



**BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

PHẠM HÙNG

**NGHIÊN CỨU KỸ THUẬT LẬP LỊCH CHO
MẠNG THÔNG TIN DI ĐỘNG THẾ HỆ MỚI**

Chuyên ngành: **Kỹ thuật viễn thông**

Mã số: **9.52.02.08**

LUẬN ÁN TIẾN SỸ KỸ THUẬT

Hà nội - 2021



**BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

LUẬN ÁN TIẾN SĨ

**NGHIÊN CỨU KỸ THUẬT LẬP LỊCH CHO
MẠNG THÔNG TIN DI ĐỘNG THẾ HỆ MỚI**

**Chuyên ngành: Kỹ thuật viễn thông
Mã số: 9.52.02.08**

**Nghiên cứu sinh: Phạm Hùng
Giảng viên hướng dẫn: PGS. TS. Nguyễn Tiến Ban
PGS. TS. Đặng Hoài Bắc**

Hà nội - 2021

LỜI CẢM ƠN

Trải qua 4 năm học ở Học Viện Công Nghệ Bru Chính Viễn Thông, được tiếp nhận những kiến thức và kinh nghiệm quý báu của các thầy cô đã giúp em có được ngày hôm nay, trưởng thành hơn và có kiến thức vững vàng hơn.

Em xin trân thành cảm ơn khoa Viễn Thông và khoa Sau Đại Học, Học Viện Công Nghệ Bru Chính Viễn Thông đã tạo điều kiện thuận lợi và giúp đỡ em rất nhiều trong quá trình học tập cũng như làm luận án tốt nghiệp.

Em cũng xin gửi lời cảm ơn trân thành tới các thầy cô trong Học Viện Công Nghệ Bru Chính Viễn Thông đã tận tình giảng dạy và trang bị cho em những kiến thức quý báu trong những năm học vừa qua.

Và đặc biệt em xin trân thành cảm ơn thầy Nguyễn Tiến Ban và thầy Đặng Hoài Bắc về những hỗ trợ quý báu giúp em vượt qua những giai đoạn khó khăn trong quá trình học tập, cung cấp nhiều kiến thức để em có thể hoàn thành xong luận án này.

Cảm ơn các anh chị đồng nghiệp và bạn bè đã góp ý và giúp đỡ tận tình trong quá trình thực hiện luận án này.

Em xin trân thành cảm ơn!

Hà Nội, tháng năm 2021

Nghiên cứu sinh

Phạm Hùng

Mục lục

Danh sách hình vẽ	iv
Danh sách bảng	iv
Danh mục từ viết tắt	viii
Từ điển chú giải	xi
Mở Đầu	1
1 Tính cấp thiết của luận án	1
2 Những vấn đề còn tồn tại	4
3 Ý nghĩa khoa học và thực tiễn	5
4 Mục tiêu nghiên cứu	5
5 Đối tượng nghiên cứu	6
6 Phương pháp nghiên cứu	7
7 Phạm vi nghiên cứu	7
8 Những đóng góp của luận án	7
9 Bố cục luận án	9
1 Tổng quan về kỹ thuật lập lịch	12
1.1 Hệ thống thông tin di động 5G	12
1.2 Yêu cầu kỹ thuật của hệ thống thông tin di động 5G	14
1.3 Hệ thống thông tin di động nhiều ăng ten đa người dùng MU-MIMO	15

1.3.1	Ước lượng kênh truyền theo chiều lên	16
1.3.2	Truyền dữ liệu chiều xuống	17
1.3.3	Xử lý tuyến tính	17
1.4	Hệ thống thông tin nhiều ăng ten cỡ rất lớn Massive MIMO	19
1.4.1	Ưu điểm của hệ thống thông tin nhiều ăng ten cỡ rất lớn .	20
1.4.2	Thách thức của hệ thống thông tin nhiều ăng ten cỡ rất lớn	21
1.5	Hệ thống thông tin nhiều ăng ten cỡ rất lớn mmWave Massive MIMO	22
1.5.1	Các thách thức của bước sóng cỡ mmWave Massive MIMO .	23
1.5.2	Mảng ăng ten cho mmWave Massive MIMO	24
1.5.3	Kiến trúc bộ RF	25
1.5.4	Ước lượng kênh truyền trong hệ thống mmWave Massive MIMO	26
1.5.5	Mô hình kênh truyền	27
1.5.6	Các giải pháp tiền mã hóa	27
1.6	Kỹ thuật lập lịch	30
1.6.1	Cấu trúc khung của mạng vô tuyến	30
1.6.2	Kỹ thuật lập lịch ở mạng vô tuyến	31
1.6.3	Thách thức	31
1.6.4	Một số kỹ thuật lập lịch	33
1.7	Kết luận chương 1	37

**2 Kỹ thuật lập lịch đảm bảo tốc độ tối thiểu của thuê bao dựa trên
thông tin lão hóa kênh truyền 39**

2.1	Ý tưởng kỹ thuật lập lịch đảm bảo tốc độ tối thiểu cho thuê bao khi sử dụng thông tin lão hóa kênh truyền	39
2.2	Mô hình hệ thống	42
2.3	Ước lượng kênh truyền	43
2.4	Truyền dữ liệu chiều xuống	44
2.5	Lão hóa kênh truyền	45
2.6	Tốc độ tổng có thể đạt được với hiện tượng lão hóa kênh truyền . .	46
2.6.1	Tiền mã hóa MRT	47
2.6.2	Tiền mã hóa ZF	47
2.7	Kỹ thuật lập lịch đề xuất QoS-Aware	48
2.8	Kết quả mô phỏng	51
2.9	Kết luận chương 2	57

