathcal{M} \setminus \mathcal{M}_2, \end{cases}$$

với $\mathcal{M}_1 = \{1, \dots, m_1 - 1\}$ and $\mathcal{M}_2 = \{1, \dots, m_2 - 1\}$. Trong đó, $m_1 = \Theta(n^{1-\delta})$ và $m_2 = \Theta \left(n^{\gamma-(\gamma-\beta)\frac{3}{2\alpha}} \right)$.

Chứng minh. Như đã chỉ ra ở Mệnh đề 2.4.1, áp dụng các điều kiện KKT cho (2.23) và (2.24), ta có

$$A_m^* = \frac{p_m^{\frac{2}{3}}}{(2\lambda^*)^{\frac{2}{3}}} \text{ trong đó } m \in \mathcal{M}_1 \setminus \mathcal{D}_1 \quad (2.32)$$

và

$$B_m^* = \frac{p_m^{\frac{2}{3}}}{(2\mu^*)^{\frac{2}{3}}} \text{ trong đó } m \in \mathcal{M}_3 \setminus \mathcal{D}_2, \quad (2.33)$$

với $\mathcal{D}_1 \subset \mathcal{M}_1$ và $\mathcal{D}_2 \subset \mathcal{M}_3$ là các tập dữ liệu thỏa mãn điều kiện $A_m^* = n$ và $B_m^* = f(n)$ tương ứng. Từ phần chứng minh của Bổ đề 2.4.1, ta có $\mathcal{D}_1 = \{1, \dots, m_0 - 1\}$, với m_0 ký hiệu cho giá trị nhỏ nhất của tham số m thỏa mãn điều kiện $A_m^* < n$. Sử dụng (2.28) và (2.30), ta có $\sum_{m=1}^M A_m^* = nK_n$ và $\sum_{m=1}^M B_m^* = f(n)K_{BS}$ do $\lambda^* > 0$ và $\mu^* > 0$ tương ứng. Từ đây ta có thể tính tổng các tập dữ liệu $A_m^* + B_m^*$ với tất cả các tham số $m \in \mathcal{M}$.

Bài toán tối ưu ở (2.22) có thể được xem xét ở hai trường hợp tùy thuộc theo sự tương quan của hệ số Lagrange λ^* ở (2.25) và μ^* ở (2.29). Cụ thể, ta xem xét hai trường hợp như sau $\lambda^* = \Theta(\mu^*)$ và $\lambda^* \neq \Theta(\mu^*)$, mỗi điều kiện trên sẽ tương ứng với các trường hợp khi $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$ và $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$.

Trường hợp 1: $\lambda^* = \Theta(\mu^*)$.

Từ Chương 1, ta gọi lại công thức sau

$$H_\alpha(M) = \begin{cases} \Theta(1) & \text{với } \alpha > 1 \\ \Theta(\log M) & \text{với } \alpha = 1 \\ \Theta(M^{1-\alpha}) & \text{với } \alpha < 1. \end{cases} \quad (2.34)$$

Sử dụng (2.32) và (2.33), ta có $A_m^* + B_m^*$ có thể được tính như sau

$$A_m^* + B_m^* = \frac{p_m^{\frac{2}{3}}}{\xi^{\frac{2}{3}}}, \quad (2.35)$$

với $\xi = \Theta(\lambda^*) = \Theta(\mu^*)$ và $\forall m \in \mathcal{M} \setminus \mathcal{D}_1$. Bằng cách tính tổng $A_m^* + B_m^*$ ở (2.35) với $\forall m \in \mathcal{M} \setminus \mathcal{D}_1$ và sử dụng điều kiện thực tế là $\xi^{\frac{2}{3}} = \frac{\sum_{l=m_0}^M p_l^{\frac{2}{3}}}{\sum_{l=m_0}^M (A_l^* + B_l^*)}$, ta có

$$\begin{aligned} A_m^* + B_m^* &= \frac{p_m^{\frac{2}{3}}}{\sum_{l=m_0}^M p_l^{\frac{2}{3}}} \sum_{l=m_0}^M (A_l^* + B_l^*) \\ &= \Theta \left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})} \right). \end{aligned} \quad (2.36)$$

Ở đây, công thức thứ hai được đưa ra nhờ giá trị của tập \mathcal{D}_1 là $O(1)$ do không gian lưu trữ dữ liệu $K_n = \Theta(1)$; $\sum_{l=m_0}^M (A_l^* + B_l^*) = \Theta(n^{\delta+\beta})$ có được từ các giả thuyết $\delta + \beta \geq 1$ khi $K_{BS} = \Theta(n^\beta)$ và $f(n) = \Theta(n^\delta)$; từ $p_m = \frac{m^{-\alpha}}{H_\alpha(M)}$ và (2.34), $\sum_{l=m_0}^M p_l^{\frac{2}{3}} = \sum_{l=m_0}^M \frac{l^{-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)} = \Theta \left(\frac{M^{1-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)} \right)$ với $\alpha < \frac{3}{2}$; và $M = \Theta(n^\gamma)$.

Tiếp theo, tập dữ liệu \mathcal{D}_1 sẽ được chứng minh rằng không tồn tại. Giả sử rằng có tồn tại một tập dữ liệu \mathcal{D}_1 (tương đương với $m_0 > 1$). Số lớn nhất của tập dữ liệu \mathcal{M}_2 thỏa mãn điều kiện $A_m^* + B_m^* = \Omega(f(n))$ được ký hiệu là $m_2 - 1$, từ đó ta thấy $A_{m_2-1}^* + B_{m_2-1}^* = \Theta(f(n))$. Sử dụng (2.36), ta có $f(n) = \Theta \left((m_2 - 1)^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})} \right)$, kết quả này dẫn đến

$$m_2 = \Theta \left(n^{\gamma-(\gamma-\beta)\frac{3}{2\alpha}} \right). \quad (2.37)$$

Trong khi đó, ta có $\mathcal{M}_1 \setminus \mathcal{D}_1 = \{m_0, \dots, m_2 - 1\}$. Kết quả này là do \mathcal{M}_2 thuộc tập \mathcal{M}_1 có điều kiện là $B_m^* \leq f(n)$. Bằng cách tính tổng $A_m^* + B_m^*$ ở (2.35) với mọi $m \in \mathcal{M}_1 \setminus \mathcal{D}_1$, ta nhận được

$$A_m^* + B_m^* = \Theta(A_m^*) = \frac{p_m^{\frac{2}{3}}}{\sum_{l=m_0}^{m_2-1} p_l^{\frac{2}{3}}} \sum_{l=m_0}^{m_2-1} \Theta(A_l^*),$$

từ đó ta có

$$A_{m_0}^* = \Theta \left(\frac{\sum_{l=m_0}^{m_2-1} A_l^*}{(m_2 - 1)^{1-\frac{2\alpha}{3}}} \right) \quad (2.38)$$

với $m = m_0$ nhận được từ các công thức $p_m = \frac{m^{-\alpha}}{H_\alpha(M)}$ và (2.34), $\sum_{l=m_0}^{m_2-1} p_l^{\frac{2}{3}} = \sum_{l=m_0}^{m_2-1} \frac{l^{-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)} = \Theta\left(\frac{(m_2-1)^{1-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)}\right)$ với $\alpha < \frac{3}{2}$. Do $(m_2 - 1)^{1-\frac{2\alpha}{3}} = \omega(1)$ từ (2.37) và $\sum_{l=m_0}^{m_2-1} A_l^* = O(n)$, ta có $A_{m_0}^* = o(n)$ từ (2.38), điều này trái với điều kiện là $A_{m_0}^* = \Theta(n)$. Do đó, ta có thể kết luận rằng, giả thuyết \mathcal{D}_1 tồn tại là không hợp lý (tức là có thể kết luận giá trị $m_0 = 1$).

Phần tiếp theo sau đây, ta sẽ chỉ ra rằng điều kiện $\lambda^* = \Theta(\mu^*)$ là tương đương với $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$. Sử dụng (2.36) và (2.37) cũng như $m_0 = 1$, ta có $\sum_{m=1}^{m_2-1} (A_m^* + B_m^*) = \frac{\sum_{m=1}^{m_2-1} p_m^{\frac{2}{3}}}{\sum_{l=1}^M p_l^{\frac{2}{3}}} \sum_{l=1}^M (A_l^* + B_l^*) = \Theta\left(n^{\delta+\beta-(\gamma-\beta)\left(\frac{3}{2\alpha}-1\right)}\right)$. Sử dụng thực tế là $\sum_{m=1}^{m_2-1} (A_m^* + B_m^*) = \sum_{m=1}^{m_2-1} \Theta(A_m^*) = O(n)$, ta có bất đẳng thức sau đây: $\delta + \beta - (\gamma - \beta) \left(\frac{3}{2\alpha} - 1\right) \leq 1$, tương đương với $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$. Do đó, nếu $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$, thì ta có

$$A_m^* + B_m^* = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma\left(1-\frac{2\alpha}{3}\right)}\right) \text{ trong đó } m \in \mathcal{M}.$$

Trường hợp 2: $\lambda^* \neq \Theta(\mu^*)$.

Từ (2.32) và (2.33), ta nhận thấy rằng hai tập $\mathcal{M}_1 \setminus \mathcal{D}_1$ và $\mathcal{M}_3 \setminus \mathcal{D}_2$ không giao nhau. Do đó, ta có $\mathcal{M}_1 \subset \mathcal{M}_2$. Từ nhận xét này, ta có

$$A_m^* + B_m^* = \Theta(B_m^*) = \Theta(n^\delta) \quad (2.39)$$

với $m \in \mathcal{M}_2 \setminus \mathcal{M}_1$. $m_1 - 1$ và $m_2 - 1$ được ký hiệu là các số lớn nhất của các tập dữ liệu \mathcal{M}_1 và \mathcal{M}_2 tương ứng, ta có thể tính $A_m^* + B_m^*$ với $m \in \mathcal{M}_1$ và $m \in \mathcal{M} \setminus \mathcal{M}_2$.

Trước hết, ta sẽ tính $A_m^* + B_m^*$ và m_2 với $m \in \mathcal{M} \setminus \mathcal{M}_2$, ở đây $\mathcal{M} \setminus \mathcal{M}_2 = \{m_2, \dots, M\}$. Bằng cách tính tổng B_m^* ở (2.33) với mọi giá trị $m \in \mathcal{M} \setminus \mathcal{M}_2$,

ta có

$$B_m^* = \frac{p_m^{\frac{2}{3}}}{\sum_{l=m_2}^M p_l^{\frac{2}{3}}} \sum_{l=m_2}^M B_l^* \quad (2.40a)$$

$$= O\left(m^{-\frac{2\alpha}{3}} n^{\delta+\beta-\gamma(1-\frac{2\alpha}{3})}\right), \quad (2.40b)$$

trong đó, công thức thứ hai có được là do $\sum_{l=m_2}^M B_l^* = O(n^{\delta+\beta})$; từ $p_m = \frac{m^{-\alpha}}{H_\alpha(M)}$ và (2.34), $\sum_{l=m_2}^M p_l^{\frac{2}{3}} = \sum_{l=m_2}^M \frac{l^{-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)} = \Theta\left(\frac{M^{1-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)}\right)$ với $\alpha < \frac{3}{2}$; và $M = \Theta(n^\gamma)$. Từ (2.40b) và $B_{m_2}^* = \Theta(f(n))$, ta có $f(n) = O\left(m_2^{-\frac{2\alpha}{3}} n^{\delta+\beta-\gamma(1-\frac{2\alpha}{3})}\right)$, từ đây ta có $m_2 = O\left(n^{\gamma - (\gamma-\beta)\frac{3}{2\alpha}}\right)$. Do $(\gamma - \beta)\left(1 - \frac{3}{2\alpha}\right) < 0$, ta nhận thấy $\gamma - (\gamma - \beta)\frac{3}{2\alpha} < \beta$, kết quả là $m_2 = o(K_{BS})$. Do $\sum_{l=m_2}^M B_l^* = f(n)K_{BS} - O(m_2 f(n)) = \Theta(n^{\delta+\beta})$, (2.40a) có thể được trình bày như sau

$$A_m^* + B_m^* = \Theta(B_m^*) = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\delta+\beta-\gamma(1-\frac{2\alpha}{3})}\right) \quad (2.41)$$

với $m \in \mathcal{M} \setminus \mathcal{M}_2$. Hơn nữa, do $A_{m_2}^* + B_{m_2}^* = \Theta\left(m_2^{-\frac{2\alpha}{3}} n^{\delta+\beta-\gamma(1-\frac{2\alpha}{3})}\right)$ từ (2.41) và $A_{m_2}^* + B_{m_2}^* = \Theta(f(n))$, ta có kết quả sau

$$m_2 = \Theta\left(n^{\gamma - (\gamma-\beta)\frac{3}{2\alpha}}\right). \quad (2.42)$$

Tiếp theo, chúng ta sẽ tìm $A_m^* + B_m^*$ và m_1 với $m \in \mathcal{M}_1$. Bằng cách tính tổng A_m^* ở (2.32) với các giá trị $\forall m \in \mathcal{M}_1 \setminus \mathcal{D}_1 (= \{m_0, \dots, m_1 - 1\})$, ta có

$$A_m^* = \frac{p_m^{\frac{2}{3}}}{\sum_{l=m_0}^{m_1-1} p_l^{\frac{2}{3}}} \sum_{l=m_0}^{m_1-1} A_l^*. \quad (2.43)$$

Tương tự như Trường hợp 1, chúng ta có thể chứng minh tập dữ liệu \mathcal{D}_1 không tồn tại (điều này tương đương với việc $m_0 = 1$). Từ kết quả $f(n) = \Theta\left(\frac{m_1^{-\frac{2\alpha}{3}} \sum_{l=1}^{m_1-1} A_l^*}{(m_1-1)^{1-\frac{2\alpha}{3}}}\right)$ do $\sum_{l=1}^{m_1-1} p_l^{\frac{2}{3}} = \Theta\left(\frac{(m_1-1)^{1-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)}\right)$ và $A_{m_1-1}^* = \Theta(f(n))$, ta có kết quả sau

$$m_1 = \Theta\left(\frac{\sum_{l=1}^{m_1-1} A_l^*}{f(n)}\right). \quad (2.44)$$

Tiếp theo, chúng ta cần xem xét $\sum_{l=1}^{m_1-1} A_l^*$ để tìm giá trị của m_1 . Với $\alpha < \frac{3}{2}$, bằng cách thiết lập $m_0 = 1$, công thức bài toán tối ưu(2.22a) có thể được trình bày lại như sau

$$\begin{aligned}
& \sum_{m=1}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}} \\
&= \sum_{m=1}^{m_1-1} \frac{p_m}{\sqrt{A_m^* + B_m^*}} + \sum_{m=m_1}^{m_2-1} \frac{p_m}{\sqrt{A_m^* + B_m^*}} + \sum_{m=m_2}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}} \\
&= \Theta \left(\frac{\left(\sum_{m=1}^{m_1-1} p_m^{\frac{2}{3}} \right)^{\frac{3}{2}}}{\sqrt{\sum_{m=1}^{m_1-1} A_m^*}} \right) + \Theta \left(\frac{\sum_{m=m_1}^{m_2-1} p_m}{\sqrt{f(n)}} \right) + \Theta \left(\frac{\left(\sum_{m=m_2}^M p_m^{\frac{2}{3}} \right)^{\frac{3}{2}}}{\sqrt{\sum_{m=m_2}^M B_m^*}} \right) \\
&= \Theta \left(\frac{m_1^{1-\alpha}}{H_\alpha(M) \sqrt{f(n)}} \right) + \Theta \left(\frac{\max\{H_\alpha(m_1), H_\alpha(m_2-1)\}}{H_\alpha(M) \sqrt{f(n)}} \right) \\
&+ \Theta \left(\frac{M^{3/2-\alpha}}{H_\alpha(M) \sqrt{f(n)} K_{BS}} \right), \tag{2.45}
\end{aligned}$$

trong đó, công thức thứ hai nhận được là kết quả từ (2.39) với $m \in \mathcal{M}_2 \setminus \mathcal{M}_1$, (2.40a) với $m \in \mathcal{M} \setminus \mathcal{M}_2$, và (2.43) với $m \in \mathcal{M}_1$; công thức thứ ba nhận được khi $m_2 = o(M)$ từ (2.42) và $\sum_{l=1}^{m_1-1} p_l^{\frac{2}{3}} = \Theta \left(\frac{(m_1-1)^{1-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)} \right)$ và $\sum_{l=m_2}^M p_l^{\frac{2}{3}} = \Theta \left(\frac{M^{1-\frac{2\alpha}{3}}}{H_\alpha^{\frac{2}{3}}(M)} \right)$ từ $p_m = \frac{m^{-\alpha}}{H_\alpha(M)}$ và (2.34). Ta nhận thấy rằng, thành phần thứ hai ở (2.45) biến thiên chậm hơn các thành phần khác do $\alpha \leq 1$, (2.22a) được tính bởi thành phần thứ ba ở (2.45) và với $1 < \alpha < \frac{3}{2}$, (2.22a) sẽ có giá trị phụ thuộc vào độ lớn của thành phần thứ nhất hoặc thứ ba tại công thức (2.45). Do $1 < \alpha < \frac{3}{2}$, thành phần thứ nhất (bao gồm cả m_1) có thể tối thiểu hóa khi m_1 đạt giá trị nhỏ nhất, từ đó ta có $\sum_{l=1}^{m_1-1} A_l^* = \Theta(n)$ từ (2.3) (tổng các giới hạn) và (2.44). Từ đó, ta có

$$m_1 = \Theta(n^{1-\delta}).$$

Hơn nữa, từ thiết lập $m_0 = 1$, $\sum_{l=1}^{m_1-1} p_l^{\frac{2}{3}} = \Theta\left(\frac{m_1^{1-\frac{2\alpha}{3}}}{H_{\frac{2}{3}}^{\alpha}(M)}\right)$, và $\sum_{l=1}^{m_1-1} A_l^* = \Theta(n)$, (2.43) có thể được trình bày như sau

$$A_m^* + B_m^* = \Theta(A_m^*) = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\delta+(1-\delta)\frac{2\alpha}{3}}\right) \text{ trong đó } m \in \mathcal{M}_1.$$

Trong trường hợp sau cùng, với $m \in \mathcal{M}_2 \setminus \mathcal{M}_1$, $A_m^* + B_m^*$ đạt giá trị $\Theta(n^\delta)$ theo định nghĩa của \mathcal{M}_1 và \mathcal{M}_2 .

Kết luận là, từ phân kết luận của Trường hợp 1, chúng ta có $\lambda^* \neq \Theta(\mu^*)$ khi $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$. \square

Các nghiệm tối ưu ở Định lý 2.4.1 được thể hiện ở Hình 2.4. Với $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$, số lượng các bản ghi tối ưu của tập dữ liệu m , $A_m^* + B_m^*$, giảm đều khi tham số m tăng lên. Với $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$, ta thấy có tồn tại một tập các tập dữ liệu sao cho $A_m^* + B_m^*$ có giá trị $\Theta(f(n))$.

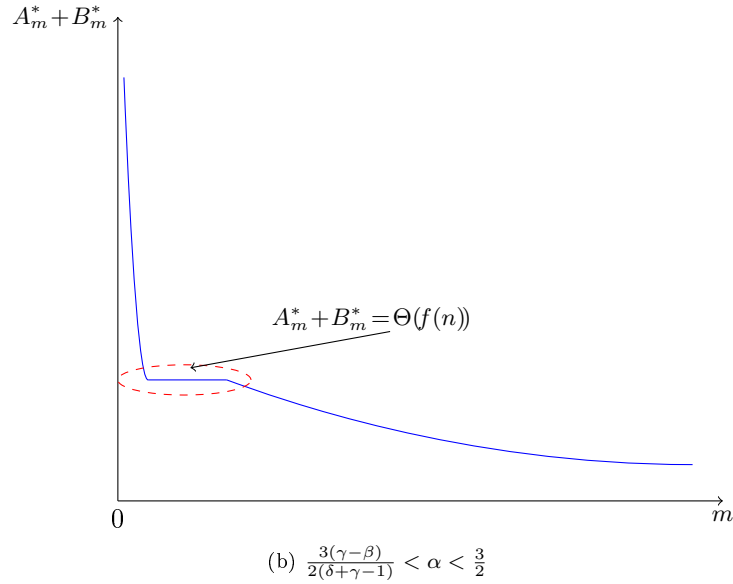
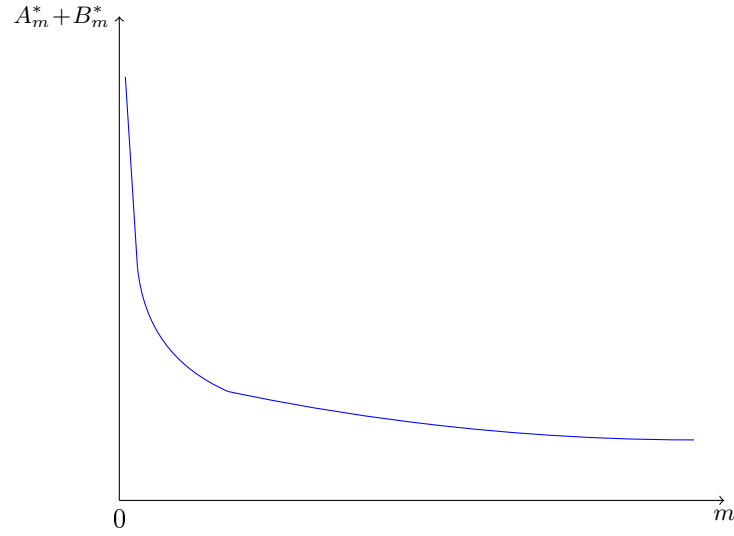
Từ kết quả trên, chúng ta có thể lựa chọn các giá trị tối ưu bản ghi $\{A_m^*\}_{m=1}^M$ và $\{B_m^*\}_{m=1}^M$ sao cho vẫn đảm bảo được giá trị tối ưu ở các Định lý trên, ta có Mệnh đề sau.

Mệnh đề 2.4.1. *Giả sử mạng vô tuyến hỗn hợp hướng nội dung sử dụng phương pháp truyền tin được đề xuất tại Chương 2 và $\alpha < 3/2$, giá trị tối ưu của $\{A_m^*\}_{m=1}^M$ và $\{B_m^*\}_{m=1}^M$ được tính như sau*

$$A_m^* = \begin{cases} \Theta\left(m^{-\frac{2\alpha}{3}} n^{\min\{\beta+\delta-\gamma(1-\frac{2\alpha}{3}), \delta+(1-\delta)\frac{2\alpha}{3}\}}\right) & \text{trong đó } m \in \mathcal{M}_1 \cap \mathcal{M}_2, \\ 0 & \text{trong đó } m \in \mathcal{M} \setminus (\mathcal{M}_1 \cap \mathcal{M}_2) \end{cases} \quad (2.46)$$

và

$$B_m^* = \begin{cases} \Theta(n^\delta) & \text{trong đó } m \in \mathcal{M}_2, \\ \Theta\left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})}\right) & \text{trong đó } m \in \mathcal{M} \setminus \mathcal{M}_2, \end{cases}$$



Hình 2.4: Phương pháp tối ưu lưu trữ dữ liệu tương ứng với sự biến thiên của tham số m .

tương ứng, với $\mathcal{M}_1 = \{1, \dots, m_1 - 1\}$ và $\mathcal{M}_2 = \{1, \dots, m_2 - 1\}$. Trong đó, $m_1 = \Theta(n^{1-\delta})$ và $m_2 = \Theta\left(n^{\gamma-(\gamma-\beta)\frac{3}{2\alpha}}\right)$.

Chứng minh. Việc lựa chọn các tập giá trị tối ưu $\{A_m^*\}_{m=1}^M$ và $\{B_m^*\}_{m=1}^M$ thỏa mãn nghiệm tối ưu của $\{A_m^* + B_m^*\}_{m=1}^M$ trong Định lý 2.4.1 tùy theo giá trị của tham số α .

Trước hết, ta xem xét trường hợp $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$. Với $m \in \mathcal{M}_2$ và $A_m^* +$

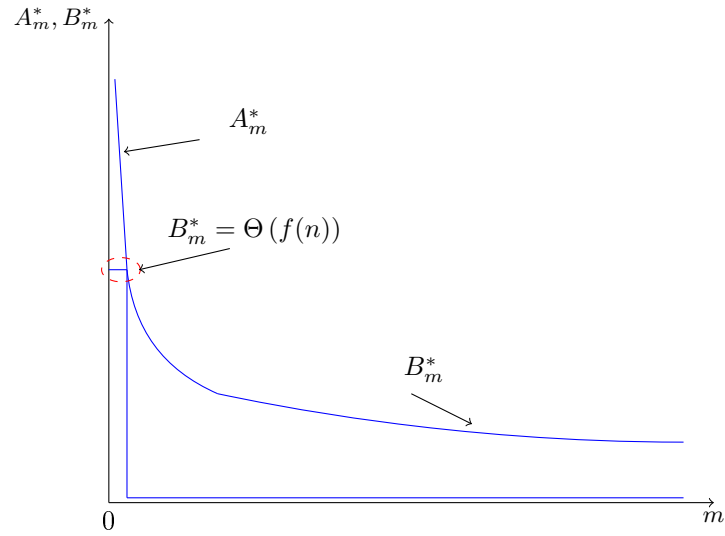
$B_m^* = \Omega(f(n)) (= \Omega(n^\delta))$, ta có $A_m^* = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})}\right)$ và $B_m^* = \Theta(n^\delta)$. Với $\mathcal{M} \setminus \mathcal{M}_2$ và $A_m^* + B_m^* = o(f(n))$, ta có $A_m^* = 0$ và $B_m^* = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})}\right)$.

Tiếp theo, ta xem xét trường hợp $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$. Với $m \in \mathcal{M}_1$ và $A_m^* + B_m^* = \Theta(A_m^*)$, ta có $A_m^* = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\delta+(1-\delta)\frac{2\alpha}{3}}\right)$ and $B_m^* = \Theta(n^\delta)$. Với $m \in \mathcal{M}_2 \setminus \mathcal{M}_1$ và $A_m^* + B_m^* = \Theta(f(n))$, ta có $A_m^* = 0$ và $B_m^* = \Theta(n^\delta)$. Tương tự trường hợp trên, ta có $A_m^* = 0$ và $B_m^* = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})}\right)$ với $m \in \mathcal{M} \setminus \mathcal{M}_2$.

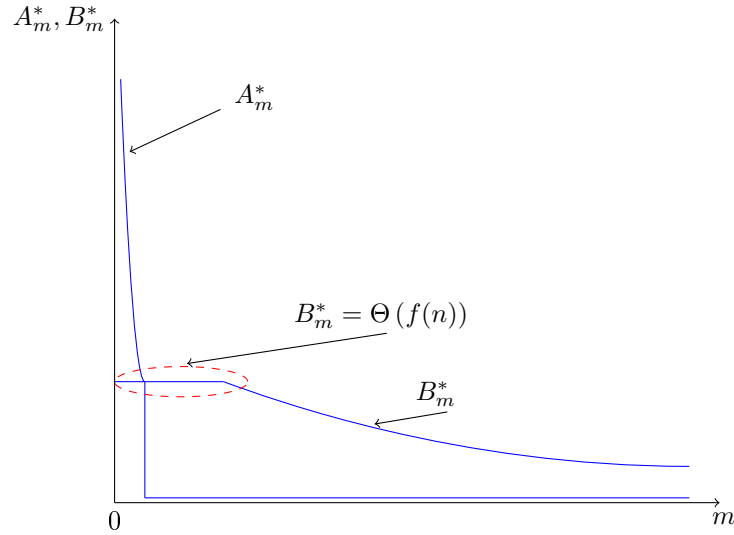
Từ thực tế là $m_1 = \Theta(n^{1-\delta})$ và $m_2 = \Theta\left(n^{\gamma-(\gamma-\beta)\frac{3}{2\alpha}}\right)$, ta có $\mathcal{M}_1 \cap \mathcal{M}_2 = \mathcal{M}_2$ trong trường hợp $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$. Đối với các trường hợp còn lại, $\mathcal{M}_1 \cap \mathcal{M}_2 = \mathcal{M}_1$. Như vậy, với các giá trị được đưa ra như trên, A_m^* có thể được thể hiện như (2.46). \square

Phương pháp lưu trữ dữ liệu tối ưu tại Mệnh đề 2.4.1 được thể hiện ở Hình 2.5. Từ kết quả trên, ta có nhận xét như sau.

Chú ý 2.4.1. Từ Hình 2.5, ta nhận thấy rằng, các tập dữ liệu có tính chất phổ biến cao và đang là xu hướng được quan tâm lớn, số lượng bản ghi được lưu trong mạng được tính bởi $\omega(f(n))$, chủ yếu sẽ được truyền đi bởi phương pháp truyền tin đa chặng Người dùng tới người dùng. Trong khi đó, các tập dữ liệu nhận được sự quan tâm ít hơn sẽ chủ yếu được phục vụ bởi các trạm gốc thông tin. Tức là, từ Mệnh đề 2.4.1, các tập dữ liệu $m \in \mathcal{M}_1 \cap \mathcal{M}_2$ được lưu trữ chủ yếu tại các thiết bị người dùng di động sẽ giúp cho mạng đạt được mức thông lượng và độ trễ tốt nhất.



$$(a) \alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$$



$$(b) \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$$

Hình 2.5: Các tập dữ liệu $\{A_m^*\}_{m=1}^M$ và $\{B_m^*\}_{m=1}^M$ biến thiên theo tham số m .

2.4.3. Nghiệm tối ưu hóa sử dụng phần mềm tính toán trên máy tính

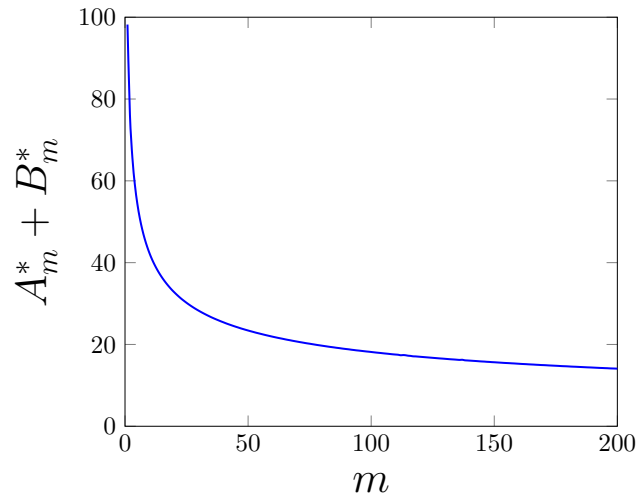
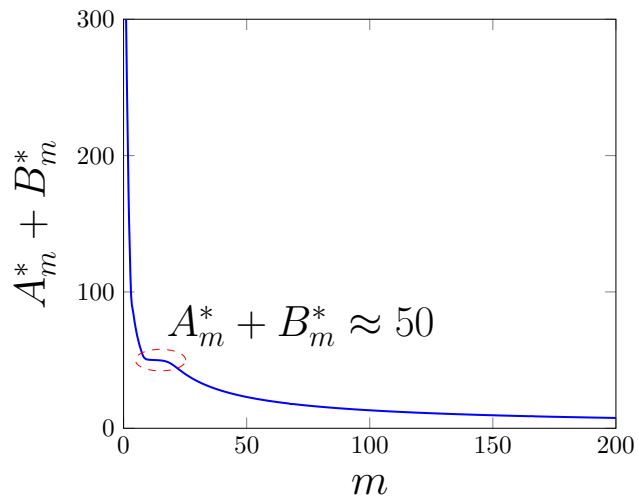
Trong phần này, để kiểm tra lại tính đúng đắn của các kết quả phân tích thể hiện ở Phần trước 2.4.2, phần mềm giải toán Mathematica trên máy tính sẽ được sử dụng nhiều lần để tìm ra nghiệm tối ưu của (2.22) tương ứng theo các tham số hệ thống được cho trước n , M , K_n , K_{BS} , và $f(n)$ ở Bảng 2.1. Từ các tham số này cho trước, theo định nghĩa của các tham số mạng

Bảng 2.1: Bảng các tham số mạng được sử dụng để tính toán trên máy tính

Ký hiệu	Mô tả	Giá trị thiết lập
n	Số lượng thiết bị di động người dùng	300
M	Số lượng tệp dữ liệu trong thư viện mạng	200
$f(n)$	Số lượng trạm gốc thông tin	50
K_{BS}	Độ lớn lưu trữ của trạm gốc thông tin	50
K_n	Độ lớn lưu trữ của thiết bị di động người dùng	2

M , $f(n)$, K_{BS} , và $f(n)$ ta có các tham số γ , β , và δ có giá trị lần lượt là 0.93, 0.69, và 0.69, tương ứng. Như phần nội dung 2.4.2, có hai trường hợp $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$ và $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$, với $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} = 0.59$. Với hai trường hợp xảy ra trong quá trình tìm nghiệm tối ưu ở (2.22), hai giá trị của α được sử dụng để tính toán bởi phần mềm giải toán Mathematica là: $\alpha = 0.55$ và $\alpha = 1.2$. Ở Hình 2.6, tập nghiệm $\{A_m^* + B_m^*\}_{m=1}^{200}$ được thể hiện tương ứng với tham số m . Có thể nhận thấy rằng, kết quả tính toán bởi phần mềm máy tính giải toán Mathematica là phù hợp với các kết quả phân tích được đưa ra tại Định lý 2.4.1, như thể hiện ở Hình 2.4. Cụ thể hơn, với trường hợp $\alpha = 0.55$, Hình 3.5(a) thể hiện rằng $A_m^* + B_m^*$ có đặc tính giảm tuyến tính khi giá trị của tham số m tăng lên. Bên cạnh đó, Như thể hiện ở hình 3.5(b), luôn tồn tại một tập nghiệm các tệp dữ liệu m sao cho $A_m^* + B_m^*$ có giá trị xấp xỉ $f(n) = 50$ với trường hợp $\alpha = 1.2$.

Ở Hình 2.6, các tập nghiệm $\{A_m^*\}_{m=1}^{200}$ và $\{B_m^*\}_{m=1}^{200}$ được trình bày tương ứng với sự biến thiên của tham số m . Hướng biến thiên của các tập nghiệm là tương tự như kết quả phân tích được đưa ra tại Mệnh đề 2.4.1, như thể hiện ở Hình 2.5. Các kết quả tính toán máy tính cho thấy các tệp dữ liệu phổ biến nhất trong mạng có xu hướng được lưu trữ và truyền đi chủ yếu bởi các thiết bị di động người dùng. Kết quả chỉ ra rằng khi B_m^* giữ nguyên giá trị

(a) $\alpha = 0.55$ (b) $\alpha = 1.2$

Hình 2.6: Tập nghiệm tối ưu $\{A_m^* + B_m^*\}_{m=1}^{200}$ tương ứng với tập dữ liệu m .

xấp xỉ $f(n) = 50$ khi tham số m nhỏ và sau đó giảm dần khi giá trị của m tăng lên.

2.4.4. Thông lượng và độ trễ tối ưu

Từ các Định lý 2.3.1, Định lý 2.4.1 và Mệnh đề 2.4.1, mức cân bằng tối ưu thông lượng và độ trễ của mạng có thể tính được. Cụ thể, hàm tối ưu (2.22a)

được tính như sau

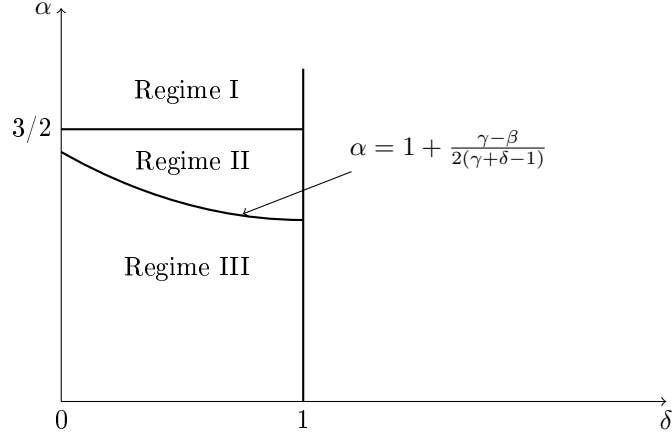
$$\sum_{m=1}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}} = \Theta \left(\frac{n^{\gamma(\frac{3}{2}-\alpha) - \frac{\beta+\delta}{2}}}{H_\alpha(M)} \right),$$

với $H_\alpha(M) = \sum_{i=1}^M i^{-\alpha}$. Trong đó $\frac{3(\gamma-\beta)}{2(\delta+\gamma-1)} < \alpha < \frac{3}{2}$, ta có

$$\begin{aligned} & \sum_{m=1}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}} \\ &= \sum_{m=1}^{m_1-1} \frac{p_m}{\sqrt{A_m^* + B_m^*}} + \sum_{m=m_1}^{m_2-1} \frac{p_m}{\sqrt{A_m^* + B_m^*}} + \sum_{m=m_2}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}} \\ &= \Theta \left(\frac{n^{(1-\delta)(\frac{3}{2}-\alpha) - \frac{1}{2}}}{H_\alpha(M)} \right) + \Theta \left(\frac{n^{-\frac{\delta}{2}} \max\{H_\alpha(m_1), H_\alpha(m_2)\}}{H_\alpha(M)} \right) \\ &+ \Theta \left(\frac{n^{\gamma(\frac{3}{2}-\alpha) - \frac{\beta+\delta}{2}}}{H_\alpha(M)} \right). \end{aligned} \quad (2.47)$$

Các giá trị thứ nhất, thứ hai và thứ ba của vế phải của công thức (2.47) được ký hiệu bởi F_1 , F_2 , và F_3 , tương ứng. Theo đó, có thể nhận thấy là $F_2 = \Theta(F_1)$ với $1 < \alpha < \frac{3}{2}$ và $F_2 = O(F_3)$ với $\alpha \leq 1$. Thêm vào đó, với $\alpha \leq \frac{3(\gamma-\beta)}{2(\delta+\gamma-1)}$, hàm tối ưu $\sum_{m=1}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}}$ ở (2.22a) sẽ có giá trị biến thiên theo F_3 . Do đó, chúng ta cần so sánh giá trị biến thiên của F_1 và F_3 theo các tham số mạng α , δ , β , và γ để tìm ra giá trị biến thiên của (2.22a). Trong nghiên cứu này, chúng tôi chia các trường hợp có thể, bao gồm cả trường hợp $\alpha \geq \frac{3}{2}$, thành các chế độ hoạt động của mạng (regime) như sau (Xem Hình 2.7).

- Regime I (Chế độ Zipf cao): $\{\alpha | \alpha \geq \frac{3}{2}\}$
- Regime II (Chế độ Zipf thông thường):
 $\left\{ \alpha \left| 1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)} \leq \alpha < \frac{3}{2} \right. \right\}$
- Regime III (Chế độ Zipf thấp):



Hình 2.7: Các chế độ hoạt động của mạng theo mối quan hệ của các tham số α , δ , β , và γ .

$$\left\{ \alpha \mid \alpha < 1 + \frac{\gamma - \beta}{2(\gamma + \delta - 1)} \right\}$$

Từ việc phân chia thành các regime như trên, chúng ta có Định lý thể hiện giá trị tối ưu của mức cân bằng thông lượng và độ trễ như sau.