3 Kỹ thuật lập lịch cho hệ thống ăng ten cỡ rất lớn đảm bảo đa

	tốc độ tối thiểu	59
3.1	Ý tưởng về kỹ thuật lập lịch đảm bảo đa tốc độ tối thiểu cho người dùng	59
3.2	Mô hình hệ thống	61
3.3	Ước lượng kênh truyền	61
3.4	Truyền tín hiệu chiều xuống	62
3.5	Bài toán tối ưu của kỹ thuật lập lịch	63
3.5.1	Các tiêu chí tối ưu	63
3.5.2	Xây dựng hàm mục tiêu	66
3.6	Giải pháp đề xuất	67
3.7	Kết quả mô phỏng	69

3.8	Kết luận chương 3	75
4	Kỹ thuật lập lịch cho hệ thống ăng ten bước sóng mm sử dụng	
	các bộ ADC phân giải thấp	76
4.1	Ý tưởng kỹ thuật lập lịch cho hệ thống ăng ten cỡ rất lớn bước sóng mm khi sử dụng các bộ ADC phân giải thấp	76
4.2	Mô hình hệ thống	78
4.2.1	Truyền dẫn chiều xuống	79
4.2.2	Mô hình lượng tử hóa	80
4.2.3	Tiền mã hóa lai đa người dùng hai giai đoạn cho kênh truyền một đường	81
4.3	Kỹ thuật lập lịch đề xuất	82
4.4	Kết quả mô phỏng	83
4.5	Kết luận chương 4	87
	Kết luận	89
	Danh mục các công trình nghiên cứu khoa học	91
	Tài liệu tham khảo	91

Danh sách hình vẽ

1	Lưu lượng sử dụng di động mỗi tháng	1
1.1	Các thành phần mạng vô tuyến 5G.	14
1.2	Mô hình hệ thống thông tin nhiều ăng ten đa người dùng	16
1.3	Mô hình xử lý tuyến tính tại trạm gốc	18
1.4	Kiến trúc của mạng 5G dựa trên bước sóng cỡ mm	23
1.5	Kiến trúc của mảng ăng ten lai bước sóng cỡ mm	25
1.6	Kiến trúc của bộ RF	26
1.7	Cấu trúc tiền mã hóa kỹ thuật số cho một thuê bao của hệ thống mmWave	28
1.8	Cấu trúc tạo búp sóng tương tự	29
1.9	Cấu trúc tạo búp sóng lai cho hệ thống một thuê bao	30
1.10	Cấu trúc khung của mạng vô tuyến NR	30
1.11	Cấu trúc khung con mạng vô tuyến NR	31
1.12	So sánh hiệu quả sử dụng phổ	35
1.13	Nhóm các ăng ten sử dụng	36
2.1	So sánh dung lượng hữu ích khi sử dụng MRT	52
2.2	So sánh dung lượng không hữu ích khi sử dụng MRT	53
2.3	Dung lượng hữu ích của hệ thống khi sử dụng MRT và $M = 128$.	54
2.4	So sánh dung lượng hữu ích khi sử dụng tiền mã hóa ZF và $K_a = 40$	55

2.5	So sánh lưu lượng không hữu ích khi sử dụng tiền mã hóa ZF với $K_a = 40$	56
2.6	Số thuê bao được phục vụ khi sử dụng ZF và MRT	57
3.1	Tốc độ mỗi thuê bao khi sử dụng kỹ thuật QoS-Assurance	71
3.2	Tốc độ tổng của hệ thống	72
3.3	So sánh tổng tốc độ	72
3.4	So sánh dung lượng hữu ích giữa hai kỹ thuật	73
3.5	So sánh về dung lượng giữa ba kỹ thuật	74
4.1	Hiệu quả sử dụng phổ theo SNR	85
4.2	So sánh hiệu quả sử dụng phổ giữa bộ ADC phân giải thấp và No-ADC	86
4.3	Hiệu quả sử dụng phổ theo K_S	87

Danh sách bảng

3.1	Tham số mô phỏng	70
-----	----------------------------	----

Danh mục từ viết tắt

- ADC** Analog To Digital Converter - bộ biến đổi tương tự số
- AAU** Active Antenna Unit - Đơn vị ăng ten chủ động
- BS** Base Station - trạm gốc
- BB** Baseband - băng cơ sở
- BBU** Baseband Unit - Đơn vị băng cơ sở
- BPF** Bandpass Filter - bộ lọc thông dải
- CSI** Channel State Information - thông tin trạng thái kênh truyền
- CU** Central Unit - Đơn vị tập trung
- DPC** Dirty Paper Coding - mã hóa thuật toán giấy bẩn
- DU** Distributed Unit - Đơn vị phân tán
- FDD** Frequency Division Duplex - ghép kênh phân chia theo tần số
- GFBR** Guaranteed Flow Bit Rate - tốc độ luồng bit được đảm bảo
- GBR** Guaranteed Bit Rate- tốc độ bit được đảm bảo
- gNB** Next Generation NodeB - trạm gốc thế hệ tiếp theo
- IoT** Internet of Things - kết nối vạn vật
- IP** Internet Protocol - giao thức internet
- OFDM** Orthogonal Frequency Division Multiplexing - ghép kênh phân chia theo tần số trực giao
- LNA** Low Noise Amplifier - bộ khuếch đại nhiễu thấp