Định lý 2.4.2. *Giả sử mạng vô tuyến hỗn hợp hướng nội dung sử dụng phương pháp truyền tin được đề xuất tại Chương 2 và phương pháp lưu trữ các bản ghi của các tệp dữ liệu trong mạng được tối ưu theo Mệnh đề 2.4.1, tùy theo sự biến thiên của hệ số Zipf α và các tham số γ , δ , và β , mức cân bằng thông lượng và độ trễ được tính như sau*

$$\lambda(n) = \Theta\left(\frac{D(n)}{n^b}\right), \text{ với } \lambda(n) = O\left(\frac{1}{\sqrt{n^{b+\epsilon}}}\right),$$

với hằng số $\epsilon > 0$. Trong đó,

$$b = \begin{cases} 0 & \text{tại Regime I,} \\ (1 - \delta)(3 - 2\alpha) & \text{tại Regime II,} \\ 1 - \delta - \beta + \min\{3 - 2\alpha, 1\} \gamma & \text{tại Regime III.} \end{cases}$$

Chứng minh. Ở chế độ Regime I, $\lambda(n) = \Theta(D(n))$. Ở chế độ Regime II, ta thấy rằng (2.22a) biến thiên theo $F_1 = \Theta\left(n^{(1-\delta)(\frac{3}{2}-\alpha)-\frac{1}{2}}\right)$, dẫn đến kết

Bảng 2.2: Bảng giá trị thông lượng và độ trễ mạng tối ưu

α	$\lambda(n)$	$D(n)$
$\alpha \geq \frac{3}{2}$	$\Theta\left(\frac{1}{\log n}\right)$	$\Theta(1)$
$\frac{3}{2} > \alpha \geq 1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)}$	$\Theta\left(\frac{1}{n^{(1-\delta)(\frac{3}{2}-\alpha)}\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{(1-\delta)(\frac{3}{2}-\alpha)}}{\sqrt{\log n}}\right)$
$1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)} > \alpha \geq 1$	$\Theta\left(\frac{1}{n^{\gamma(\frac{3}{2}-\alpha)-\frac{\delta+\beta-1}{2}}\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{\gamma(\frac{3}{2}-\alpha)-\frac{\delta+\beta-1}{2}}}{\sqrt{\log n}}\right)$
$1 > \alpha$	$\Theta\left(\frac{1}{n^{\frac{\gamma}{2}-\frac{\delta+\beta-1}{2}}\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{\frac{\gamma}{2}-\frac{\delta+\beta-1}{2}}}{\sqrt{\log n}}\right)$

quả là $b = (1 - \delta)(3 - 2\alpha)$ ở công thức (2.18). Đối với chế độ Regime III: với $1 < \alpha < 1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)}$, (2.22a) biến thiên theo $F_3 = \Theta\left(n^{\gamma(\frac{3}{2}-\alpha)-\frac{\delta+\beta}{2}}\right)$; và với $\alpha \leq 1$, (2.22a) biến thiên theo $F_3 = \Theta\left(n^{\frac{\gamma}{2}-\frac{\delta+\beta}{2}}\right)$. Do đó, ta có $b = 1 - \delta - \beta + \min\{3 - 2\alpha, 1\}$. \square

Xuất phát từ những tính toán ở phần chứng minh trên, thông lượng và độ trễ mạng tối ưu phụ thuộc theo các tham số mạng $\alpha, \beta, \gamma, \delta$ có thể được tổng hợp theo Bảng 2.2.

2.5. Hiệu năng mạng vô tuyến hướng nội dung sử dụng phương pháp lưu trữ dữ liệu cơ bản

Do mô hình mạng đề xuất tại Chương này là nghiên cứu đầu tiên đối với mạng hỗn hợp vô tuyến sử dụng kỹ thuật đệm dữ liệu, trong đó bao gồm n thiết bị di động người dùng và $f(n)$ trạm gốc thông tin di động được trang bị các bộ nhớ lưu trữ chia sẻ có giới hạn, một giải pháp lưu trữ cơ bản sẽ được đưa ra phân tích và so sánh với giải pháp tối ưu đã đưa ra. Cụ thể, các tập nghiệm tối ưu của số lượng bản sao của các tệp dữ liệu tại các thiết bị di động $\{A_m\}_{m=1}^M$ và trạm gốc thông tin $\{B_m\}_{m=1}^M$ sẽ được phân tích và tối ưu riêng lẻ thay vì giải bài toán tối ưu chung ở phương trình (2.22). Các

bài toán tối ưu trước hết sẽ được xác định và dựa trên đó để tìm ra các tập nghiệp tối ưu một cách độc lập với nhau. Sau đó, các so sánh về hiệu năng mạng sẽ được thực hiện giữa hai phương pháp sử dụng kỹ thuật đệm dữ liệu được đưa ra trong Chương này.

2.5.1. Xây dựng bài toán tối ưu hóa thông lượng và độ trễ

Trong phương pháp tối ưu cơ bản, giả sử rằng tất cả các tệp dữ liệu đều được lưu trữ tại *các thiết bị di động người dùng* (tương đương với trường hợp $K_{BS} = 0$) hoặc *chỉ tại các trạm gốc thông tin* (tương ứng với trường hợp $K_n = 0$). Khi đó, hàm tối ưu (2.22a) trở thành $\sum_{m=1}^M \frac{p_m}{\sqrt{A_m}}$ hoặc $\sum_{m=1}^M \frac{p_m}{\sqrt{B_m}}$ khi mà các giá trị $B_m = 0$ và $A_m = 0$ tương ứng. Chi tiết hơn, hai hàm tối ưu được đưa ra như sau:

$$\min_{\{A_m\}_{m=1}^M} \sum_{m=1}^M \frac{p_m}{\sqrt{A_m}} \quad (2.48a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M A_m \leq nK_n, \quad (2.48b)$$

$$1 \leq A_m \leq n \text{ trong đó } m \in \mathcal{M} \quad (2.48c)$$

và

$$\min_{\{B_m\}_{m=1}^M} \sum_{m=1}^M \frac{p_m}{\sqrt{B_m}} \quad (2.49a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M B_m \leq f(n)K_{BS}, \quad (2.49b)$$

$$1 \leq B_m \leq f(n) \text{ trong đó } m \in \mathcal{M}. \quad (2.49c)$$

Sử dụng phương pháp phân tích giải bài toán tối ưu tương tự như ở Mục 2.4, hai tập nghiệp A_m và B_m được xem xét là các số thực $[1, \infty)$, từ đó các hàm (2.48a) và (2.49a) có thể được chứng minh là hàm lồi và có thể giải và

tìm được nghiệm tương ứng.

2.5.2. Giải bài toán tối ưu hóa và so sánh

Tương tự như Mục 2.4.2, phương pháp Lagrange có thể sử dụng để giải các bài toán tối ưu được nêu ra tại (2.48) và (2.49). Từ đó, hai Định lý sau đây về số lượng các bản sao tối ưu của tệp dữ liệu $m \in \mathcal{M}$ được thành lập.

Định lý 2.5.1. *Giả sử rằng $\alpha < 3/2$ và phương pháp truyền tin ở mục 2.2 được sử dụng cho mạng vô tuyến hỗn hợp hướng nội dung đề xuất. Giá trị tối ưu của (2.48) và (2.49) là*

$$A_m^* + B_m^* = \begin{cases} \Theta \left(m^{-\frac{2\alpha}{3}} n^{1-\gamma(1-\frac{2\alpha}{3})} \right) & \text{với } m \in \mathcal{M}_4, \\ \Theta(n^\delta) & \text{với } m \in \mathcal{M}_2 \setminus \mathcal{M}_4, \\ \Theta \left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})} \right) & \text{với } m \in \mathcal{M} \setminus \mathcal{M}_2, \end{cases}$$

với $\mathcal{M}_2 = \{1, \dots, m_2 - 1\}$ và $\mathcal{M}_4 = \{1, \dots, m_4 - 1\}$. Trong đó, $m_2 = \Theta \left(n^{\gamma-(\gamma-\beta)\frac{3}{2\alpha}} \right)$ và $m_4 = \Theta \left(n^{\gamma-(\gamma+\delta-1)\frac{3}{2\alpha}} \right)$.

Chứng minh. Do bài toán tối ưu thứ nhất (2.48) tương ứng với trường hợp mạng vô tuyến không chứa trạm gốc thông tin, tập nghiệm tối ưu của $\{A_m^*\}_{m=1}^M$ được cho bởi

$$A_m^* = \Theta \left(m^{-\frac{2\alpha}{3}} n^{1-\gamma(1-\frac{2\alpha}{3})} \right) \quad (2.50)$$

với $\alpha < \frac{3}{2}$ [36].

Tiếp theo, bài toán thứ hai ở (2.49) sẽ được giải, tương ứng với trường hợp mạng vô tuyến hướng nội dung chỉ cho phép sử dụng kỹ thuật đệm dữ liệu đối với các bộ nhớ lưu trữ chia sẻ của các trạm gốc thông tin. Sử dụng các lập luận tương tự như phần Chứng minh của Định lý 2.4.1 dẫn đến kết quả

$B_m^* = \Theta(f(n)) = \Theta(n^\delta)$ với $m \in \mathcal{M}_2 \triangleq \{1, \dots, m_2 - 1\}$ và

$$B_m^* = \Theta\left(m^{-\frac{2\alpha}{3}} n^{\beta+\delta-\gamma(1-\frac{2\alpha}{3})}\right) \quad (2.51)$$

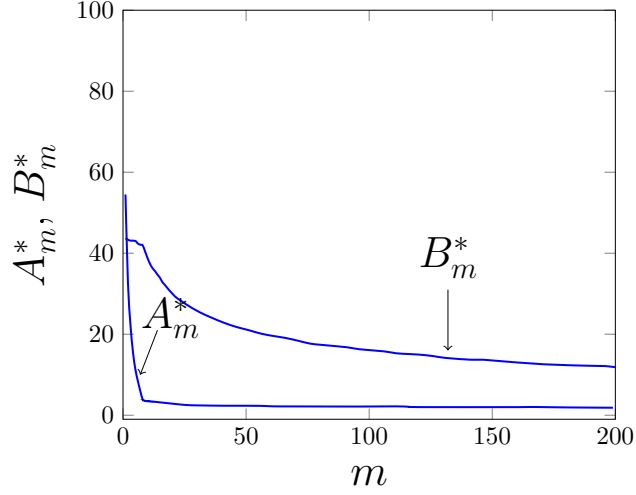
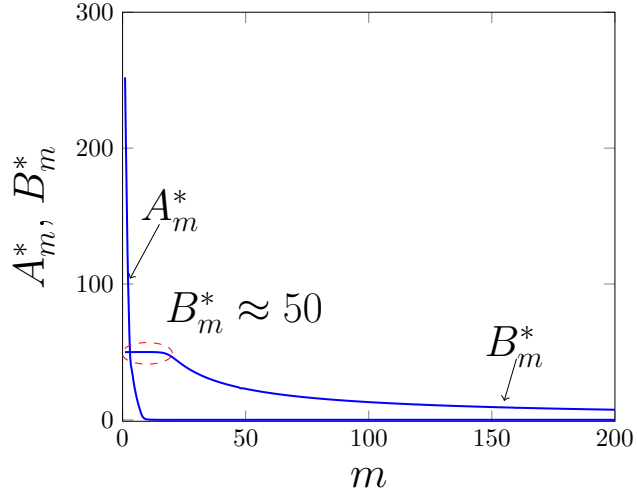
với $m \in \mathcal{M} \setminus \mathcal{M}_2$, trong đó $m_2 = \Theta\left(n^{\gamma-(\gamma-\beta)\frac{3}{2\alpha}}\right)$.

Từ (2.50) và (2.51), ta cần xem xét và so sánh tính biến thiên và độ lớn tương ứng của các tập nghiệm A_m^* và B_m^* với $m \in \mathcal{M}$. Sử dụng điều kiện $\delta + \beta \geq 1$, ta có $B_m^* = \Omega(A_m^*)$ với $m \in \mathcal{M} \setminus \mathcal{M}_2$. Ký hiệu $\mathcal{M}_4 \triangleq \{1, \dots, m_4 - 1\}$ là tập các tập dữ liệu sao cho $A_m^* + B_m^* = \Theta(A_m^*)$ với trường hợp lưu trữ cơ bản, với $m_4 - 1$ là hệ số lớn nhất của tập \mathcal{M}_4 . Từ đó, sử dụng (2.50) và thực tế rằng $A_{m_4-1}^* = \Theta(B_{m_4-1}^*) = \Theta(n^\delta)$, ta có $n^\delta = \Theta\left(m_4^{-\frac{2\alpha}{3}} n^{1-\gamma(1-\frac{2\alpha}{3})}\right)$, dẫn đến kết quả $m_4 = \Theta\left(n^{\gamma-(\gamma+\delta-1)\frac{3}{2\alpha}}\right)$. Như vậy, Định lý đã được chứng minh. \square

Kết quả về phương pháp cơ bản được đưa ra ở trên sẽ được so sánh với kết quả tối ưu tại Mục 2.4. Độ cân bằng thông lượng và độ trễ của phương pháp cơ bản có thể được tìm ra bằng cách sử dụng kết quả của Định lý 2.5.1 vào hàm (2.22a). Từ đó, ta có

$$\begin{aligned} & \sum_{m=1}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}} \\ &= \sum_{m=1}^{m_4-1} \frac{p_m}{\sqrt{A_m^* + B_m^*}} + \sum_{m=m_4}^{m_2-1} \frac{p_m}{\sqrt{A_m^* + B_m^*}} + \sum_{m=m_2}^M \frac{p_m}{\sqrt{A_m^* + B_m^*}} \\ &= \Theta\left(\frac{n^{(\gamma-(\gamma+\delta-1)\frac{1}{\alpha})(\frac{3}{2}-\alpha)-\frac{1}{2}}}{H_\alpha(M)}\right) + \Theta\left(\frac{n^{-\frac{\delta}{2}} \max\{H_\alpha(m_2), H_\alpha(m_4)\}}{H_\alpha(M)}\right) \\ &+ \Theta\left(\frac{n^{\gamma(\frac{3}{2}-\alpha)-\frac{\beta+\delta}{2}}}{H_\alpha(M)}\right). \end{aligned} \quad (2.52)$$

Ký hiệu các thành phần thứ nhất, thứ hai, và thứ ba của phía bên phải phương trình (2.52) là F_4 , F_5 , và F_3 , tương ứng. Theo đó, luôn có $F_5 = \Theta(F_4)$

(a) $\alpha = 0.55$ (b) $\alpha = 1.2$

Hình 2.8: Các tập nghiệm tối ưu $\{A_m^*\}_{m=1}^{200}$ và $\{B_m^*\}_{m=1}^{200}$ tương ứng với sự biến thiên của tham số m .

với $1 < \alpha < \frac{3}{2}$ và $F_5 = O(F_3)$ với $\alpha \leq 1$. Do đó, bằng cách so sánh tương ứng với F_3 và F_4 , nếu $\frac{3(\gamma+\delta-1)}{3(\gamma+\delta-1)-(\gamma-\beta)} \leq \alpha < \frac{3}{2}$, thì $F_4 = \Omega(F_3)$; và $F_4 = o(F_3)$ trong các trường hợp khác. Tiếp theo, để so sánh với phương pháp lưu trữ tối ưu được đưa ra tại Mục 2.4, chúng ta xem xét ba trường hợp sau đây tùy theo giá trị của hệ số α . Với $\alpha < \frac{3(\gamma+\delta-1)}{3(\gamma+\delta-1)-(\gamma-\beta)}$, giá trị của (2.22a) là F_3 , cho thấy rằng, cả hai phương pháp lưu trữ dữ liệu đều đạt được cùng giá trị cân bằng thông lượng và độ trễ của mạng. Với $\frac{3(\gamma+\delta-1)}{3(\gamma+\delta-1)-(\gamma-\beta)} \leq \alpha < 1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)}$

và $1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)} \leq \alpha < \frac{3}{2}$, (2.22a) đạt giá trị F_3 và F_1 , tương ứng theo phương pháp lưu trữ tối ưu đề xuất, trong khi lại chỉ đạt giá trị F_4 tương ứng theo phương pháp lưu trữ tối ưu cơ bản. Do F_4 lớn hơn F_3 và F_1 trong cả hai trường hợp, có thể nói rằng mức cân bằng thông lượng và độ trễ của phương pháp lưu trữ tối ưu đề xuất là tốt hơn trong trường hợp $\frac{3(\gamma+\delta-1)}{3(\gamma+\delta-1)-(\gamma-\beta)} \leq \alpha < \frac{3}{2}$.

Chú ý 2.5.1. *Độ lợi về hiệu năng trong phương pháp lưu trữ tối ưu đề xuất so với phương pháp lưu trữ cơ bản tới từ thực tế là trong mô hình mạng đề xuất, các thiết bị di động người dùng có thể được ưu tiên lưu trữ các tệp dữ liệu có mức độ phổ biến cao, với số lượng các bản sao trong mạng lớn hơn số lượng các trạm gốc thông tin. Nhờ đó, hiệu năng mạng là thông lượng và độ trễ được cải thiện trong trường hợp $A_m^* + B_m^* = \omega(n^\delta)$.*

Chú ý rằng, với $\alpha \geq \frac{3}{2}$ (tương ứng với trường hợp không cần thiết phải sử dụng thêm các bộ nhớ lưu trữ chia sẻ của các trạm gốc thông tin cho kỹ thuật đệm dữ liệu), giá trị của hàm tối ưu (2.22a) là $\frac{1}{\sqrt{n}}$ khi sử dụng phương pháp lưu trữ cơ bản. Đây cũng là trường hợp đưa đến mức cân bằng thông lượng và độ trễ tốt nhất của mạng như kết quả của Định lý 2.3.1.