LTE Long Term Evolution - mạng tiến hóa dài hạn

LOS Line-of-Sight - truyền sóng trong tầm nhìn thẳng

MRT Maximum Ratio Transmission - truyền dẫn tốc độ cực đại

MU-MIMO Multi-user MIMO - hệ thống nhiều ăng ten nhiều người dùng

MIMO Multiple Input Multiple Output - nhiều ăng ten thu nhiều ăng ten phát

Massive MIMO hệ thống ăng ten nhiều ăng ten đầu vào và nhiều ăng ten đầu ra cỡ rất lớn

MR Maximum Rate - tối đa tốc độ

MMSE Minimum Mean Square Error - Lỗi bình phương trung bình nhỏ nhất

MS Mobile Subscriber - thuê bao

N-LOS Non-Line-of-Sight - truyền sóng không trong tầm nhìn thẳng

NR New Radio - Mạng vô tuyến thế hệ mới

ZF Zero Forcing - cưỡng bức bằng không

PF Proportional Fair - công bằng giữa các thuê bao

QoS Quality of Service - chất lượng dịch vụ

RF Radio Frequency - tần số vô tuyến điện

RRH Remote Radio Head - Khối vô tuyến từ xa

SNR Signal-To-Noise Ratio - tỷ lệ tín hiệu trên nhiễu

SDAP Service Data Adaptation Protocol - giao thức thích nghi dữ liệu dịch vụ

TDD Time Division Duplex - ghép kênh phân chia theo thời gian

ULA Uniform Linear Array - mảng tuyến tính đồng nhất

Từ điển chú giải

M Số ăng ten. x

T Chu kỳ khung. x

β_k Hệ số fa đình chậm của kênh truyền. x

$\mathbb{E}[\cdot]$ Phép toán kỳ vọng. x

\mathbb{H} Ma trận kênh truyền. x

\mathbb{K}_a Tập thuê bao trong tế bào gồm K_a thuê bao. x

\mathbb{K}_p Tập thuê bao được ướng lượng kênh gồm K_p thuê bao. x

\mathbb{K}_s Tập thuê bao được phục vụ gồm K_s thuê bao. x

\mathbf{A}^* Ma trận chuyển vị liên hợp của ma trận \mathbf{A} . x

\mathbf{A}^T Ma trận chuyển vị của ma trận \mathbf{A} . x

\mathbf{F} Ma trận tiền mã hóa. x

\mathbf{I}_N Ma trận đơn vị kích thước $N \times N$. x

\mathbf{g} Véc tơ kênh truyền Rayleigh fa đình. x

\mathbf{h} Véc tơ kênh truyền. x

$\mathcal{CN}(0, \sigma_r^2)$ Phân bố chuẩn trung bình không, phương sai σ_r^2 . x

σ Công suất nhiễu đường xuống. x

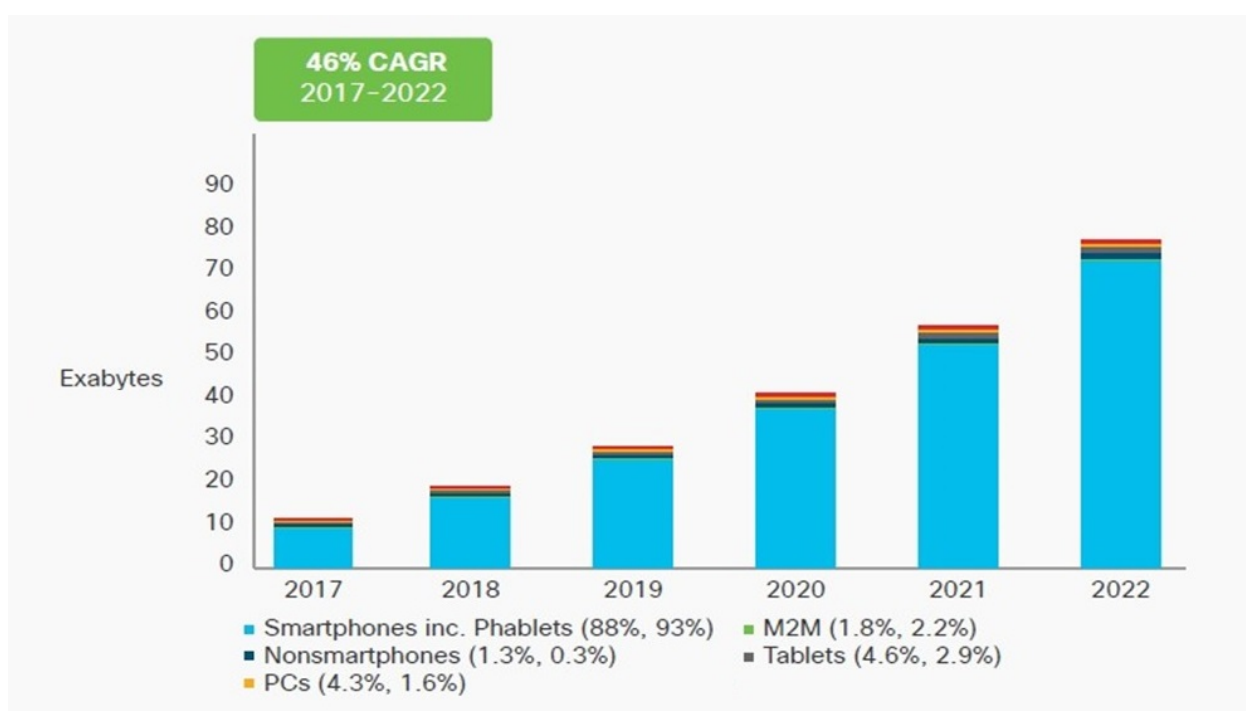
σ_r Công suất nhiễu đường lên. x

τ_p Khoảng thời gian ước lượng kênh. x

Mở Đầu

1 Tính cấp thiết của luận án

Hiện nay nhu cầu sử dụng dịch vụ di động ngày càng lớn do tính phổ biến và tiện dụng của các thiết bị đầu cuối di động. Theo dự báo của Cisco năm 2019, lưu lượng di động được dự báo tiếp tục tăng đều qua các năm từ mức 10 exabytes một tháng trong năm 2017 lên mức hơn 80 exabytes một tháng trong năm 2022, như trong hình 1 [24].



Hình 1: Lưu lượng sử dụng di động mỗi tháng

Bên cạnh đó, các ứng dụng mới trong kỷ nguyên của Internet of Things - kết

nổi vạn vật (IoT) sẽ tiếp tục đặt ra nhiều cơ hội cũng như thách thức cho sự phát triển của ngành di động. Các hệ thống camera giám sát sẽ rất phổ biến và theo dõi xã hội mọi nơi và mọi lúc. Các động vật cũng được giám sát và theo dõi sức khỏe từ xa như con người. Các ô tô tự lái sẽ được nối mạng để nhận biết môi trường cũng như các nguy cơ tiềm ẩn theo thời gian thực. Các bác sĩ có thể cứu chữa cho bệnh nhân từ xa qua mạng thông tin di động, không còn phụ thuộc môi trường địa lý. Việc tương tác hai chiều sẽ đòi hỏi mạng thông tin di động có chất lượng cao hơn về tốc độ, độ tin cậy và độ trễ.

Mạng di động các thế hệ đầu tiên như 2G và 3G chủ yếu phục vụ nhu cầu thoại của người dùng với công nghệ chuyển mạch kênh là chủ yếu. Chỉ đến 4G thì tốc độ dữ liệu của người dùng mới được cải thiện và mạng lõi đã hoàn toàn sử dụng công nghệ chuyển mạch gói, tuy nhiên tài nguyên vô tuyến vẫn rất hữu hạn. Một đặc điểm khác của mạng vô tuyến là sự biến thiên theo thời gian, theo tần số và cả không gian. Vì vậy, lập lịch được sử dụng như một lớp thích nghi với sự biến đổi của môi trường vô tuyến và các yêu cầu khác nhau về chất lượng dịch vụ của mạng lõi chuyển mạch gói. Tuy nhiên, các kỹ thuật lập lịch cho mạng 4G được giới hạn chủ yếu vào các nhóm sau [57]:

- **Tối ưu tốc độ hệ thống:** kỹ thuật lập lịch sẽ chọn ra các thuê bao có chất lượng kênh truyền tốt nhất để phục vụ nhằm đạt được tốc độ hệ thống cao nhất. Tuy nhiên, nhược điểm của kỹ thuật này là các thuê bao ở trạng thái kênh truyền kém sẽ không được phục vụ.
- **Bình quân về tốc độ:** kỹ thuật lập lịch sẽ phục vụ mọi thuê bao sao cho dung lượng sử dụng của mọi thuê bao sẽ gần bằng nhau nhất. Đây là kỹ thuật được sử dụng chủ yếu khi triển khai thực tế của mạng 4G. Tuy nhiên, nhược điểm của kỹ thuật này là tốc độ hệ thống đạt được sẽ kém hơn nhiều so với

kỹ thuật tối ưu tốc độ hệ thống.

Mạng 5G đem lại sự thay đổi rất lớn so với các mạng trước đây cả về độ rộng băng tần sử dụng và đặc biệt là sử dụng số lượng ăng ten cỡ rất lớn phía trạm gốc. Dung lượng hệ thống tăng tỷ lệ thuận với số lượng ăng ten sử dụng nên dung lượng của trạm gốc trong 5G sẽ tiệm cận với 10 Gbps tức là gấp 10 lần tốc độ của mạng 4G. Sử dụng sóng mmWave với dải tần từ 28 Ghz đến 300 Ghz khiến tài nguyên vô tuyến lớn hơn rất nhiều so với mạng 4G. Vì vậy, kỹ thuật lập lịch áp dụng cho mạng 5G cần hướng tới việc đảm bảo chất lượng người dùng hoặc tối ưu về hiệu quả sử dụng hơn là đơn thuần nhằm tăng tốc độ hệ thống hoặc phục vụ bình quân cho mọi thuê bao.