2.6. So sánh và đánh giá

Để có những đánh giá và góc nhìn đầy đủ hơn về thông lượng và độ trễ tối ưu của mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu đề xuất, trong đó các trạm gốc thông tin được trang bị bộ nhớ lưu trữ chia sẻ có giới hạn, các bảng 2.3 và 2.4 được đưa ra để so sánh các giá trị thông lượng và độ trễ tối ưu của mạng trong các trường hợp mô hình mạng đề xuất $K_{BS} \neq \infty$, mô hình mạng trong đó các trạm gốc thông tin được trang bị bộ nhớ lưu trữ chia sẻ có dung lượng vô hạn $K_{BS} = \infty$ [36], và mô

Bảng 2.3: Bảng so sánh độ trễ mạng tối ưu

α	$K_{BS} \neq \infty$	$K_{BS} = \infty$ [36]	$K_{BS} = 0$ [1, 15, 36, 43, 46]
$\alpha \geq \frac{3}{2}$	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$
$\frac{3}{2} > \alpha \geq 1 + \frac{\gamma - \beta}{2(\gamma + \delta - 1)}$	$\Theta\left(\frac{n^{(1-\delta)(\frac{3}{2}-\alpha)}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{(1-\delta)(\frac{3}{2}-\alpha)}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{(\frac{3}{2}-\alpha)\gamma}}{\sqrt{\log n}}\right)$
$1 + \frac{\gamma - \beta}{2(\gamma + \delta - 1)} > \alpha \geq 1$	$\Theta\left(\frac{n^{\gamma(\frac{3}{2}-\alpha) - \frac{\delta + \beta - 1}{2}}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{(1-\delta)(\frac{3}{2}-\alpha)}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{(\frac{3}{2}-\alpha)\gamma}}{\sqrt{\log n}}\right)$
$1 > \alpha$	$\Theta\left(\frac{n^{\frac{\gamma}{2} - \frac{\delta + \beta - 1}{2}}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{\frac{1-\delta}{2}}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{\frac{\gamma}{2}}}{\log n}\right)$

Bảng 2.4: Bảng so sánh thông lượng mạng tối ưu

α	$K_{BS} \neq \infty$	$K_{BS} = \infty$ [36]	$K_{BS} = 0$ [1, 15, 36, 43, 46]
$\alpha \geq \frac{3}{2}$	$\Theta\left(\frac{1}{\log n}\right)$	$\Theta\left(\frac{1}{\log n}\right)$	$\Theta\left(\frac{1}{\log n}\right)$
$\frac{3}{2} > \alpha \geq 1 + \frac{\gamma - \beta}{2(\gamma + \delta - 1)}$	$\Theta\left(\frac{1}{n^{(1-\delta)(\frac{3}{2}-\alpha)}\sqrt{\log n}}\right)$	$\Theta\left(\frac{1}{n^{(1-\delta)(\frac{3}{2}-\alpha)}\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{(\alpha-\frac{3}{2})\gamma}}{\sqrt{\log n}}\right)$
$1 + \frac{\gamma - \beta}{2(\gamma + \delta - 1)} > \alpha \geq 1$	$\Theta\left(\frac{1}{n^{\gamma(\frac{3}{2}-\alpha) - \frac{\delta + \beta - 1}{2}}\sqrt{\log n}}\right)$	$\Theta\left(\frac{1}{n^{(1-\delta)(\frac{3}{2}-\alpha)}\sqrt{\log n}}\right)$	$\Theta\left(\frac{n^{(\alpha-\frac{3}{2})\gamma}}{\sqrt{\log n}}\right)$
$1 > \alpha$	$\Theta\left(\frac{1}{n^{\frac{\gamma}{2} - \frac{\delta + \beta - 1}{2}}\sqrt{\log n}}\right)$	$\Theta\left(\frac{1}{n^{\frac{1-\delta}{2}}\sqrt{\log n}}\right)$	$\Theta\left(\frac{1}{n^{\frac{\gamma}{2}}\sqrt{\log n}}\right)$

hình mạng không sử dụng kỹ thuật đệm dữ liệu đối với các trạm gốc thông tin $K_{BS} = 0$ [1, 15, 36, 43, 46].

Ở chế độ Regime I (i.e., chế độ Zipf cao), mức tối ưu tốt nhất của độ trễ và thông lượng đạt được nhờ sử dụng phương pháp truyền tin đa chặng Người dùng tới người dùng, và do đó, việc sử dụng thêm bộ nhớ lưu trữ của các trạm gốc thông tin để lưu trữ thêm các tệp dữ liệu là không cần thiết. Lý do bởi vì phần lớn các tệp dữ liệu trong thư viện của mạng là các tệp dữ liệu có tính phổ biến cao.

Mặt khác, ở chế độ Regimes II và III (i.e., các chế độ Zipf thông thường và thấp), việc sử dụng thêm dung lượng lưu trữ $f(n)K_{BS}$ của các trạm gốc thông tin di động chia sẻ trong mạng vô tuyến hỗn hợp hướng nội dung giúp tăng đáng kể hiệu năng mạng so với trường hợp mạng không có sự hiện diện của các trạm gốc thông tin ($K_{BS} = 0$).

Trong đó, ở chế độ Regime II ($\frac{3}{2} > \alpha \geq 1 + \frac{\gamma - \beta}{2(\gamma + \delta - 1)}$), mức tối ưu thông

lượng và độ trễ của mô hình mạng đề xuất đạt được tương đương với trường hợp mô hình mạng vô tuyến hỗn hợp hướng nội dung tĩnh sử dụng các trạm gốc thông tin được trang bị bộ nhớ lưu trữ chia sẻ có dung lượng vô hạn (tương đương với việc kết nối trực tiếp liên tục, không gián đoạn với đường truyền dẫn mạng lõi back-haul chứa tất cả các tệp dữ liệu của mạng) [36].

Như vậy, lợi ích của việc trang bị các bộ nhớ lưu trữ tại trạm gốc thông tin trong trường hợp thông lượng và độ trễ của mạng được tối ưu đã được thể hiện hết sức rõ ràng theo các chế độ hoạt động khác nhau của mạng như đã trình bày.

2.7. Kết luận Chương 2

Xuất phát từ thực tế là các trạm gốc thông tin di động được trang bị các bộ nhớ lưu trữ giới hạn có dung lượng lớn, có thể được áp dụng cùng với kỹ thuật đệm dữ liệu, Chương này đã đưa ra mô hình mạng vô tuyến hỗn hợp hướng nội dung, trong đó mỗi thiết bị di động và trạm gốc thông tin đều được trang bị các bộ nhớ lưu trữ chia sẻ hữu hạn và tìm ra được giải pháp tối ưu hóa được thông lượng và độ trễ mạng. Để tối ưu hóa được các tham số hiệu năng mạng, phương pháp truyền tin phù hợp với mô hình mạng đã được đề xuất, từ đó xây dựng được công thức tính thông lượng và độ trễ, phân tích đưa ra bài toán tối ưu hóa cần thực hiện để tìm ra số lượng tối ưu các bản sao của các tệp dữ liệu trong mạng lưu trữ tại bộ nhớ chia sẻ của các thiết bị di động người dùng và trạm gốc thông tin tương ứng nhờ sử dụng kỹ thuật giải phân tách biến. Các kết quả phân tích đã được xác thực lại bởi các kết quả tính toán bởi máy tính, sử dụng phần mềm phân tích Mathematica. Hiệu năng mạng nhận được cho thấy là tối ưu khi so sánh với phương pháp

lưu trữ dữ liệu cơ bản đề xuất.

Kết quả nghiên cứu tại Chương này cho thấy, ở chế độ Zipf cao $\alpha \geq \frac{3}{2}$, thông lượng và độ trễ tối ưu đạt được nhờ sử dụng phương pháp truyền tin đa chặng Người dùng tới người dùng, và do đó, việc sử dụng thêm bộ nhớ lưu trữ của các trạm gốc thông tin để lưu trữ thêm các tệp dữ liệu là không cần thiết. Lý do bởi vì phần lớn các tệp dữ liệu trong thư viện của mạng là các tệp dữ liệu có tính phổ biến cao và đang được lưu trữ tại phần lớn các thiết bị người dùng. Mặt khác, ở các chế độ Zipf khác, việc sử dụng thêm dung lượng lưu trữ $f(n)K_{BS}$ của các trạm gốc thông tin di động chia sẻ trong mạng vô tuyến hỗn hợp hướng nội dung giúp tăng đáng kể hiệu năng mạng so với trường hợp mạng không có sự hiện diện của các trạm gốc thông tin [1, 15, 36, 43, 46]. Bên cạnh đó, khi hệ số Zipf $\alpha \geq 1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)}$, mức tối ưu thông lượng và độ trễ mạng đạt được tương đương với mức tối ưu đạt được tại trường hợp mô hình mạng vô tuyến hỗn hợp hướng nội dung tĩnh sử dụng các trạm gốc thông tin được trang bị bộ nhớ lưu trữ chia sẻ có dung lượng vô hạn (tương đương với việc kết nối trực tiếp liên tục, không gián đoạn với đường truyền dẫn mạng lõi back-haul chứa tất cả các tệp dữ liệu của mạng) [36].

Để đảm bảo rằng giải pháp tối ưu đồng thời số lượng bản sao của các tệp dữ liệu được lưu trữ tại các thiết bị di động $\{A_m\}_{m=1}^M$ và trạm gốc thông tin $\{B_m\}_{m=1}^M$ trong mô hình mạng đề xuất là giải pháp tối ưu phù hợp, Chương nghiên cứu này cũng đã đưa ra được hiệu năng mạng tối ưu dựa trên giải pháp tối ưu độc lập các tập $\{A_m\}_{m=1}^M$ và $\{B_m\}_{m=1}^M$. Kết quả nhận được thể hiện rằng, hiệu năng mạng tối ưu đạt được theo giải pháp đề xuất ban đầu là tốt nhất nhờ tận dụng được điểm mạnh của việc phân bổ bản sao của các

tệp dữ liệu đồng thời tại thiết bị người dùng và trạm gốc thông tin. Trong đó các thiết bị di động người dùng có thể được ưu tiên lưu trữ các tệp dữ liệu có mức độ phổ biến cao (lớn hơn $f(n)$) và các tệp dữ liệu có mức độ phổ biến thấp hơn sẽ được lưu trong bộ nhớ chia sẻ của các trạm gốc thông tin.

Các đóng góp của luận án trong chương này đã được công bố trong 01 bài báo tại Hội nghị khoa học quốc tế hàng đầu của ngành Lý thuyết thông tin ISIT 2016 [IC1] và 01 bài báo đăng trên tạp chí quốc tế ISI (IEEE Access) [IJ1]. Ngoài ra, một số đóng góp khác của luận án trong chương này cũng được công bố trong các hội nghị khoa học KICS Summer 2015 [IC2], KICS Winter 2016 [IC3], và KICS Summer 2016 [IC4] được tổ chức tại Hàn Quốc.

Trong nghiên cứu ở Chương tiếp theo, luận án sẽ nghiên cứu giải pháp tối ưu hóa thông lượng và độ trễ của mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu với giả thuyết về mô hình mạng nghiên cứu mới, trong đó khác với mô hình mạng dòng chảy được sử dụng tại Chương này, kích thước của các tệp dữ liệu trong mạng có độ lớn đáng kể sẽ được xem xét và nghiên cứu tới. Việc thêm hệ số liên quan nữa là kích thước của tệp dữ liệu và số lượng các mảnh tin được phân tách từ mỗi tệp dữ liệu sẽ khiến cho việc bài toán tối ưu trở nên khó hơn cần giải quyết.

Chương 3

TỐI ƯU HÓA THÔNG LƯỢNG VÀ ĐỘ TRỄ CỦA MẠNG VÔ TUYẾN HƯỚNG NỘI DUNG SỬ DỤNG PHƯƠNG PHÁP PHÂN MẢNH TỆP DỮ LIỆU

Giới thiệu chung: Nghiên cứu của Chương này đề xuất mô hình mạng nghiên cứu là mạng vô tuyến hướng nội dung sử dụng kỹ thuật phân mảnh tệp dữ liệu, trong đó mỗi thiết bị đầu cuối di chuyển theo phương pháp bước ngẫu nhiên và các tệp dữ liệu thư viện mạng có kích thước lớn. Từ đó, giải pháp tối ưu hóa thông lượng và độ trễ mạng sẽ được trình bày. Cụ thể hơn, thay vì sử dụng mô hình dòng chảy tương tự như ở Chương trước và các nghiên cứu [1, 15, 36, 43, 46] khi không xem xét kích thước tệp dữ liệu là tham số mạng quan trọng có ảnh hưởng tới hiệu năng mạng, trong mô hình đề xuất, kích thước tệp dữ liệu được giả sử là rất lớn đến mức không thể truyền đi hoàn toàn trong mỗi khoảng thời gian tương ứng một khe thời gian trong mạng. Giả thiết này phù hợp với thực tế hiện nay khi mà các tệp dữ liệu có thể chứa những nội dung đa phương tiện có dung lượng rất lớn. Khi đó, mỗi tệp dữ liệu này sẽ được chia thành K mảnh tin khác nhau và có kích thước bằng nhau sao cho mỗi mảnh tin có thể được truyền đi hoàn toàn trong một khe thời gian từ thiết bị nguồn tới một trong những thiết bị người dùng khác ở vị trí lân cận. Có hai phương pháp thu nhận các mảnh tin và tái tạo lại tệp dữ liệu gốc sẽ được trình bày trong nội dung của Chương này là phương pháp thu nhận tuần tự và phương pháp thu nhận ngẫu nhiên. Dựa trên hai

phương pháp này, ta sẽ phân tích sự biến thiên theo số lượng nút mạng n của độ trễ và thông lượng của mạng. Từ đó, bài toán tối ưu hóa phương pháp lưu trữ tối ưu tại bộ nhớ chia sẻ của các thiết bị đầu cuối sẽ được xây dựng sao cho hiệu năng mạng đạt được là tốt nhất (độ cân bằng giữa độ trễ và thông lượng là tốt nhất). Kết quả phân tích và giải bài toán tối ưu thu được sẽ được kiểm tra lại bằng các tính toán trên máy tính. Kết quả cho thấy rằng, hiệu năng mạng nhận được nhờ sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên tốt hơn rất nhiều so với việc hiệu năng mạng sử dụng phương pháp tuần tự.

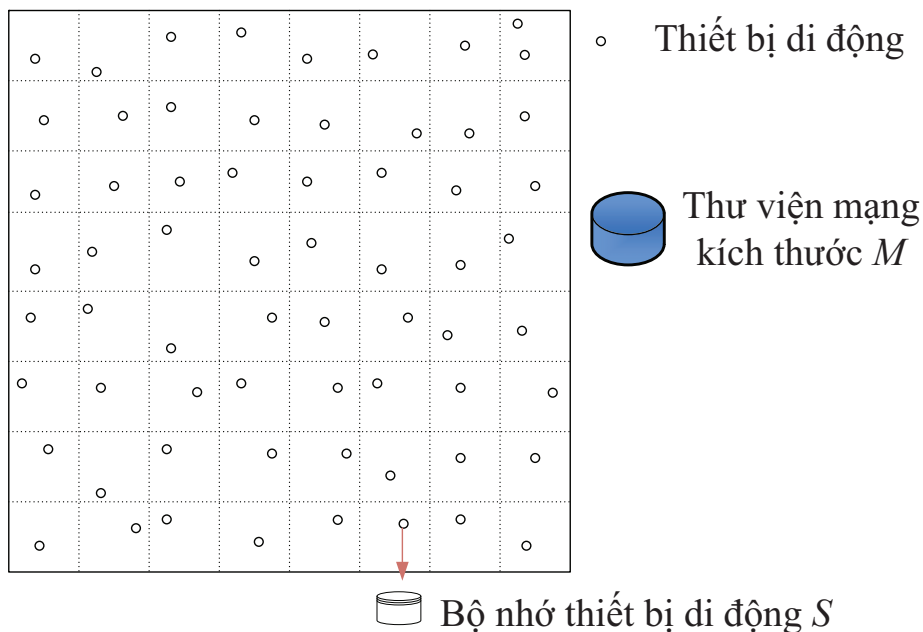
Những đóng góp chính của Chương này có thể được tóm tắt lại như sau:

- Đề xuất mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu, trong đó các thiết bị người dùng di động được coi là các nút mạng có khả năng lưu trữ các tệp dữ liệu trong mạng với dung lượng lưu trữ bị giới hạn. Các nút mạng nguồn cần tải lần lượt K mảnh tin rời rạc của tệp dữ liệu m để tổng hợp lại thành thông tin mong muốn.
- Từ mô hình mạng được đề xuất, mức cân bằng thông lượng và độ trễ của mạng sẽ được phân tích, tìm ra công thức tính toán với hai phương pháp thu nhận mảnh tin đề xuất là tuần tự và ngẫu nhiên.
- Đưa ra giải pháp tối ưu hóa thông lượng và độ trễ thông qua việc tối ưu số lượng bản sao của mỗi tệp dữ liệu lưu trữ và phân bố tại các thiết bị người dùng trong mạng. Các kết quả phân tích và tính toán sẽ được kiểm tra lại bởi các kết quả được giải bởi chương trình toán học được lập trình trên máy tính Mathematica. Kết quả nhận được cho thấy rằng các kết quả phân tích và đánh giá lý thuyết là phù hợp và chính xác.

3.1. Mô hình mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu

Trong phần này sẽ nghiên cứu mạng ad hoc di động hướng nội dung thể hiện ở Hình 3.1, trong đó n thiết bị di động người dùng có thể coi là các nút mạng di động di chuyển theo mô hình bước ngẫu nhiên RWMM trong mô hình mạng đơn vị chuẩn (unit torus), tương tự như Chương 2. Tại thời điểm của mỗi khe thời gian, mỗi nút mạng yêu cầu tải một tệp dữ liệu nằm trong thư viện có kích thước $M = \Theta(n^\gamma)$ của mạng với $0 < \gamma < 1$ một cách độc lập và ngẫu nhiên. Trong Chương này, thay vì giả thiết rằng kích thước của các tệp dữ liệu đủ nhỏ để quá trình truyền tải luôn kết thúc trong một khe thời gian như trường hợp sử dụng mô hình dòng chảy tại Chương 2, chúng tôi giả thiết rằng mỗi tệp dữ liệu được cấu thành bởi $K = \Theta(n^\beta)$, với $0 < \beta < \gamma$, mảnh tin rời rạc có kích cỡ bằng nhau sao cho mỗi mảnh tin có thể được truyền đi hoàn toàn tới nút mạng đích trong khoảng thời gian tương ứng với một khe thời gian. Mỗi nút mạng được trang bị bộ nhớ đệm dữ liệu có khả năng lưu trữ $K_n = \Theta(K)$ mảnh tin khác nhau.

Giả sử rằng xác suất yêu cầu đối với tệp dữ liệu $m \in \{1, \dots, M\}$ tuân theo phân bố Zipf tương tự như Chương 2, được tính theo công thức $p_m = \frac{m^{-\alpha}}{H_\alpha(M)}$, trong đó α là hệ số Zipf và $H_\alpha(M) = \sum_{i=1}^M i^{-\alpha}$. Trong mạng vô tuyến hướng nội dung, kỹ thuật đệm dữ liệu được thực hiện theo hai bước, bước đệm dữ liệu và bước truyền tin. Hai bước này tương ứng với quá trình lập kịch bản lưu trữ thông tin tại bộ nhớ chia sẻ của các nút mạng và quá trình định tuyến đường đi cho các mảnh tin được gửi tới thành công nút mạng đích. Trước hết, chúng ta xem xét bước đệm dữ liệu, bước quyết định mỗi mảnh tin sẽ được



Hình 3.1: Mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu.

lưu tại bộ nhớ đệm của các nút mạng. $X_{m,i}$ là ký hiệu của số lượng bản sao của mảnh tin i thuộc tệp dữ liệu $m \in \mathcal{M}$ với $i = 1, \dots, K$. Theo đó, chúng ta có thể thấy rằng $X_{m,i} = X_m$ với mọi $i = 1, \dots, K$, với X_m là số lượng bản sao của tệp dữ liệu m trên toàn mạng. Với giới hạn tổng bộ nhớ đệm tại các nút mạng là nK_n , chúng ta có biểu thức giới hạn sau:

$$\sum_{m=1}^M KX_m \leq nK_n. \quad (3.1)$$

Tương tự như các nghiên cứu [15, 43, 47], chúng tôi sử dụng phương pháp lưu trữ ngẫu nhiên sao cho các bản sao của mỗi mảnh tin được lưu trữ đều và ngẫu nhiên tại bộ nhớ đệm của các nút mạng. Xin nhắc lại rằng, tại bước truyền tin, vị trí của X_m nút mạng đang lưu giữ mảnh tin quan tâm sẽ bị thay đổi một cách độc lập và ngẫu nhiên theo thời gian do đặc tính của mô hình di động. Ở bước truyền tin, mỗi nút mạng sẽ tải từng mảnh tin của tệp dữ liệu quan tâm từ một trong những nút mạng đang lưu giữ trên toàn mạng

bằng phương pháp truyền tin đa chặng. Chúng tôi sử dụng phương pháp định tuyến được áp dụng ở [38] để định nghĩa về điều kiện đối với sự kiện mảnh tin được truyền nhận thành công. Cụ thể, nếu $d(u, v)$ là ký hiệu của khoảng cách Euclidean giữa nút u và v , thì quá trình truyền tin giữa nút u và v được cho là thành công khi và chỉ khi $d(u, v) \leq r$ và không có bất kỳ một nút mạng nào thực hiện truyền tin trong bán kính $(1 + \Delta)r$ từ nút mạng v , với r và $\Delta > 0$ là các tham số phương pháp định tuyến được thiết lập trước.