Các hướng nghiên cứu chính về kỹ thuật lập lịch trong mạng 5G hiện nay là:

- Tăng dung lượng hệ thống nhờ đặc tính kênh truyền hệ thống ăng ten cỡ rất lớn là ổn định theo thời gian nên có thể sử dụng thông tin lão hóa kênh truyền để ước lượng kênh truyền hiện tại của thuê bao. Nhờ đó, giảm được tài nguyên ước lượng kênh truyền mà vẫn có thể phục vụ thêm thuê bao.
- Do số lượng ăng ten là rất lớn nên có thể lựa chọn các ăng ten có kênh truyền tốt nhất để phục vụ mà không cần tiêu tốn công suất và các ăng ten không ở trong điều kiện tối ưu cho thuê bao. Xuất hiện hướng nghiên cứu của kỹ thuật lập lịch về các tiêu chí để lựa chọn tập thuê bao và tập ăng ten phục vụ.
- Do dung lượng hệ thống của 5G đã đạt được tốc độ cao nên bắt đầu quan tâm đến việc bảo đảm tốc độ tối thiểu cho thuê bao. Lúc này bài toán lập lịch không chỉ phục vụ thuê bao bằng mọi giá mà còn phải đánh giá xem tốc độ phục vụ của thuê bao có đạt được yêu cầu của chất lượng dịch vụ QoS hay

không.

- Lập lịch đảm bảo độ trễ tối thiểu cho thuê bao. Đây là hướng lập lịch đảm bảo chất lượng dịch vụ QoS hướng đến các ứng dụng đòi hỏi độ trễ thấp trong mạng 5G.

Tuy nhiên, do đặc tính về chất lượng dịch vụ trong mạng lõi dựa trên công nghệ IP phức tạp hơn rất nhiều so với những kỹ thuật lập lịch hiện nay trong các nghiên cứu về mạng 5G nên cần có những cải tiến hơn nữa về kỹ thuật lập lịch để đáp ứng được nhu cầu của người dùng.

2 Những vấn đề còn tồn tại

Việc áp dụng thông tin lão hóa kênh truyền để làm tăng dung lượng hệ thống cho mạng 5G đã góp phần tăng mạnh dung lượng hệ thống tuy nhiên các nghiên cứu hiện nay về kỹ thuật lập lịch áp dụng thông tin lão hóa kênh truyền chưa quan tâm đến việc sử dụng thông tin lão hóa kênh truyền có thể dẫn đến tình trạng tốc độ đạt được của thuê bao là không cao khi sử dụng thông tin lão hóa kênh truyền quá cũ. Trường hợp khác có thể xảy ra là các nghiên cứu về kỹ thuật lập lịch sử dụng thông tin lão hóa kênh truyền không chỉ ra được số thuê bao tối ưu cần được phục vụ là bao nhiêu. Dẫn đến trong nhiều trường hợp số lượng thuê bao được phục vụ là quá lớn dẫn đến tốc độ trung bình của các thuê bao sẽ bị giảm xuống. Vì vậy, cần tham chiếu các thông số về chất lượng dịch vụ của phần mạng lõi IP để tác động lên quá trình lập lịch nhằm đảm bảo được chất lượng dịch vụ cho người dùng.

Hiện tại, có một số nghiên cứu về kỹ thuật lập lịch đang chú trọng về đảm bảo một tốc độ tối thiểu cho toàn bộ thuê bao và sử dụng các thuật toán vét cạn để tìm lời giải tối ưu. Tuy nhiên thực tế là các ứng dụng chạy trên mạng thông tin di

động có nhiều yêu cầu về tốc độ tối thiểu khác nhau. Việc sử dụng các tính toán quá phức tạp sẽ dẫn đến độ trễ xử lý trong việc lập lịch của thuê bao. Vì vậy, cần thiết đề xuất ra kỹ thuật lập lịch cho người dùng với nhiều tốc độ tối thiểu khác nhau và kỹ thuật đề xuất cần đơn giản trong tính toán nhằm giảm thời gian tính toán, đáp ứng được yêu cầu về độ trễ của người dùng.

Hơn nữa, các thông số về chất lượng dịch vụ trong mạng lõi không chỉ là độ trễ mà còn có các thông tin về độ ưu tiên của dịch vụ. Độ ưu tiên của dịch vụ khác nhau có thể dẫn đến sự khác biệt hoàn toàn về giá cước dịch vụ giữa các người dùng. Vì vậy, thông số về độ ưu tiên của dịch vụ trong mạng lõi nên là một tham số đánh giá trong quá trình lập lịch.

3 Ý nghĩa khoa học và thực tiễn

Ý nghĩa khoa học của luận án mà luận án hướng tới là xây dựng các mô hình toán học và chương trình mô phỏng để đánh giá ảnh hưởng của kỹ thuật lập lịch lên hiệu năng hệ thống thông tin di động 5G. Các mô hình toán học, chương trình tính toán và mô phỏng sẽ là công cụ cho việc thiết kế, đánh giá hiệu quả của các kỹ thuật lập lịch đề xuất. Ý nghĩa thực tiễn của luận án mà nghiên cứu sinh kỳ vọng đạt được thể hiện ở các kỹ thuật lập lịch đề xuất có thể áp dụng được vào thực tế để đảm bảo tốc độ tối thiểu cho thuê bao, khai thác đặc tính trực giao của kênh truyền vô tuyến và dung lượng của hệ thống thông tin di động 5G.

4 Mục tiêu nghiên cứu

Xuất phát từ những phân tích ở trên, luận án tập trung thực hiện các mục tiêu chính sau đây:

- Đề xuất kỹ thuật lập lịch và tính toán ảnh hưởng của việc sử dụng thông tin

lão hóa kênh truyền lên tốc độ của từng thuê bao cũng như toàn hệ thống để đảm bảo rằng các thuê bao được phục vụ sẽ được cung cấp một tốc độ tối thiểu trong khi vẫn cải thiện dung lượng hệ thống.

- Đề xuất kỹ thuật lập lịch và thực hiện mô hình hóa toán học để đảm bảo các thuê bao được phục vụ sẽ có được tốc độ tối thiểu theo yêu cầu và chọn ra được tập thuê bao tối ưu để dung lượng hệ thống là cao nhất. Hơn nữa kỹ thuật lập lịch sẽ tính đến độ ưu tiên của dữ liệu, đảm bảo rằng khi các yếu tố khác là như nhau thì thuê bao có độ ưu tiên cao hơn sẽ được phục vụ trước.
- Đề xuất kỹ thuật lập lịch để khai thác đặc tính của kênh truyền trực giao trong mạng 5G. Dựa vào đặc tính kênh truyền để xây dựng ngưỡng giới hạn để tìm ra tập thuê bao có đặc tính kênh truyền cần tìm kiếm, từ đó giảm bớt được thời gian tìm tập thuê bao phục vụ tối ưu.

5 Đối tượng nghiên cứu

Đối tượng nghiên cứu của luận án là kỹ thuật lập lịch sử dụng tại Next Generation NodeB - trạm gốc thế hệ tiếp theo (gNB) của mạng thông tin di động thế hệ mới 5G nhằm mục đích lựa chọn tập thuê bao phục vụ và phân phối tài nguyên vô tuyến cho các thuê bao một cách tối ưu để đạt dung lượng toàn hệ thống là cao nhất. Để đạt được mục tiêu đó cần nghiên cứu các kỹ thuật chính của hệ thống thông tin di động 5G như ước lượng kênh truyền và tiền mã hóa. Đồng thời nghiên cứu các yêu cầu của thuê bao trong mạng thông tin di động 5G để thỏa mãn các yêu cầu đó khi xây dựng các kỹ thuật lập lịch.

6 Phương pháp nghiên cứu

- Từ kết quả khảo sát các nghiên cứu liên quan sẽ chỉ ra các hạn chế còn tồn tại, xác định bài toán cần nghiên cứu và tiến đến đưa ra các đề xuất nhằm giải quyết bài toán đặt ra.
- Sử dụng các công cụ toán học để tính toán hiệu quả của kỹ thuật đề xuất.
- Mô phỏng hoạt động của kỹ thuật đề xuất nhằm kiểm định lại tính đúng đắn của đề xuất.

7 Phạm vi nghiên cứu

Các nghiên cứu của luận án được giới hạn trong phạm vi dưới đây:

- Nghiên cứu kỹ thuật lập lịch trong hệ thống ăng ten nhiều đầu vào và nhiều đầu ra cỡ rất lớn.
- Nghiên cứu hệ thống sử dụng tiền mã hóa Maximum Ratio Transmission - truyền dẫn tốc độ cực đại (MRT) và Zero Forcing - cưỡng bức bằng không (ZF).
- Thực hiện các kỹ thuật để chọn thuê bao tối ưu về tốc độ hệ thống nhưng phải đảm bảo chất lượng dịch vụ cho toàn bộ thuê bao.
- Can thiệp vào công suất thu phát của thuê bao để thay đổi tốc độ của thuê bao theo yêu cầu về tốc độ tối thiểu khác nhau.

8 Những đóng góp của luận án

Luận án đã đạt được các kết quả nghiên cứu và đóng góp chính như sau:

- **Trong đóng góp đầu tiên, luận án đề xuất một kỹ thuật lập lịch có thể đảm bảo tốc độ tối thiểu cho thuê bao khi hệ thống sử dụng thông tin lão hóa kênh truyền**

Khắc phục các hạn chế của các nghiên cứu trước đó khi đề xuất việc lập lịch sử dụng thông tin lão hóa kênh truyền chỉ nhằm vào mục đích tăng tối đa dung lượng hệ thống, luận án đề xuất bên cạnh mục tiêu tăng dung lượng hệ thống, cần đánh giá xem liệu việc sử dụng thông tin lão hóa kênh truyền có thể đảm bảo được yêu cầu về tốc độ tối thiểu cho thuê bao hay không khi sử dụng các thuật toán tiền mã hóa MRT và ZF.

Kết quả mô phỏng cho thấy giải pháp đề xuất có thể đảm bảo tốc độ tối thiểu cho các thuê bao đã sử dụng thông tin lão hóa kênh, dung lượng hữu ích của hệ thống hoàn toàn vượt trội kỹ thuật lập lịch được đem ra so sánh, đặc biệt khi số lượng thuê bao phục vụ ở mức rất lớn.

Kết quả nghiên cứu liên quan được công bố trong công trình [J1].