Để nhận được tệp dữ liệu mong muốn, nút mạng cần phải tìm kiếm và tải đủ K mảnh tin khác nhau của tệp dữ liệu được phân bố trên toàn mạng. Trong đó, độ trễ truyền tin đối với mỗi mảnh tin được tính là khoảng thời gian từ thời điểm nút mạng gửi bản tin yêu cầu tới nút mạng đích đang lưu trữ mảnh tin cho tới thời điểm mảnh tin đó được truyền thành công tới nút mạng yêu cầu. Dựa vào đó, chúng ta có thể tính được độ trễ trung bình của mạng đối với M tệp dữ liệu khác nhau và n nút mạng trong mạng.

3.2. Phương pháp thu nhận mảnh tin và đề xuất phương pháp định tuyến truyền tin

3.2.1. Phương pháp thu nhận mảnh tin

Trong mục này, chúng tôi sẽ mô tả hai phương pháp thu nhận tệp dữ liệu được mô tả trong Hình 3.2, thể hiện cách thức K mảnh tin của tệp dữ liệu yêu cầu được thu thập.

- *Thu nhận tuần tự*: Tất cả K mảnh tin của tệp dữ liệu sẽ được tải một cách tuần tự bởi nút mạng yêu cầu. Như thể hiện ở Hình 1(a), nút mạng sẽ tìm mảnh tin số 1 gần nhất của tệp dữ liệu m , tiếp theo là mảnh tin số 2 của tệp dữ liệu m được lưu trữ tại bộ nhớ chia sẻ của một nút mạng

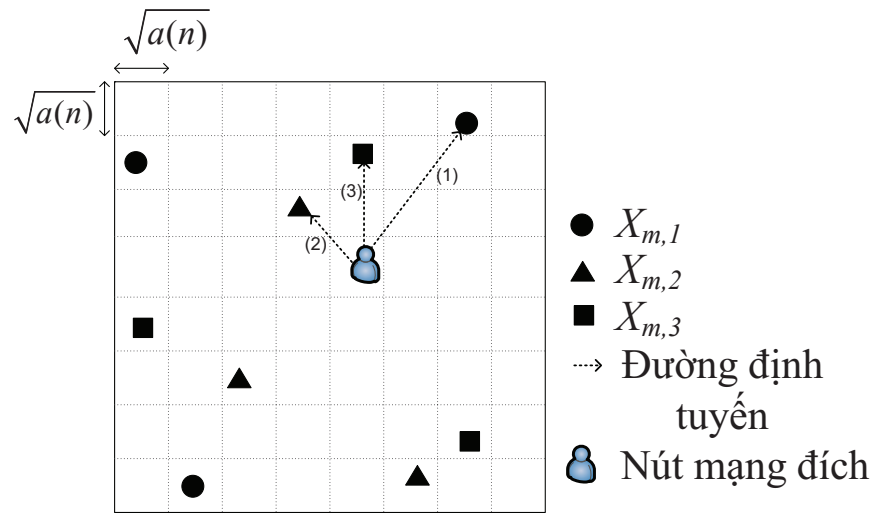
khác ở vị trí gần nhất đối với nút mạng sau khi nút mạng đã nhận được mảnh tin 1, và cứ thế tiếp tục cho đến khi tải đủ K mảnh tin của tệp dữ liệu mong muốn.

- *Thu nhận ngẫu nhiên*: Nút mạng sẽ tải mảnh tin một cách ngẫu nhiên. Như thể hiện trên Hình 1(b), nút mạng sẽ trước hết nhận được mảnh tin số 1 của tệp dữ liệu m , là mảnh tin có vị trí gần nó nhất, tiếp theo đó nút mạng sẽ tải mảnh tin số 2, là mảnh tin gần nó thứ hai và cứ tiếp tục như thế cho đến khi tải đủ K mảnh tin của tệp dữ liệu mong muốn.

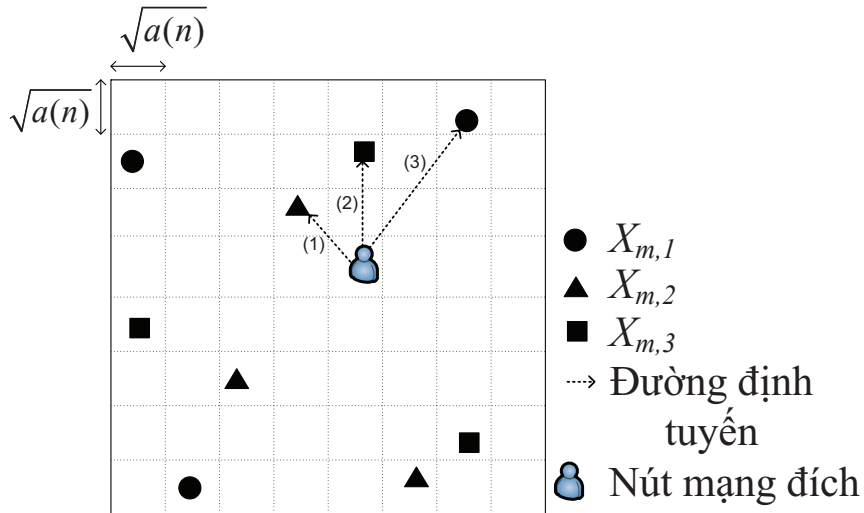
Dựa trên phương pháp định tuyến truyền tin được giới thiệu ở phần nội dung tiếp theo, mỗi phương pháp thu nhận mảnh tin sẽ dẫn đến sự khác nhau về hiệu năng mạng.

3.2.2. Phương pháp định tuyến truyền tin

Trong phần này, phương pháp định tuyến truyền tin được đề xuất sử dụng để truyền tải các mảnh tin tới các nút mạng đích đang yêu cầu tải mảnh tin đó sẽ được mô tả chi tiết. Phương pháp truyền tin đề xuất tương đương với phương pháp định tuyến truyền các tệp dữ liệu theo dạng người dùng tới người dùng ở Chương 2. Cụ thể, thiết bị di động người dùng có yêu cầu tải mảnh tin i đầu tiên sẽ tìm nút mạng gần nhất trong $X_{m,i}$ nút mạng đang lưu trữ tệp dữ liệu mong muốn (theo khoảng cách Euclidean). Sau đó, tin nhắn yêu cầu được tải mảnh tin sẽ được gửi tới nút mạng đó thông qua các ô tế bào lân cận theo phương pháp truyền tin đa chặng, tương đương với bước đầu tiên của quá trình truyền tin. Tương tự, mảnh tin mong muốn sẽ được gửi tới nút mạng đích có yêu cầu tải mảnh tin theo phương pháp truyền tin đa chặng theo chiều ngược lại. Nhớ rằng các nút mạng đích yêu cầu tải thông tin



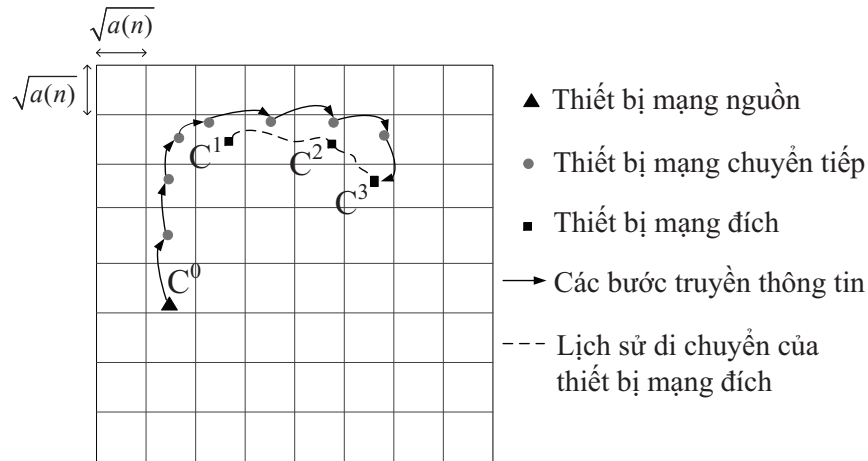
(a) Thu nhận tuần tự



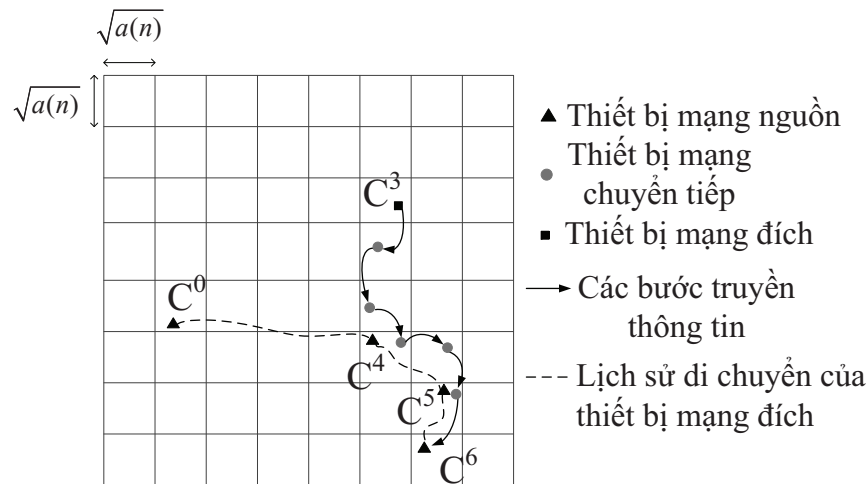
(b) Thu nhận ngẫu nhiên

Hình 3.2: Phương pháp thu nhận mảnh tin của tệp dữ liệu m .

cũng đang di chuyển trong mạng theo mô hình bước đi ngẫu nhiên RWMM. Mỗi khe thời gian sẽ được chia thành hai khe thời gian nhỏ hơn, trong đó bước thứ nhất và thứ hai của thủ tục truyền tin sẽ được thực hiện tương ứng với khoảng thời gian của hai khe thời gian nhỏ này. Trong trường hợp nút mạng yêu cầu tải mảnh tin có vị trí đứng ở trong phạm vi định tuyến truyền tin cơ bản của nút mạng lưu trữ mảnh tin, yêu cầu tải mảnh tin sẽ được đáp ứng ngay nhờ phương pháp truyền tin đơn chặng chỉ trong khoảng thời gian



Hình 3.3: Giai đoạn thứ nhất của quá trình định tuyến truyền tin.



Hình 3.4: Giai đoạn thứ hai của quá trình định tuyến truyền tin.

của một khe thời gian tiêu chuẩn. Chi tiết thủ tục truyền tin được thể hiện ở Hình 3.3 và Hình 3.4 và được mô tả chi tiết như sau:

Step 1) Giai đoạn thứ nhất - Bước yêu cầu tải mảnh tin

Như thể hiện ở Hình vẽ 3.3, xuất phát từ tế bào định tuyến C^0 , bản tin yêu cầu tải tin sẽ được truyền theo phương thức đa chặng tới các tế bào định tuyến lân cận tới tế bào định tuyến C^1 đang chứa thiết bị đích, với khoảng cách mỗi bước nhảy truyền tin được tính bởi $\Theta\left(\sqrt{a(n)}\right)$. Khi tin nhắn yêu cầu tải mảnh tin được truyền tới tế bào định tuyến C^1 , thiết bị

đích chứa tệp dữ liệu mong muốn đã di chuyển tới vị trí khác C^2 do đặc tính di động của thiết bị theo mô hình bước nhảy ngẫu nhiên RWMM. Do đó, tin nhắn yêu cầu tải mảnh tin phải tiếp tục nhảy từ tế bào định tuyến C^1 tới tế bào định tuyến C^2 . Quy trình này sẽ tiếp tục cho tới khi bản tin yêu cầu tải mảnh tin đuổi kịp thiết bị đích tại tế bào định tuyến C^3 .

Step 2) Giai đoạn thứ hai - Bước truyền tin Khi thực thể mạng đích chứa mảnh tin nhận được yêu cầu tải tin, thiết bị di động người dùng yêu cầu tải mảnh tin đã di chuyển tới vị trí khác so với ban đầu C^4 do tính chất di động theo mô hình bước đi ngẫu nhiên RWMM. Như thể hiện tại Hình vẽ 3.4, mảnh tin được yêu cầu tải sẽ được truyền tới thiết bị đích bằng cách di chuyển theo thiết bị đích này theo cách thức tương tự như tin nhắn yêu cầu tải mảnh tin phải đi trong bước truyền tin thứ nhất.

Dựa trên hai phương pháp thu nhận mảnh tin của tệp dữ liệu và phương pháp định tuyến truyền tin đề xuất ở Nội dung 3.2, công thức tính thông lượng và độ trễ của mạng, cũng như công thức cân bằng thông lượng và độ trễ sẽ được tính toán và phân tích ở phần nội dung tiếp theo của Luận án.

3.3. Thông lượng và độ trễ

Từ định nghĩa các tham số hiệu năng mạng tổng quan ở Lời mở đầu luận án, các tham số thông lượng và độ trễ của mô hình mạng đề xuất trong Chương 3 được định nghĩa cụ thể như sau.

Thông lượng: Thông lượng của mạng là giá trị trung bình của dung lượng các mảnh tin mà thiết bị người dùng nhận được trong một khe thời gian.

Độ trễ: Độ trễ của mạng là thời gian trung bình tính từ thời điểm tin nhắn yêu cầu tải mảnh tin đầu tiên của tệp dữ liệu mong muốn gửi đến thiết

bị đích chứa mảnh tin đó cho đến khi thiết bị người dùng yêu cầu tải tin nhận được đầy đủ tất cả các mảnh tin của tệp dữ liệu mong muốn.

Sử dụng các phương pháp thu nhận các mảnh tin để tổng hợp thành tệp dữ liệu mong muốn tải xuống của mỗi nút mạng và phương pháp định tuyến truyền tin được giới thiệu ở Mục trước, trong nội dung của Mục này đưa ra công thức tính thông lượng và độ trễ của mạng tương ứng với mỗi trường hợp sử dụng phương pháp thu nhận mảnh tin khác nhau, từ đó làm cơ sở để xây dựng bài toán tối ưu hóa thông lượng và độ trễ của mạng ở mục tiếp theo. Chi tiết như sau:

3.3.1. Tối ưu hóa thông lượng và độ trễ trường hợp sử dụng phương pháp thu nhận mảnh tin tuần tự

Độ trễ truyền tin luôn phụ thuộc vào khoảng cách giữa nguồn và đích truyền tin. Tương tự chương 2, đối với nút mạng di động yêu cầu tải mảnh tin i của tệp tin $m \in \mathcal{M}$, khoảng cách ban đầu trung bình giữa nút mạng yêu cầu tải tin và nút mạng gần nhất đang lưu trữ mảnh tin mong muốn là $\Theta\left(\frac{1}{\sqrt{X_{m,i}}}\right)$, trong đó $X_{m,i}$ là số lượng bản sao của mảnh tin tải mảnh tin i của tệp tin m được lưu trữ tại bộ nhớ đệm của các nút mạng trong mạng.

Sử dụng phương pháp định tuyến đã được đề xuất và mô tả ở Mục 3.2.2, dễ thấy rằng số lượng bản ghi của mỗi mảnh tin dao động ở mức $a(n)^{-1}$ là đủ để đảm bảo để mảnh tin mong muốn có thể được truyền đến nút mạng đích trong khoảng thời gian hữu hạn $\Theta(1)$. Do đó, với $X_{m,i} = X_m$ khi $i \in \{1, \dots, K\}$ và $m \in \mathcal{M}$, các giới hạn đối với số lượng bản sao của mỗi tệp tin trong mạng

như sau:

$$\begin{cases} X_m \leq a(n)^{-1}, \\ X_m \geq 1. \end{cases} \quad (3.2)$$

Thời gian để nút mạng đích nhận được một mảnh tin bất kỳ của tập tin m thông qua phương pháp định tuyến đa chặng là $\Theta\left(\frac{1}{\sqrt{a(n)X_m}}\right)$. Theo đó, để nhận được toàn bộ K mảnh tin của tập tin m sẽ cần khoảng thời gian là $D(n) = \Theta\left(\frac{K}{\sqrt{a(n)X_m}}\right)$. Trong trường hợp đặc biệt khi $X_m = \Theta(a(n)^{-1})$ với $m \in \mathcal{M}$, $D(n) = \Theta(K)$, là mức trễ truyền tin tốt nhất mà mô hình mạng đề xuất có thể nhận được.

Áp dụng cùng phương pháp phân tích như Mục 2.3, ta thu được Định lý về mức cân bằng thông lượng và độ trễ như sau:

Định lý 3.3.1. *Giả sử mạng vô tuyến hướng nội dung sử dụng phương pháp truyền tin được trình bày tại Chương 3 bằng phương pháp thu nhận tuần tự, mức cân bằng thông lượng và độ trễ của mạng được tính bởi công thức như sau*

$$\lambda(n) = \Theta\left(\frac{D(n)}{Kn \left(\sum_{m=1}^M \frac{p_m}{\sqrt{X_m}}\right)^2}\right) \quad (3.3)$$

trong đó, giới hạn biên trên của $\lambda(n)$ là

$$\lambda(n) = O\left(\frac{1}{\sum_{m=1}^M p_m \sqrt{\frac{n \log n}{X_m}}}\right)$$

và p_m là xác suất yêu cầu tải tin của tập dữ liệu $m \in \mathcal{M}$.

Chứng minh. Trước hết, cần tính toán khoảng cách ban đầu của một cặp nguồn đích truyền tin ngẫu nhiên trong mô hình mạng hướng nội dung được đề xuất. Chiều dài định tuyến của toàn bộ quãng đường cần đi của bản tin

yêu cầu tải dữ liệu được quyết định bởi khoảng cách ban đầu của cặp truyền tin nguồn đích, trong đó khoảng cách trung bình giữa nút mạng đích với mỗi mảnh tin của tập dữ liệu yêu cầu m được cho bởi Bổ đề 2.3.1 như sau $\Theta\left(\frac{1}{\sqrt{X_m}}\right)$. Từ đó, tổng số bước nhảy của quãng đường đi định tuyến của bản tin yêu cầu tải mảnh tin mong muốn biến thiên theo $\Theta\left(\frac{1}{\sqrt{a(n)X_m}}\right)$. Độ trễ mạng $D(n)$ được xác định từ thời điểm bản tin yêu cầu tải dữ liệu rời đi từ thiết bị yêu cầu tải dữ liệu cho tới khi thiết bị này nhận được mảnh tin mong muốn. Thêm vào đó, độ trễ luôn tỷ lệ thuận với chiều dài quãng đường định tuyến và mỗi bước nhảy được tính là mất quãng thời gian tương đương với một khe thời gian đơn vị. Theo phương pháp thu nhận tuần tự, số lượng nút mạng có chứa mảnh tin bất kỳ trong mạng là độc lập với các mảnh tin khác và có giá trị $X_{m,i} = X_m$ không đổi. Do đó, để thu nhận đủ K mảnh tin của tập dữ liệu mong muốn, khoảng thời gian trung bình là

$$D(n) = \Theta\left(\sum_{m=1}^M \frac{Kp_m}{\sqrt{a(n)X_m}}\right), \quad (3.4)$$

trong đó p_m là xác suất thiết bị người dùng bất kỳ yêu cầu tải tập dữ liệu $m \in \mathcal{M}$.

Tương tự như luận giải ở Chương 2, để nhận được một tập dữ liệu bất kỳ tương đương với việc tải đủ K mảnh tin, tổng quãng đường di chuyển trung bình của K mảnh tin là $\Theta\left(\frac{K}{\sqrt{X_m}}\right)$. Từ đó, có thể tính được số lượng mảnh tin trung bình đi qua một tế bào định tuyến bất kỳ ở mỗi khe thời gian được tính bởi công thức $O\left(n \sum_{m=1}^M p_m \sqrt{\frac{a(n)}{X_m}}\right)$. Tại mỗi khe thời gian, các mảnh tin (hoặc bản tin yêu cầu tải tin) đều được truyền qua các tế bào định tuyến. Theo định nghĩa của mô hình truyền tin được mô tả tại Mục 3.1, dễ nhận thấy rằng luôn tồn tại cách thức lập lịch truyền tin sao cho mỗi tế bào định

tuyến sẽ được kích hoạt hoạt động sau mỗi $1 + c$ khe thời gian, tương đương với việc thông lượng dữ liệu truyền qua mỗi tế bào định tuyến là $\Theta(1)$. Tổng dữ liệu truyền qua mỗi tế bào định tuyến phụ thuộc vào số lượng thông tin truyền qua nó ở mỗi khe thời gian. Tương tự như [1], Thông lượng trung bình được tính bởi

$$\lambda(n) = \Theta \left(\frac{1}{n \sum_{m=1}^M p_m \sqrt{\frac{a(n)}{X_m}}} \right), \quad (3.5)$$

thông lượng đạt được tối ưu khi $a(n) = \Theta \left(\frac{\log n}{n} \right)$. Do đó, sử dụng (3.4) và (3.5) sẽ dẫn tới kết quả (3.3). Định lý đã được chứng minh. \square

Mức tối ưu thông lượng và độ trễ của mạng vô tuyến hỗn hợp di động hướng nội dung sử dụng kỹ thuật đệm dữ liệu sẽ được tính toán dựa trên việc lựa chọn tối ưu các tham số về số lượng bản sao $\{X_m\}_{m=1}^M$. Bài toán đề xuất tối ưu hóa mức cân bằng thông lượng và độ trễ mạng tương đương với việc tối đa hóa Thông lượng $\lambda(n)$ và tối thiểu hóa Độ trễ $D(n)$. Do đó, từ các mức giới hạn ở các phương trình (3.1) và (3.2), bài toán tối ưu hóa được xây dựng nên như sau:

$$\max_{\{X_m\}_{m=1}^M} \lambda(n) \quad (3.6a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M X_m \leq n \frac{K_n}{K},$$

$$1 \leq X_m \leq a(n)^{-1} \text{ trong đó } m \in \mathcal{M}.$$

Do tối ưu hóa Thông lượng $\lambda(n)$ và Độ trễ mạng $D(n)$ tương đương với tối thiểu hóa $n \left(\sum_{m=1}^M \frac{p_m}{\sqrt{X_m}} \right)$ ở công thức (3.3) và từ định nghĩa $K_n = \Theta(K)$, bài toán ở (3.6) trở nên không phụ thuộc vào giá trị K và có thể được viết lại như sau:

Bài toán 1:

$$\min_{\{X_m\}_{m=1}^M} \sum_{m=1}^M \frac{p_m}{\sqrt{X_m}} \quad (3.7a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M X_m \leq n, \quad (3.7b)$$

$$1 \leq X_m \leq a(n)^{-1} \text{ trong đó } m \in \mathcal{M}. \quad (3.7c)$$

Ta nhận thấy rằng bài toán (3.7) chính là bài toán được xây dựng và giải tại [1, 15, 36, 43, 46] với nghiệm tối ưu như sau:

$$X_m = \begin{cases} \frac{n}{\log n} & \text{với } m = 1, \dots, m_1 - 1 \\ \frac{p_m^{\frac{2}{3}}}{\sum_{l=m_1}^{m_2} p_l^{\frac{2}{3}}} n & \text{với } m_1, \dots, m_2 - 1 \\ 1 & \text{với } m_2, \dots, M. \end{cases}$$

trong đó,

$$m_1 = \begin{cases} \Theta(\log n) & \text{với } \alpha > \frac{3}{2} \\ \Theta(1) & \text{với } \alpha \leq \frac{3}{2}. \end{cases}$$

và

$$m_2 = \begin{cases} \Theta\left(\min\left(M, n^{\frac{3}{2\alpha}} (\log n)^{1-\frac{3}{2\alpha}}\right)\right) & \text{với } \alpha > \frac{3}{2} \\ M & \text{với } \alpha \leq \frac{3}{2}. \end{cases}$$

Từ đây, thông lượng và độ trễ trong mạng được tính như sau:

- $\alpha \geq \frac{3}{2}$: $D(n) = \Theta(K)$, $\lambda(n) = \Theta\left(\frac{1}{\log n}\right)$;
- $1 \leq \alpha < \frac{3}{2}$: $D(n) = \Theta\left(\frac{KM^{\frac{3}{2}-\alpha}}{\sqrt{\log n}}\right)$, $\lambda(n) = \Theta\left(\frac{M^{\alpha-\frac{3}{2}}}{\sqrt{\log n}}\right)$;
- $\alpha < 1$: $D(n) = \Theta\left(K \frac{\sqrt{M}}{\log n}\right)$, $\lambda(n) = \Theta\left(\frac{1}{\sqrt{M \log n}}\right)$;

3.3.2. Tối ưu hóa thông lượng và độ trễ trường hợp sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên

Trong trường hợp này, mỗi nút mạng yêu cầu tải tin sẽ nhận được K mảnh tin của tệp dữ liệu mong muốn theo thứ tự ngẫu nhiên. Với tệp dữ liệu $m \in \mathcal{M}$, tổng KX_m mảnh tin được phân bố ngẫu nhiên trong mạng. Từ bổ đề 2.3.1, khoảng cách trung bình mảnh tin đầu tiên đến nút mạng đích là $\Theta\left(\frac{1}{\sqrt{KX_m}}\right)$. Do số lượng các mảnh tin còn lại trong mạng thuộc tệp dữ liệu yêu cầu là $(K-1)X_m$, ta có quãng đường trung bình mảnh tin thứ hai phải di chuyển để đến nút mạng đích là $\Theta\left(\frac{1}{\sqrt{(K-1)X_m}}\right)$. Tương tự với các mảnh tin tiếp theo.

Trước hết giả sử rằng $KX_m \geq a(n)^{-1}$. Từ lập luận ở trên, Ký hiệu l_m là số thứ tự của mảnh tin nhỏ nhất thỏa mãn điều kiện $(K-l_m+1)X_m \geq a(n)^{-1}$, với l_m mảnh tin đầu tiên, nút mạng đích luôn nhận được mỗi mảnh tin trong thời gian của mỗi khe thời gian, tương ứng với khoảng cách trung bình $\Theta\left(\sqrt{a(n)}\right)$. Với l_m-1 mảnh tin còn lại, khoảng cách trung bình để nhận được mỗi mảnh tin lần lượt là $\Theta\left(\frac{1}{\sqrt{(K-l_m-1)X_m}}\right)$, $\Theta\left(\frac{1}{\sqrt{(K-l_m-2)X_m}}\right)$, \dots , $\Theta\left(\frac{1}{\sqrt{X_m}}\right)$. Từ đó, tổng quãng đường trung bình để thu nhận đủ K mảnh tin của tệp dữ liệu m được cho bởi $\Theta\left(l_m\sqrt{a(n)}\right) + \sum_{j=l_m-1}^{K-1} \Theta\left(\frac{1}{\sqrt{(K-j)X_m}}\right)$. Từ đặc tính của hàm Riemann Zeta, ta có kết quả $\sum_{j=l_m-1}^{K-1} \Theta\left(\frac{1}{\sqrt{(K-j)X_m}}\right) = \Theta\left(\frac{\sqrt{K-l_m+1}}{\sqrt{X_m}}\right)$. Từ định nghĩa của hệ số l_m , ta có $(K-l_m+1)X_m = a(n)^{-1}$. Theo đó, $\Theta\left(l_m\sqrt{a(n)}\right) + \Theta\left((K-l_m+1)\sqrt{a(n)}\right) = \Theta\left(K\sqrt{a(n)}\right)$ là tổng quãng đường trung bình, tương ứng với tổng khoảng cách trung bình nhỏ nhất để thu nhận đủ K mảnh tin của tệp dữ liệu m . Từ đây, ta nhận thấy rằng, với trường hợp sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên, số lượng

bản sao của mỗi mảnh tin của tệp dữ liệu trong mạng không nhất thiết phải lớn hơn $\frac{a(n)^{-1}}{K}$. Ta có giới hạn về số lượng bản sao của tệp dữ liệu $m \in \mathcal{M}$ được cho bởi

$$\begin{cases} X_m \leq \frac{a(n)^{-1}}{K}, \\ X_m \geq 1. \end{cases} \quad (3.8)$$

Từ những phân tích ở trên, ta có Định lý sau:

Định lý 3.3.2. *Giả sử mạng vô tuyến hướng nội dung sử dụng phương pháp truyền tin được đề xuất tại Chương 3 sử dụng phương pháp thu nhận ngẫu nhiên, mức cân bằng thông lượng và độ trễ của mạng được tính bởi công thức như sau*

$$\lambda(n) = \Theta \left(\frac{D(n)}{n \left(\sum_{m=1}^M \frac{p_m}{\sqrt{X_m}} \right)^2} \right) \quad (3.9)$$

trong đó, giới hạn biên trên của $\lambda(n)$ là

$$\lambda(n) = O \left(\frac{1}{\sum_{m=1}^M p_m \sqrt{\frac{n \log n}{K X_m}}} \right)$$

và p_m là xác suất yêu cầu tải tin của tệp dữ liệu $m \in \mathcal{M}$.

Chứng minh. Với $X_m \leq \frac{a(n)^{-1}}{K}$, tổng khoảng cách trung bình của tất cả các mảnh tin di chuyển để đến được nút mạng đích được tính bởi công thức như sau:

$$\sum_{i=1}^{K-1} \Theta \left(\frac{1}{\sqrt{(K-i)X_m}} \right) = \Theta \left(\frac{1}{\sqrt{X_m}} \right) \sum_{i=1}^{K-1} \frac{1}{\sqrt{K-i}} \quad (3.10)$$

$$= \Theta \left(\frac{\sqrt{K}}{\sqrt{X_m}} \right). \quad (3.11)$$

Độ trễ được tính bởi

$$D(n) = \Theta \left(\sum_{m=1}^M \frac{\sqrt{K} p_m}{\sqrt{a(n) X_m}} \right), \quad (3.12)$$

Với tổng khoảng quãng đường di chuyển $\Theta\left(\frac{\sqrt{K}}{\sqrt{X_m}}\right)$, ta thu được tập dữ liệu có kích thước là K . Sử dụng phương pháp chứng minh tương tự như Định lý , ta nhận được Thông lượng trung bình được tính bởi

$$\lambda(n) = \Theta\left(\frac{1}{n \sum_{m=1}^M p_m \sqrt{\frac{a(n)}{KX_m}}}\right), \quad (3.13)$$

thông lượng đạt được tối ưu khi $a(n) = \Theta\left(\frac{\log n}{n}\right)$. Do đó, sử dụng (3.15) và (3.16) sẽ dẫn tới kết quả (3.9). Định lý đã được chứng minh. \square

Từ Định lý 3.3.2, có thể thấy rằng tối ưu hóa mức cân bằng thông lượng và độ trễ mạng tương đương với việc tối đa hóa Thông lượng $\lambda(n)$ và tối thiểu hóa Độ trễ $D(n)$. Do đó, từ các mức giới hạn tại các phương trình (3.1) và (3.8), bài toán tối ưu hóa được xây dựng nên như sau:

$$\max_{\{X_m\}_{m=1}^M} \lambda(n) \quad (3.14a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M X_m \leq n \frac{K_n}{K}, \quad (3.14b)$$

$$1 \leq X_m \leq \frac{a(n)^{-1}}{K} \text{ trong đó } m \in \mathcal{M}. \quad (3.14c)$$

Do tối ưu hóa Thông lượng $\lambda(n)$ và Độ trễ mạng $D(n)$ tương đương với tối thiểu hóa $n \left(\sum_{m=1}^M \frac{p_m}{\sqrt{X_m}}\right)$ ở công thức (3.9), bài toán ở (3.14) có thể được viết lại như sau:

Bài toán 2:

$$\min_{\{X_m\}_{m=1}^M} \sum_{m=1}^M \frac{p_m}{\sqrt{X_m}} \quad (3.15a)$$

$$\text{với các điều kiện: } \sum_{m=1}^M KX_m \leq nK_n, \quad (3.15b)$$

$$1 \leq X_m \leq \frac{a(n)^{-1}}{K} \text{ trong đó } m \in \mathcal{M}. \quad (3.15c)$$

Tham khảo [42], tương tự như những công trình nghiên cứu trước đây [1, 15, 36, 43, 46], khả vi cấp hai của hàm tối ưu có giá trị dương, ta có hàm tối ưu là hàm lồi. Từ đó, phương pháp Lagrange có thể được sử dụng để giải bài toán ở (3.15). Trước hết, nghiệm tối ưu X_m^* sẽ được chứng minh là không tăng khi $m \in \mathcal{M}$. Hàm Lagrange tương ứng với bài toán trên được tính như sau:

$$\begin{aligned} & \mathcal{L}(\{X_m\}_{m \in \mathcal{M}}, \lambda, \{\mu_m\}_{m \in \mathcal{M}}) \\ &= \sum_{m \in \mathcal{M}} \frac{p_m}{\sqrt{X_m}} + \lambda \left(\sum_{m=1}^M X_m - n \frac{K_n}{K} \right) + \sum_{m \in \mathcal{M}} \mu_m \left(X_m - \frac{a(n)^{-1}}{K} \right), \end{aligned} \quad (3.16)$$

trong đó $\mu_m, \lambda \in \mathbb{R}$. Các điều kiện Karush–Kuhn–Tucker (KKT) đối với (3.15) được tính như sau

$$\frac{\partial \mathcal{L}(\{X_m^*\}_{m \in \mathcal{M}}, \lambda^*, \{\mu_m^*\}_{m \in \mathcal{M}})}{\partial X_m^*} = 0 \quad (3.17)$$

$$\lambda^* \geq 0 \quad (3.18)$$

$$\mu_m^* \geq 0 \quad (3.19)$$

$$\mu_m^* \left(X_m^* - \frac{a(n)^{-1}}{K} \right) = 0 \quad (3.20)$$

$$\lambda^* \left(\sum_{m=1}^M X_m^* - n \frac{K_n}{K} \right) = 0 \quad (3.21)$$

với $\forall m \in \mathcal{M}$. Từ (3.17), ta có

$$-\frac{p_m}{2\sqrt{X_m^{*3}}} + \lambda^* + \mu_m^* = 0 \text{ trong đó } m \in \mathcal{M}. \quad (3.22)$$

Với $M = \omega(K_n \log n)$, luôn tồn tại ít nhất 01 tập dữ liệu $m \in \mathcal{M}$ sao cho nghiệm $X_m^* < \frac{a(n)^{-1}}{K}$. Theo đó, ta nhận được $\mu_m^* = 0$ từ (3.20), dẫn đến giá trị $\lambda^* > 0$ ở (3.22). Sử dụng (3.21) ta nhận được $\sum_{m=1}^M X_m^* = n \frac{K_n}{K}$.

Giả sử D là tập hợp các tập dữ liệu thỏa mãn điều kiện $X_m^* = \frac{a(n)^{-1}}{K}$. Trước hết cần xem xét tập hợp $m \in \mathcal{M} \setminus D$. Sử dụng (3.20) và (3.22), ta có

được:

$$X_m^* = \frac{p_m^{\frac{2}{3}}}{(2\lambda^*)^{\frac{2}{3}}}. \quad (3.23)$$

Như vậy, X_m^* tỷ lệ thuận với p_m và giảm dần đều khi $m \in \mathcal{M} \setminus D$ tăng lên.

Xem xét tập dữ liệu $k \in D$ và $m_0 \in \mathcal{M} \setminus D$ là các giá trị nhỏ nhất thỏa mãn điều kiện $X_{m_0}^* = \frac{a(n)^{-1}}{K}$. Sử dụng các phương trình tại (3.18), (3.19), và (3.22), ta có $k < m_0$. Từ đây, ta có nghiệm tối ưu X_m^* không phải là tập giá trị tăng khi $m \in \mathcal{M}$ tăng dần. Sử dụng (3.23) với mọi $m \in \mathcal{M} \setminus D$, ta có phương trình sau:

$$X_m^* = \frac{p_m^{\frac{2}{3}}}{\sum_{l=m_0}^M p_l^{\frac{2}{3}}} \sum_{l=m_0}^M X_l^*. \quad (3.24)$$

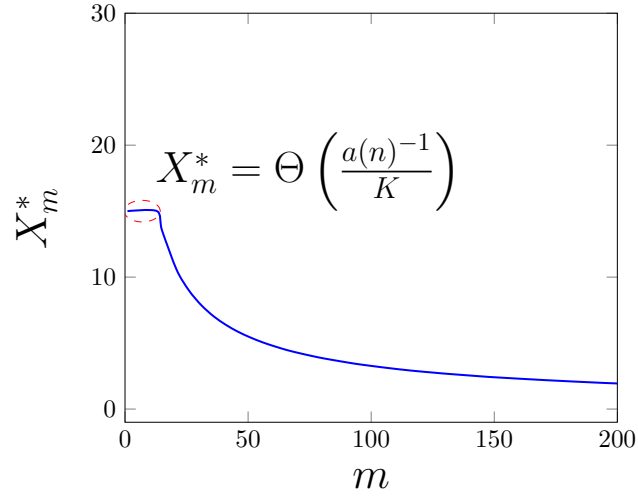
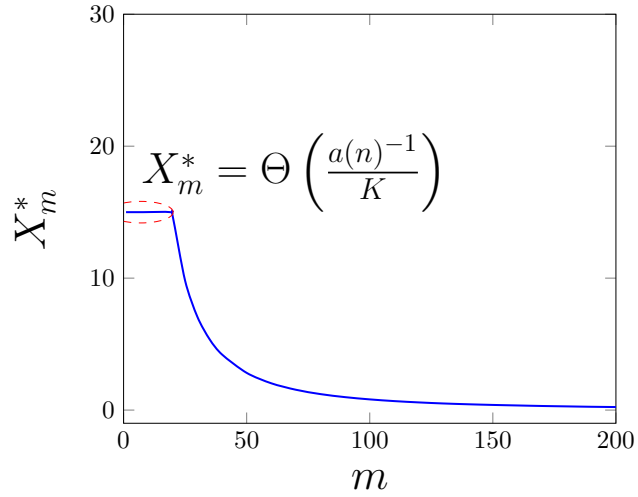
Nếu $X_M^* = \frac{a(n)^{-1}}{K}$, ta có $\sum_{l=1}^M X_l^* = \frac{Ma(n)^{-1}}{K}$, điều này trái với $\sum_{l=1}^M X_m^* = \frac{nK_n}{K}$ như đã chỉ ra ở trên. Do đó, ta luôn có $X_M^* < \frac{a(n)^{-1}}{K}$. Nghiệm tối ưu tìm được thể hiện ở Hình 3.5 theo các giá trị của hệ số Zipf $\alpha = 0.5$ và $\alpha = 1.2$ tương ứng.

Xuất phát từ định nghĩa của m_0 , là giá trị nhỏ nhất thỏa mãn $X_{m_0}^* = \frac{a(n)^{-1}}{K}$, ta có $M = \omega(m_0)$. Từ (3.24) có phương trình sau:

$$\sum_{m=1}^M \frac{p_m}{\sqrt{X_m^*}} = \sum_{m=1}^{m_0} \frac{p_m}{\sqrt{\frac{a(n)^{-1}}{K}}} + \frac{\left(\sum_{m=m_0}^M p_m^{\frac{2}{3}}\right)^{\frac{3}{2}}}{\sqrt{\sum_{m=m_0}^M X_m^*}}. \quad (3.25)$$

Tùy vào giá trị của tham số α , có thể xem xét các trường hợp như sau:

- $\alpha \geq \frac{3}{2}$

(a) $\alpha = 0.5$ (b) $\alpha = 1.2$ Hình 3.5: Nghiệm tối ưu số lượng bản sao lưu trữ của các tệp dữ liệu theo tham số m .