- **Trong đóng góp thứ hai, luận án đề xuất một kỹ thuật lập lịch cung cấp được nhiều tốc độ tối thiểu khác nhau tùy theo yêu cầu của người dùng**

Khác với các nghiên cứu về đảm bảo chất lượng QoS chỉ tập trung về độ trễ, luận án đề xuất bên cạnh tham số về độ trễ cần quan tâm tới các tham số khác trong QoS như độ ưu tiên và tốc độ tối thiểu cho thuê bao, và đặc biệt là cho phép phục vụ nhiều thuê bao có tốc độ tối thiểu khác nhau trong cùng một chu kỳ khung. Kỹ thuật lập lịch được đề xuất sẽ tính toán công suất tối ưu cho từng thuê bao để đạt được tốc độ yêu cầu và không ảnh hưởng đến chất lượng của thuê bao khác khi sử dụng tiền mã hóa MRT. Thuật toán cũng

đã tham chiếu đến mức độ ưu tiên của lưu lượng, từ đó đảm bảo độ ưu tiên cho các lưu lượng quan trọng của hệ thống.

Kết quả mô phỏng cho thấy, kỹ thuật lập lịch đề xuất đã đáp ứng đúng tốc độ tối ưu mà thuê bao yêu cầu, và có thể phục vụ nhiều mức tốc độ tối thiểu khác nhau trong cùng một chu kỳ khung.

Kết quả nghiên cứu liên quan được công bố trong công trình [J2].

- **Trong đóng góp thứ ba, luận án đề xuất một kỹ thuật lập lịch khai thác đặc tính trực giao của kênh truyền trong mạng mmWave sử dụng Analog To Digital Converter - bộ biến đổi tương tự số (ADC) phân giải thấp**

Các nghiên cứu cho thấy việc tiết kiệm công suất bằng các bộ ADC phân giải thấp nên áp dụng tại phía thuê bao người dùng tức là nghiên cứu áp dụng cho hệ thống mmWave theo chiều xuống. Luận án tính toán tốc độ của từng thuê bao cũng như toàn hệ thống khi sử dụng tiền mã hóa lai hai giai đoạn. Kỹ thuật đề xuất đã xác định điều kiện về trực giao của kênh truyền thuê bao để tìm ra tập thuê bao phục vụ bán trực giao từng đôi một với nhau.

Kết quả mô phỏng cho thấy, kỹ thuật lập lịch đề xuất có thể hoạt động tốt trong cả hai miền Signal-To-Noise Ratio - tỷ lệ tín hiệu trên nhiễu (SNR) thấp và cao cũng như hoạt động tốt với bộ ADC phân giải thấp và cao.

Kết quả nghiên cứu liên quan được công bố trong công trình [J3].

9 Bố cục luận án

Trong chương 1, luận án trình bày tổng quan về vấn đề nghiên cứu bao gồm mạng thông tin di động thế hệ mới 5G, hệ thống Multi-user MIMO - hệ thống

nhiều ăng ten nhiều người dùng (MU-MIMO). Chương này cũng trình bày các vấn đề ước lượng kênh truyền, mô hình kênh truyền và các giải pháp tiền mã hóa trong hệ thống ăng ten nhiều ăng ten đầu vào và nhiều ăng ten đầu ra cỡ rất lớn (Massive MIMO) và mmWave Massive MIMO. Các hướng nghiên cứu chính về kỹ thuật lập lịch hiện nay và các hạn chế của nó cũng được trình bày. Từ đó, chỉ ra các hướng nghiên cứu nhằm cải tiến và khắc phục các vấn đề còn tồn tại trong các kỹ thuật lập lịch hiện nay.

Trong chương 2, luận án đã đề xuất một kỹ thuật toán lập lịch đảm bảo tốc độ tối thiểu khi thuê bao sử dụng thông tin lão hóa kênh truyền. Chương này đã xây dựng mô hình kênh để tính toán ảnh hưởng của việc sử dụng thông tin lão hóa kênh truyền lên vấn đề ước lượng kênh, tiền mã hóa MRT và ZT từ đó tính ra tốc độ của thuê bao khi sử dụng thông tin lão hóa kênh. Để đảm bảo tốc độ tối thiểu của thuê bao, kỹ thuật đề xuất thêm vào một bước trong quá trình lập lịch để lựa chọn các thuê bao phù hợp để phục vụ bằng cách so sánh tốc độ có thể đạt được của cả thuê bao được ước lượng kênh và cả các thuê bao chỉ sử dụng thông tin ước lượng kênh. Thực nghiệm kỹ thuật cho nhiều tốc độ tối thiểu khác nhau và các tình huống thực tế khác nhau như khi tăng dần số thuê bao hoặc số lượng ăng ten sử dụng.

Trong chương 3, luận án đề xuất kỹ thuật lập lịch nhằm đảm bảo nhiều loại tốc độ tối thiểu cho thuê bao. Giải pháp giúp phân bổ tối ưu công suất phát theo điều kiện kênh truyền và tốc độ mà thuê bao yêu cầu, trong khi vẫn hướng đến mục tiêu tổng dung lượng hệ thống là cao nhất. Kỹ thuật còn sử dụng thông tin về độ ưu tiên của lưu lượng để đảm bảo rằng các lưu lượng quan trọng hơn sẽ được phục vụ trước. Và kỹ thuật sử dụng một tiêu chí tổng hợp các tham số nên sẽ có thời gian tính toán ít hơn nhiều so với các kỹ thuật vét cạn nên hoàn toàn

khả thi để áp dụng triển khai cho mạng thông tin di động 5G.