Áp dụng $X_{m_0}^* = \frac{a(n)^{-1}}{K}$ ở (3.24), ta có:

$$\Theta\left(\frac{a(n)^{-1}}{K}\right) = \frac{p_m^{\frac{2}{3}}}{\sum_{l=m_0}^M p_l^{\frac{2}{3}}} \sum_{l=m_0}^M X_l^* \quad (3.26)$$

$$= \Theta\left(\frac{m_0^{-\frac{2\alpha}{3}}}{m_0^{1-\frac{2\alpha}{3}}} \sum_{l=m_0}^M X_l^*\right) \quad (3.27)$$

$$= \Theta\left(\frac{\sum_{l=m_0}^M X_l^*}{m_0}\right) \quad (3.28)$$

Từ đây, phần tử thứ hai ở (3.25) có thể được tính toán như sau

$$\frac{\left(\sum_{m=m_0}^M p_m^{\frac{2}{3}}\right)^{\frac{3}{2}}}{\sqrt{\sum_{m=m_0}^M X_m^*}} = \Theta\left(\frac{m_0^{\frac{3}{2}-\alpha}}{\sqrt{\frac{a(n)^{-1}}{K}m_0}}\right) = \Theta\left(\frac{m_0^{1-\alpha}}{\sqrt{\frac{a(n)^{-1}}{K}}}\right).$$

Do $m_0 \geq 1$, luôn có giá trị của $\sum_{m=1}^M \frac{p_m}{\sqrt{X_m^*}}$ biến thiên theo giá trị vượt trội của phần tử thứ nhất ở (3.25) là

$$\sum_{m=1}^{m_0} \frac{p_m}{\sqrt{X_m^*}} = \Theta\left(\frac{1}{\sqrt{X_{m_0}^*}}\right) = \Theta\left(\sqrt{\frac{K}{a(n)^{-1}}}\right).$$

Do vậy, sử dụng (3.12) và (3.13), các giá trị đạt được tối ưu của thông lượng và độ trễ lần lượt là:

$$\lambda(n) = \Theta\left(\frac{1}{\log n}\right) \text{ và } D(n) = \Theta(K) \text{ với } a(n) = \Theta\left(\frac{\log n}{n}\right).$$

- $\frac{3}{2} > \alpha > 1$

Từ (3.25) ta có:

$$\sum_{m=1}^M \frac{p_m}{\sqrt{X_m^*}} = \Theta\left(\sqrt{\frac{K}{a(n)^{-1}}}\right) + \frac{M^{\frac{3}{2}-\alpha}}{\sqrt{\sum_{m=m_0}^M X_m^*}}.$$

Tùy theo sự biến thiên của hệ số Zipf α và các tham số mạng γ và β có thể chia thành các trường hợp như sau:

- Trường hợp $\frac{3}{2} > \alpha \geq \frac{3}{2} - \frac{\beta}{2\gamma}$:

Ta luôn có được giá trị tối ưu $\sum_{m=1}^M \frac{p_m}{\sqrt{X_m^*}} = \Theta\left(\sqrt{\frac{K}{a(n)^{-1}}}\right)$ khi giá trị của phần tử thứ nhất ở (3.25) vượt trội giá trị của phần tử thứ hai, là giá trị tốt nhất có thể đạt được. Theo đó, sử dụng (3.12) và (3.13) ta cũng nhận được các giá trị đạt được tối ưu của Thông lượng và độ trễ lần lượt là:

$$\lambda(n) = \Theta\left(\frac{1}{\log n}\right) \text{ và } D(n) = \Theta(K) \text{ với } a(n) = \Theta\left(\frac{\log n}{n}\right).$$

– Trường hợp $\frac{3}{2} - \frac{\beta}{2\gamma} > \alpha \geq 1$:

Phần tử thứ hai của (3.25) sẽ có giá trị lớn chi phối, sử dụng phương pháp tính toán tương tự đối với các công thức áp dụng các đặc tính của hàm Riemann zeta, ta có biến thiên giá trị của hàm tối ưu đạt được là $\sum_{m=1}^M \frac{p_m}{\sqrt{X_m^*}} = \Theta\left(\frac{M^{\frac{3}{2}-\alpha}}{\sqrt{n}}\right)$ với $a(n) = \Theta\left(\frac{\log n}{n}\right)$. Từ đó, sử dụng (3.12) và (3.13) ta có các giá trị đạt được tối ưu của thông lượng và độ trễ lần lượt là:

$$\lambda(n) = \Theta\left(\frac{M^{\alpha-\frac{3}{2}}}{\sqrt{\log n}}\right) \text{ và } D(n) = \Theta\left(\frac{\sqrt{KM}^{\frac{3}{2}-\alpha}}{\sqrt{\log n}}\right).$$

• $\alpha < 1$

Tương tự, ta có $\sum_{m=1}^M \frac{p_m}{\sqrt{X_m^*}} = \Theta\left(\frac{\sqrt{M}}{\sqrt{n}}\right)$ với $a(n) = \Theta\left(\frac{\log n}{n}\right)$. Từ đó sử dụng (3.12) và (3.13), các giá trị đạt được tối ưu của thông lượng và độ trễ lần lượt là:

$$\lambda(n) = \Theta\left(\frac{1}{\sqrt{M \log n}}\right) \text{ và } D(n) = \Theta\left(\frac{\sqrt{KM}}{\sqrt{\log n}}\right)$$

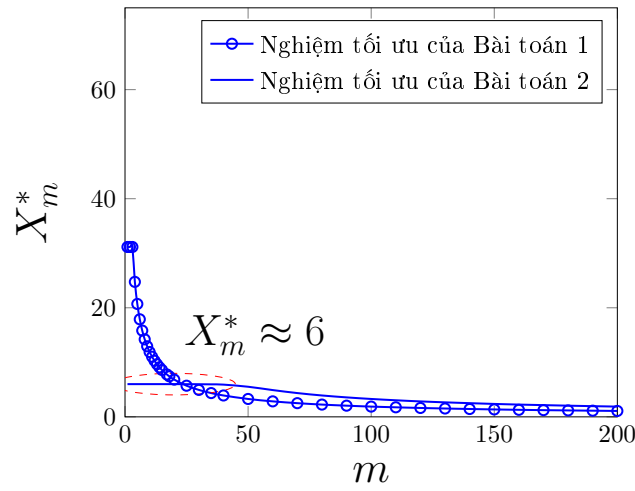
3.3.3. Nghiệm tối ưu hóa sử dụng phần mềm tính toán trên máy tính

Trong phần này, để kiểm tra lại tính đúng đắn của các kết quả phân tích thể hiện ở Nội dung trước cần thực hiện tính toán bởi phần mềm giải toán Mathematica trên máy tính nhiều lần để tìm ra nghiệm tối ưu của Bài toán 1 và Bài toán 2 tương ứng theo các tham số hệ thống được cho trước n , M , K_n , K , $a(n)^{-1}$, và α ở Bảng 3.1 tương ứng.

Có thể nhận thấy rằng, kết quả tính toán bởi phần mềm máy tính giải toán Mathematica là phù hợp với các kết quả phân tích được đưa ra tại Nội dung 3.3.1 và 3.3.2, như thể hiện ở Hình 3.6. Cụ thể, nghiệm tối ưu $\{X_m^*\}_{m=1}^M$ giảm đều khi hệ số m tăng lên đối với bài toán thứ nhất, sử dụng phương pháp

Bảng 3.1: Bảng các tham số mạng được sử dụng để tính toán trên máy tính

Ký hiệu	Mô tả	Giá trị thiết lập
n	Số lượng thiết bị di động người dùng	300
M	Số lượng tệp dữ liệu trong thư viện mạng	200
K	Số lượng các mảnh tin của mỗi tệp dữ liệu	20
$a(n)^{-1}$	Số lượng các tế bào định tuyến trong toàn mạng	120
α	Hệ số Zipf	1.2
K_n	Độ lớn lưu trữ của thiết bị di động người dùng	50



Hình 3.6: Tập nghiệm tối ưu $\{X_m^*\}_{m=1}^{200}$ của các bài toán sử dụng phương pháp phân mảnh tệp dữ liệu tương ứng với tham số m .

thu nhận mảnh tin tuần tự; Đối với bài toán thứ hai, sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên, giá trị nghiệm tối ưu $\{X_m^*\}_{m=1}^M$ là không đổi khi hệ số m nhỏ và bắt đầu giảm dần khi hệ số m đạt một giá trị ngưỡng xác định. Sự khác nhau này xuất phát từ sự khác nhau của các giá trị ngưỡng được luận giải và đưa ra tại hai bài toán tương ứng với hai trường hợp thu nhận mảnh tin.

Từ kết quả phân tích và kết quả tính toán máy tính, hiệu năng mạng sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên luôn tốt hơn khi sử dụng phương pháp thu nhận mảnh tin tuần tự. Điều này có được là nhờ sự linh động trong việc thu nhận K mảnh tin của tệp dữ liệu mong muốn theo

phương pháp ngẫu nhiên, trong đó không xét đến yếu tố thứ tự hay ưu tiên của các mảnh tin, nhờ đó dẫn đến hiệu năng mạng tốt hơn.

3.4. So sánh và đánh giá

Dựa trên kết quả tối ưu hóa Độ trễ và Thông lượng của mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu, áp dụng hai phương pháp phân mảnh tệp dữ liệu là thu nhận mảnh tin tuần tự ở Mục 3.3.1 và thu nhận mảnh tin ngẫu nhiên ở Mục 3.3.2, tùy theo sự biến thiên của hệ số Zipf α và các tham số mạng γ và β , ta có các so sánh như sau:

- $\alpha \geq \frac{3}{2}$

Cả hai phương pháp thu nhận mảnh tin đều đạt được mức tối ưu về thông lượng và độ trễ tương đương nhau. Điều này là do với trường hợp hệ số Zipf α có giá trị cao, phần lớn các tệp dữ liệu được lưu trong thư viện mạng đều là các tệp dữ liệu có tính phổ biến cao và được lưu trữ rộng rãi trong mạng, nhờ đó mà các nút mạng dễ dàng tìm được các mảnh tin của tệp dữ liệu mong muốn ở ngay các nút mạng lân cận để có thể tải trong thời gian ngắn nhất tương đương với mỗi khe thời gian.

- $\frac{3}{2} > \alpha \geq \frac{3}{2} - \frac{\beta}{2\gamma}$:

Phương pháp thu nhận mảnh tin ngẫu nhiên vẫn giúp cho mạng đạt được hiệu năng tương đương trường hợp hệ số Zipf $\alpha \geq \frac{3}{2}$. Do đó, hiệu năng mang lại là vượt trội so với hiệu năng mạng sử dụng phương pháp thu nhận mảnh tin tuần tự. Điều này là bởi vì sự linh động của phương pháp ngẫu nhiên trong việc thu nhận K mảnh tin của tệp dữ liệu mong muốn, trong đó không xét đến yếu tố thứ tự hay ưu tiên của các mảnh tin, nhờ

đó dẫn đến sự hiệu năng mạng tốt hơn.

- $\frac{3}{2} - \frac{\beta}{2\gamma} > \alpha$:

Hai phương pháp có mức thông lượng tối ưu là tương đương nhau, tuy nhiên mức tối ưu về độ trễ của phương pháp thu nhận ngẫu nhiên tốt hơn \sqrt{K} lần so với độ trễ tối ưu thu được trong trường hợp sử dụng phương pháp thu nhận mảnh tin tuần tự.

Chú ý 3.4.1. *Như vậy, ngoại trừ trường hợp đặc biệt khi phần lớn các tệp dữ liệu được lưu trong thư viện mạng đều là các tệp dữ liệu có tính phổ biến cao khi hệ số Zipf $\alpha \geq \frac{3}{2}$, hai phương pháp thu nhận tuần tự và ngẫu nhiên đều đạt được mức tối ưu về thông lượng và độ trễ tương đương nhau, phương pháp thu nhận ngẫu nhiên luôn cho thấy sự ưu việt về hiệu năng mạng so với phương pháp thu nhận tuần tự. Điều này cho thấy rằng việc yêu cầu các nút mạng phải thu thập các mảnh tin theo thứ tự quy định trước để tái tạo lại tệp dữ liệu mong muốn là hạn chế của phương pháp thu nhận mảnh tin tuần tự. Tính linh động trong việc thu thập các mảnh tin của phương pháp thu nhận ngẫu nhiên giúp cho các nút mạng có nhiều hơn sự lựa chọn cho mỗi lần tìm kiếm và thu thập mảnh tin, nhờ đó giảm thiểu đi thời gian thu thập các mảnh tin.*

Để có cái nhìn đầy đủ hơn về ảnh hưởng của kích thước tệp dữ liệu đối với các tham số tối ưu thông lượng và độ trễ mạng. Ta so sánh giá trị tối ưu hiệu năng mạng mạng khi sử dụng phương pháp phân mảnh tệp tin ngẫu nhiên khi kích thước tệp dữ liệu có giá trị đáng kể $K = \omega(1)$ với trường hợp không sử dụng phương pháp phân mảnh tệp dữ liệu (tương đương với $K = \Theta(1)$ như trong mô hình dòng chảy và các nghiên cứu trước đó [1, 15, 36, 43, 46].

Bảng 3.2: Bảng giá trị thông lượng mạng tối ưu

α	$K = \omega(1)$	$K = \Theta(1)$ [1, 15, 36, 43, 46]
$\alpha \geq \frac{3}{2}$	$\Theta\left(\frac{1}{\log n}\right)$	$\Theta\left(\frac{1}{\log n}\right)$
$\frac{3}{2} > \alpha \geq 1$	$\Theta\left(\frac{M^{\alpha - \frac{3}{2}}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{M^{\alpha - \frac{3}{2}}}{\sqrt{\log n}}\right)$
$\alpha < 1$	$\Theta\left(\frac{1}{\sqrt{M \log n}}\right)$	$\Theta\left(\frac{1}{\sqrt{M \log n}}\right)$

Bảng 3.3: Bảng giá trị độ trễ mạng tối ưu

α	$K = \omega(1)$	$K = \Theta(1)$ [1, 15, 36, 43, 46]
$\alpha \geq \frac{3}{2}$	$\Theta(K)$	$\Theta(1)$
$\frac{3}{2} > \alpha \geq 1$	$\Theta\left(\frac{\sqrt{KM}^{\frac{3}{2} - \alpha}}{\sqrt{\log n}}\right)$	$\Theta\left(\frac{M^{\frac{3}{2} - \alpha}}{\sqrt{\log n}}\right)$
$\alpha < 1$	$\Theta\left(\frac{\sqrt{KM}}{\log n}\right)$	$\Theta\left(\frac{\sqrt{M}}{\log n}\right)$

Từ bảng 3.2, ta thấy rằng khi sử dụng phương pháp phân mảnh tệp dữ liệu, kích thước tệp dữ liệu không ảnh hưởng tới thông lượng tối ưu của mạng.

Tuy nhiên, khi tệp dữ liệu có kích thước càng lớn, độ trễ mạng càng tăng (Bảng 3.3). Cụ thể, trong trường hợp $\alpha \geq \frac{3}{2}$, khi phần lớn các tệp dữ liệu trong mạng đều có tính chất phổ biến cao và được lưu tại hầu hết các thiết bị người dùng, thời gian để nhận được mỗi tệp dữ liệu mong muốn tỷ lệ thuận K lần với kích thước tệp dữ liệu. Trong các trường hợp thông thường khác, để nhận được mỗi tệp dữ liệu mong muốn, thời gian cần thiết là gấp \sqrt{K} lần so với trường hợp không sử dụng phương pháp phân mảnh tệp dữ liệu (tương đương với mô hình dòng chảy, khi mỗi tệp dữ liệu có kích thước đủ nhỏ để có thể được truyền đi hoàn toàn trong khoảng thời gian là mỗi khe thời gian giữa các nút mạng lân cận với nhau).

3.5. Kết luận Chương 3

Ở Chương này, xuất phát từ thực tế là các tệp dữ liệu hiện nay trên mạng có thể là các tệp dữ liệu đa phương tiện (như video 4K, ảnh chất lượng cao, ảnh 3D,...) có dung lượng vô cùng lớn, luận án đã thực hiện nghiên cứu trên mô hình mạng vô tuyến hướng nội dung sử dụng kỹ thuật đệm dữ liệu, trong đó xem xét ảnh hưởng của tham số kích thước tệp dữ liệu K . Trong đó, các phương pháp phân mảnh tệp tin tuần tự và ngẫu nhiên được giới thiệu và các giá trị tối ưu tương ứng của thông lượng và độ trễ của mạng đã được tính toán dựa trên việc phân tích toán học, giải bài toán tối ưu hóa số lượng các bản sao lưu trữ $\{X_m\}_{m=1}^M$ của mỗi tệp dữ liệu $m \in \mathcal{M}$ được lưu trữ phân mảnh tại các bộ nhớ chia sẻ của các nút mạng. Kết quả nhận được cho thấy rằng sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên đạt được hiệu năng mạng tối ưu tốt hơn hơn phương pháp thu nhận mảnh tin tuần tự.

Ngoài ra, kết quả nghiên cứu tại Chương 3 cho thấy rằng, khi sử dụng phương pháp phân mảnh tệp dữ liệu, kích thước tệp dữ liệu không ảnh hưởng tới thông lượng tối ưu của mạng. Đối với tham số độ trễ tối ưu của mạng, khi phần lớn các tệp dữ liệu trong mạng đều có tính chất phổ biến cao và được lưu tại hầu hết các thiết bị người dùng ($\alpha \geq \frac{3}{2}$), thời gian để nhận được mỗi tệp dữ liệu mong muốn tỷ lệ thuận K lần với kích thước tệp dữ liệu. Trong các trường hợp thông thường khác, để nhận được mỗi tệp dữ liệu mong muốn, thời gian cần thiết là gấp \sqrt{K} lần so với trường hợp không sử dụng phương pháp phân mảnh tệp dữ liệu (tương đương với mô hình dòng chảy, khi mỗi tệp dữ liệu có kích thước đủ nhỏ để có thể được truyền đi hoàn toàn trong khoảng thời gian là mỗi khe thời gian giữa các nút mạng lân cận

với nhau).

Các đóng góp của luận án trong chương này đã được công bố trong 02 bài báo khoa học đăng trên Tạp chí Khoa học và Công nghệ - Đại học Đà Nẵng [DJ1] và Tạp chí Khoa học công nghệ và Thông tin - Học viện Công nghệ Bưu chính Viễn thông [DJ2], và 01 bài báo tại hội nghị khoa học quốc tế ATC 2017 [IC5].

KẾT LUẬN

Nội dung luận án đã đạt được các mục tiêu đề ra là nghiên cứu ý nghĩa thực tế và hiệu quả của việc sử dụng kỹ thuật đệm dữ liệu cho mạng vô tuyến hướng nội dung và ảnh hưởng của các tham số của số lượng mảnh tin được phân mảnh từ mỗi tệp dữ liệu, bộ nhớ lưu trữ chia sẻ tại thiết bị di động người dùng và trạm gốc thông tin đối với hiệu năng mạng là thông lượng và độ trễ truyền tin. Các kiến thức nền tảng và các kết quả nghiên cứu đã được trình bày trong luận án với bố cục 3 Chương bao gồm: (1) Tổng quan vấn đề nghiên cứu; (2) Tối ưu hóa thông lượng và độ trễ của mạng vô tuyến hướng nội dung sử dụng mô hình dòng chảy; (3) Tối ưu hóa thông lượng và độ trễ của mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu. Các đóng góp chính đạt được của luận án có thể tóm tắt như sau:

1. Đề xuất mô hình mạng vô tuyến hướng nội dung sử dụng mô hình dòng chảy

Giải pháp tối ưu hóa thông lượng và độ trễ được đưa ra đối với mô hình mạng đề xuất, trong đó mỗi thiết bị di động và trạm gốc thông tin đều được trang bị các bộ nhớ lưu trữ chia sẻ hữu hạn. Các đóng góp của Nội dung này có thể được tóm tắt lại như sau:

- Đóng góp thứ nhất: Đề xuất được mô hình mạng trong đó cả các trạm gốc thông tin di động và thiết bị di động người dùng đều có

khả năng lưu trữ các tệp dữ liệu dữ liệu trong mạng với dung lượng lưu trữ khác nhau cũng như xem xét đến ảnh hưởng của tính di động người dùng và sử dụng phương pháp đa điểm để định tuyến truyền tải thông tin.

- Đóng góp thứ hai: Dựa trên mô hình mạng đề xuất, phương pháp định tuyến truyền tin phù hợp với mô hình mạng đã được trình bày, từ đó xây dựng được công thức tính thông lượng và độ trễ mạng, phân tích đưa ra giải pháp tối ưu hóa hiệu năng mạng. Các kết quả phân tích và tính toán được kiểm tra lại bởi các kết quả được giải bởi chương trình mô phỏng và tính toán trên máy tính Mathematica
- Đóng góp thứ ba: Phương pháp lưu trữ dữ liệu cơ bản trong đó số lượng bản sao của tệp dữ liệu được phân bố tại các thiết bị di động và trạm gốc thông tin di động được tối ưu một cách độc lập với nhau đã được trình bày và so sánh với phương pháp lưu trữ đề xuất.

Kết quả nghiên cứu tại Chương này cho thấy, ở chế độ Zipf cao $\alpha \geq \frac{3}{2}$, thông lượng và độ trễ tối ưu đạt được nhờ sử dụng phương pháp truyền tin đa chặng Người dùng tới người dùng, và do đó, việc sử dụng thêm bộ nhớ lưu trữ của các trạm gốc thông tin để lưu trữ thêm các tệp dữ liệu là không cần thiết. Lý do bởi vì phần lớn các tệp dữ liệu trong thư viện của mạng là các tệp dữ liệu có tính phổ biến cao và đang được lưu trữ tại phần lớn các thiết bị người dùng. Mặt khác, ở các chế độ Zipf khác, việc sử dụng thêm dung lượng lưu trữ $f(n)K_{BS}$ của các trạm gốc thông tin di động chia sẻ trong mạng vô tuyến hỗn hợp hướng nội dung giúp tăng đáng kể hiệu năng mạng so với trường hợp mạng không có sự

hiện diện của các trạm gốc thông tin [1, 15, 36, 43, 46]. Bên cạnh đó, khi hệ số Zipf $\alpha \geq 1 + \frac{\gamma-\beta}{2(\gamma+\delta-1)}$, mức tối ưu thông lượng và độ trễ mạng đạt được tương đương với mức tối ưu đạt được tại trường hợp mô hình mạng vô tuyến hỗn hợp hướng nội dung tĩnh sử dụng các trạm gốc thông tin được trang bị bộ nhớ lưu trữ chia sẻ có dung lượng vô hạn (tương đương với việc kết nối trực tiếp liên tục, không gián đoạn với đường truyền dẫn mạng lõi back-haul chứa tất cả các tệp dữ liệu của mạng) [36].

So sánh với phương pháp lưu trữ dữ liệu cơ bản trong đó số lượng bản sao của tệp dữ liệu được phân bố tại các thiết bị di động và trạm gốc thông tin di động được tối ưu một cách độc lập với nhau, hiệu năng mạng tối ưu đạt được theo giải pháp đề xuất ban đầu là tốt nhất nhờ tận dụng được điểm mạnh của việc phân bổ bản sao của các tệp dữ liệu đồng thời tại thiết bị người dùng và trạm gốc thông tin. Trong đó các thiết bị di động người dùng có thể được ưu tiên lưu trữ các tệp dữ liệu có mức độ phổ biến cao (lớn hơn $f(n)$) và các tệp dữ liệu có mức độ phổ biến thấp hơn sẽ được lưu trong bộ nhớ chia sẻ của các trạm gốc thông tin.

2. Đề xuất mô hình mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu

Giải pháp tối ưu hóa thông lượng và độ trễ được đưa ra đối với mô hình mạng đề xuất, trong đó kích thước tệp dữ liệu K là tham số mạng quan trọng cần được xem xét tới. Các đóng góp của Nội dung này có thể được tóm tắt lại như sau:

- Đóng góp thứ nhất: Mô hình mạng vô tuyến hướng nội dung sử dụng phương pháp phân mảnh tệp dữ liệu đã được đề xuất, trong đó mỗi

tệp dữ liệu được phân mảnh thành K mảnh tin. Các nút mạng nguồn cần tải lần lượt K mảnh tin rời rạc của tệp dữ liệu m để tổng hợp lại thành thông tin mong muốn.

- Đóng góp thứ hai: Từ mô hình mạng được đề xuất, mức cân bằng thông lượng và độ trễ của mạng đã được phân tích, tìm ra công thức tính toán và tối ưu hóa. So sánh đánh giá hiệu năng mạng giữa hai phương pháp thu nhận mảnh tin đề xuất là tuần tự và ngẫu nhiên. Các kết quả phân tích và tính toán đã được kiểm tra lại bởi các kết quả được giải bởi chương trình toán học được lập trình trên máy tính Mathematica.

Kết quả nghiên cứu tại Chương 3 cho thấy rằng, sử dụng phương pháp thu nhận mảnh tin ngẫu nhiên là giải pháp tối ưu hiệu năng mạng tốt hơn so với phương pháp thu nhận mảnh tin tuần tự. Ngoài ra, khi sử dụng kỹ thuật phân mảnh tệp dữ liệu, kích thước tệp dữ liệu không ảnh hưởng tới thông lượng tối ưu của mạng. Tuy nhiên, đối với tham số độ trễ tối ưu của mạng, khi phần lớn các tệp dữ liệu trong mạng đều có tính chất phổ biến cao và được lưu tại hầu hết các thiết bị người dùng ($\alpha \geq \frac{3}{2}$), thời gian để nhận được mỗi tệp dữ liệu mong muốn tỷ lệ K lần với kích thước tệp dữ liệu. Trong các trường hợp thông thường khác, để nhận được mỗi tệp dữ liệu mong muốn, thời gian cần thiết là gấp \sqrt{K} lần so với trường hợp không sử dụng phương pháp phân mảnh tệp dữ liệu (tương đương với $K = \Theta(1)$).

Hướng nghiên cứu tiếp theo của luận án sẽ tập trung nghiên cứu Tối ưu hóa thông lượng và độ trễ của mạng vô tuyến hướng nội dung sử dụng

các thuật toán mã hóa và sửa lỗi để nâng cao hiệu quả của kỹ thuật đệm dữ liệu.

Trên đây là một số kết luận về nội dung và những kết quả nghiên cứu của luận án. Nghiên cứu sinh chân thành cảm ơn các Thầy hướng dẫn và các nhà khoa học đã định hướng và góp ý để giúp nghiên cứu sinh hoàn thành luận án.

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

TẠP CHÍ KHOA HỌC QUỐC TẾ

- [IJ1] T.-A. Do, S.-W. Jeon, and W.-Y. Shin, “How to cache in mobile hybrid IoT networks?,” IEEE Access (Special Section on Smart Caching, Communications, Computing and Cybersecurity for Information-Centric Internet of Things), vol. 7, no. 1, pp. 27814-27828, March 2019.(Tạp chí quốc tế ISI, SCIE-Indexed)

TẠP CHÍ KHOA HỌC TRONG NƯỚC

- [DJ1] Đỗ Trung Anh, Đặng Hoài Bắc, “Độ trễ trong mạng multihop hướng nội dung sử dụng phương pháp phân mảnh tệp tin”, Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng, trang 1-5, quyển 2, 2017.
- [DJ2] Trung-Anh Do, Ngọc Chu Quang, and Hoài Bac Dang, “Optimal caching in content-centric mobile networks using file segmentation”, Journal of Science and Technology on Information and Communicaitons, vol. 1, no. 1-2, pp. 44-50, Jun. 2018.

HỘI NGHỊ KHOA HỌC

- [IC1] T.-A. Do, S.-W. Jeon, and W.-Y. Shin, "Caching in mobile HetNets: A throughput-delay trade-off perspective," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Barcelona, Spain, Jul. 2016, pp. 1247-1251. (Hội nghị quốc tế hàng đầu thường niên của ngành Lý thuyết thông tin (Information Theory))
- [IC2] A. T. Do and W.-Y. Shin, "Delay-throughput tradeoff of mobile social networks under random walk mobility," in Proc. Korea Inf. Commun. Society (KICS) Summer Conf., Jeju Island, Korea, Jun. 2015, pp. 447-448.
- [IC3] T.-A. Do, S.-W. Jeon, and W.-Y. Shin, "On the delay scaling of cache-enabled mobile networks," in Proc. Korea Inf. Commun. Society (KICS) Winter Conf., JeongSeon, Korea, Jan. 2016, pp. 420-421.
- [IC4] T.-A. Do, S.-W. Jeon, and W.-Y. Shin, "Caching in mobile HetNets: A throughput-delay trade-off perspective," in Proc. Korea Inf. Commun. Society (KICS) Summer Conf., Jeju Island, Korea, Jun. 2016, pp. 188-189.
- [IC5] Trung-Anh Do, Hoai Bac Dang, Won-Yong Shin, "On the delay of content-centric mobile multihop networks using file segmentation", in Proc. 2017 Int. Conf. Adv. Technol. Commun., Quy Nhon, Vietnam, Oct. 2017, pp. 301-305.

TÀI LIỆU THAM KHẢO

- [1] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput–delay scaling in wireless networks–Part I: The fluid model," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2568–2592, Jun. 2006.
- [2] A. Eriksson and B. Ohlman, "Scalable Object-to-Object Communication Over A Dynamic Global Network," Proc. *Future Network and Mobile-Summit '10*, June 2010.
- [3] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T.A. Thomas, J.G. Andrews, P. Xia, H.S. Jo, H.S. Dhillon, T.D. Novlan, "Heterogeneous cellular networks: From theory to practice", *IEEE Communications Magazine* vol. 50, no. 6, pp. 54–64, 2012.
- [4] A. Liu and V. Lau, "Exploiting base station caching in MIMO cellular networks: Opportunistic cooperation for video streaming," *IEEE Trans. Signal Processing*, vol. 63, no. 1, pp. 57–69, Jan 2015.
- [5] A. Malik, S. H. Lim, and W.-Y. Shin, "On the effects of subpacketization in content-centric mobile networks," *IEEE J. Sel. Areas Commun.* (Special Issue on Caching for Communication Systems and Networks), vol. 36, no. 8, pp. 1721–1736, August 2018.
- [6] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans.*

- Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.
- [7] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher and B. Ohlman, “A survey of information-centric networking,” in *IEEE Communications Magazine*, vol. 50, no. 7, pp. 26–36, July 2012.
- [8] B. Liu, Z. Liu, and D. Towsley, “On the capacity of hybrid wireless networks,” in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar./Apr. 2003, pp. 1543–1552.
- [9] C. Fricker, P. Robert, J. Roberts, and N. Sbihi, “Impact of traffic mix on caching performance in a content-centric network,” in *Proc. IEEE INFOCOM Workshop on Emerging Choices in Named-Oriented Netw. (NoMEN)*, Orlando, FL, Mar. 2012, pp. 310–315.
- [10] “Cisco Visual Networking Index: Forecast and Trends, 2017–2022,” *Cisco Public Information*, Nov. 2018.
- [11] “Cisco Annual Internet Report (2018–2023)” *Cisco Public Information*, Feb. 2020.
- [12] D. E. Knuth, “Big Omicron and big Omega and big Theta,” *ACM SIGACT News*, vol. 8, no. 2, pp. 18–24, Apr.—Jun. 1976.
- [13] E. Nygren, R. K. Sitaraman, and J. Sun, “The akamai network: a platform for high-performance internet applications”, *ACM SIGOPS Operating Systems Review*, vol. 44, no. 3, pp. 2–19, 2010.
- [14] F. Xue, L.-L. Xie, and P. R. Kumar, “The transport capacity of wireless networks over fading channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 834–847, Mar. 2005.

- [15] G. Alfano, M. Garetto, and E. Leonardi, "Content-centric wireless networks with limited buffers: When mobility hurts," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 299–311, Feb. 2016.
- [16] G. Zhang, J. Liu, J. Ren, L. Wang, and J. Zhang, "Capacity of content-centric hybrid wireless networks," *IEEE Access*, vol. 5, pp. 1449–1459, Feb. 2017.
- [17] G. Zhang, Y. Xu, X. Wang, and M. Guizani, "Capacity of hybrid wireless networks with directional antennas and delay constraint," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2097–2106, July 2010.
- [18] J. Yoon, W.-Y. Shin, and S.-W. Jeon, "Elastic routing in ad hoc networks with directional antennas," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3334–3346, Dec. 2017.
- [19] K. Poularakis, G. Iosifidis, I. Pefkianakis and L. Tassiulas, "Mobile data offloading through caching in residential 802.11 wireless networks," *IEEE Trans. Net. and Service Management*, vol. 13, no. 1, pp. 71–84, Mar. 2016.
- [20] K. Poularakis and L. Tassiulas, "Exploiting user mobility for wireless content delivery," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013.
- [21] H. Sarkissian, "The business case for caching in 4g lte networks," *Wireless 20—20*, vol. 20120, 2012.

- [22] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. IEEE INFOCOM*, 1999, vol. 1, pp. 126–134.
- [23] L. Zhou, R. Q. Hu, Y. Qian, and H.-H. Chen, "Energy-spectrum efficiency tradeoff for video streaming over mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 981–991, May. 2013.
- [24] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, Third Quart., 2016.
- [25] M. D'Ambrosio et al., "Providing Data Dissemination Services in the Future Internet," *Proc. WTC '08*, New Orleans, LA, Dec. 2008, at *IEEE GLOBECOM '08*.
- [26] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May. 2014.
- [27] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1029–1040, Aug. 2014.
- [28] M. A. Maddah-Ali and U. Niesen, "Coding for caching: Fundamental limits and practical challenges," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 23–29, Aug. 2016.
- [29] M. Cha, 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 *ACM IMC*, 2007.

- [30] M. E. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.
- [31] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.
- [32] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.
- [33] M. Ji, G. Caire, and A. F. Molisch, "The throughput-outage tradeoff of wireless one-hop caching networks," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6833–6859, Dec. 2015.
- [34] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.
- [35] M. M. Ahamed and S. Faruque, "5G Backhaul: Requirements, Challenges, and Emerging Technologies, Broadband Communications Networks" *Recent Advances and Lessons from Practice*, Abdelfatteh Haidine and Abdelhak Aqqal, *IntechOpen*, DOI: 10.5772/intechopen.78615, Nov. 2018.
- [36] M. Mahdian and E. Yeh, "Throughput and delay scaling of content-centric ad hoc and heterogeneous wireless networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 3030–3043, Aug. 2017.

- [37] M. X. Goemans, L. Li, V. S. Mirrokni, and M. Tholtan, "Market sharing games applied to content distribution in ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1020–1033, May. 2006.
- [38] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
- [39] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.
- [40] P. Li, C. Zhang, and Y. Fang, "The capacity of wireless ad hoc networks using directional antennas," *IEEE Trans. Mobile Comput.*, vol. 10, no. 10, pp. 1374–1387, Oct. 2011.
- [41] P. L. Vo, D. N. M. Dang, S. Lee, C. S. Hong and T.-Q. Le, "A Coalitional Game Approach for Fractional Cooperative Caching in Content-Oriented Networks", *Elsevier Computer Networks*, vol. 77, no. 11, pp. 144-152, Feb 2015.
- [42] S. Boyd, "Convex Optimization," *Cambridge University Press*, Aug. 2013.
- [43] S. Gitsenis, G. S. Paschos, and L. Tassiulas, "Asymptotic laws for joint content replication and delivery in wireless networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2760–2776, May. 2013.
- [44] S. Lim, W.-C. Lee, G. Cao, and C. R. Das, "A novel caching scheme for Internet based mobile ad hoc networks," in *Proc. IEEE Computer Commun. and Netw. (ICCCN)*, Dallas, TX, USA, Oct. 2003, pp. 38–43.

- [45] S. Tamoor-ul-Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-aho, "Caching in wireless small cell networks: A storage-bandwidth tradeoff," *IEEE Commun. Letters*, vol. 20, no. 6, pp. 1175–1178, June. 2016.
- [46] S.-W. Jeon, S.-N. Hong, M. Ji, G. Caire, and A. F. Molisch, "Wireless multihop device-to-device caching networks," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1662–1676, Mar. 2017.
- [47] T.-A. Do, S.-W. Jeon, and W.-Y. Shin, "Caching in mobile HetNets: A throughput-delay trade-off perspective," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1247-1251.
- [48] T.-A. Do and W.-Y. Shin, "Beamwidth scaling in wireless networks with outage constraints," in *Proc. Korea Inf. Commun. Society (KICS) Winter Conf.*, JeongSeon, Korea, Jan. 2015, pp. 130-131.
- [49] T.-A. Do and W.-Y. Shin, "Beamwidth scaling in wireless networks with outage constraints," *IEICE Trans. Commun.*, vol. E98-B, no. 11, pp. 2202-2211, Nov. 2015.
- [50] T.-A. Le, N. D. Thai, P. L. Vo, "The performance of caching strategies in content centric networking", in *Proc. IEEE Int. Conf. on Inf. Netw. (ICOIN)*, Danang, Vietnam, Apr. 2017, pp. 628-632.
- [51] T. Yamakami, "A ZIPF-like distribution of popularity and hits in the mobile web pages with short life time," in *Proc. Parallel Distrib. Comput. Appl. Technol.*, Taipei, Taiwan, Dec. 2006, pp. 240–243.
- [52] UMTS Forum, "Mobile traffic forecasts 2010–2020", no. 44, Jan. 2011.

- [53] V. Conan, J. Leguay, and T. Friedman, "Fixed point opportunistic routing in delay tolerant networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, pp. 773–782, Jun. 2008.
- [54] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," *Commun. ACM*, vol. 55, no. 1, pp. 117–124, Jan. 2012.
- [55] W.-Y. Shin, S.-Y. Chung, and Y. H. Lee, "Parallel opportunistic routing in wireless networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6290–6300, Oct. 2013.
- [56] W.-Y. Shin, S.-W. Jeon, N. Devroye, M. H. Vu, S.-Y. Chung, Y. H. Lee, and V. Tarokh, "Improved capacity scaling in wireless networks with infrastructure," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5088–5102, Aug. 2011.
- [57] X. Li, X. Wang, K. Li, and V. C. M. Leung, "CaaS: Caching as a service for 5G networks," *IEEE Access*, vol. 5, pp. 5982–5993, Mar. 2017.
- [58] X. Liu, K. Zheng, J. Zhao, X.-Y. Liu, X. Wang, and X. Di, "Information-centric networks with correlated mobility," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4256–4270, May 2017.
- [59] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5g systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 131–139, February 2014.

- [60] Y. Chen, L. Qiu, W. Chen, L. Nguyen and R. Katz, “Clustering web content for efficient replication,” *IEEE ICNP*, 2002.
- [61] Y. Cui, D. Jiang, “Analysis and Optimization of Caching and Multicasting in Large-Scale Cache-Enabled Heterogeneous Wireless Networks,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 250–264, Aug. 2017.