Trong chương 4, luận án đề xuất kỹ thuật lập lịch cho hệ thống mmWave sử dụng bộ ADC phân giải thấp. Giải pháp đề xuất giúp giảm thời gian tìm kiếm tập thuê bao phục vụ tối ưu trong khi vẫn hướng đến mục tiêu tổng dung lượng hệ thống là cao nhất. Kỹ thuật sử dụng điều kiện bán trực giao về kênh truyền để loại bỏ các thuê bao thuộc tập ứng cử viên mà không bán trực giao với tập thuê bao đã được lựa chọn.

Cuối cùng là phần kết luận chung, sẽ tổng kết lại những đóng góp, kết quả nghiên cứu đạt được trong luận án này cũng như hướng nghiên cứu trong tương lai.

Chương 1

Tổng quan về kỹ thuật lập lịch

Chương này trình bày tổng quan về kỹ thuật ước lượng kênh và xử lý tuyến tính của hệ thống thông tin di động thế hệ mới 5G. Từ đó xác định các ưu điểm và thách thức chung của mạng 5G, và tiếp theo là khảo sát các kỹ thuật lập lịch hiện nay. Đó là những lý thuyết nền tảng cho những đề xuất ở các chương tiếp theo của luận án.

1.1 Hệ thống thông tin di động 5G

Trong những năm gần đây, lưu lượng dữ liệu trong mạng thông tin di động đã tăng đột biến với sự phổ biến của thiết bị di động cá nhân như điện thoại thông minh, máy tính bảng và các loại thiết bị không dây khác. Lưu lượng dữ liệu di động toàn cầu năm 2018 đã tăng gấp sáu lần so với năm 2014. Để đáp ứng được nhu cầu này cần xây dựng các công nghệ mới. Thông lượng hệ thống thông tin di động được tính theo công thức sau:

$$T(\text{bits/s/km}^2) = D(\text{Cells/km}^2) \times B(\text{Hz}) \times E(\text{bits/s/Hz/Cell}) \quad (1.1)$$

Trong đó T là dung lượng hệ thống, D là mật độ tế bào, B là băng thông hệ thống và E là hiệu quả sử dụng phổ. Bên cạnh việc thu nhỏ bán kính vùng phủ để tăng mật độ tế bào thì hai phương án tối ưu để tăng thông lượng hệ thống là

tăng băng thông và tăng hiệu quả sử dụng phổ. Các hệ thống ăng ten thu phát nhiều đầu vào và nhiều đầu ra (MIMO) cho phép truyền nhiều luồng dữ liệu đồng thời giữa bên thu và bên phát từ đó đạt được độ lợi về ghép kênh không gian, làm tăng mạnh hiệu quả sử dụng phổ.

Các nghiên cứu về ghép kênh không gian đã chuyển từ MIMO truyền thống sang hệ thống MIMO đa người dùng (MU-MIMO), nơi mà Base Station - trạm gốc (BS) có nhiều ăng ten sẽ phục vụ nhiều thuê bao chỉ có một ăng ten. Điều này rất quan trọng trong thực tế vì lý do kích thước và pin của điện thoại mà việc tăng số ăng ten tại trạm gốc là dễ dàng hơn tại thuê bao. Càng nhiều ăng ten tại trạm gốc thì độ lợi về ghép kênh không gian càng tăng lên, đi cùng với nó là sự phức tạp trong tính toán và đắt đỏ trong phần cứng.

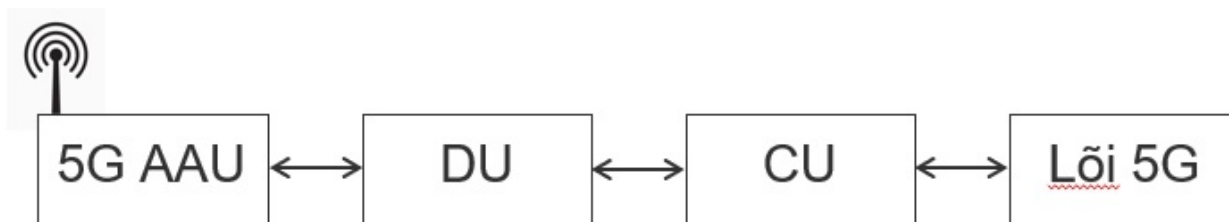
Marzetta đã chứng minh rằng khi số lượng ăng ten vô cùng lớn thì chỉ cần sử dụng các phép toán tuyến tính đơn giản như kết hợp tỷ lệ cực đại MRC cho đường lên cũng có thể đạt được hiệu quả gần tối ưu như trong MIMO truyền thống [43]. Hệ thống như vậy được gọi là Massive MIMO.

Tăng băng thông hệ thống là một cách để tăng thông lượng hệ thống. Tuy nhiên hiện nay dải phổ phù hợp nhất cho liên lạc vô tuyến ở dưới 6 GHz đã sử dụng gần hết. Vì vậy mạng thông tin di động 5G đang nghiên cứu sử dụng phổ tần ở băng sóng mmWave.

Massive MIMO và mmWave là các công nghệ hứa hẹn cho mạng thông tin 5G. Tuy nhiên, còn rất nhiều thách thức được đặt ra do có rất nhiều thuê bao được phục vụ đồng thời như vấn đề can nhiễu giữa các thuê bao, thu thập thông tin về chất lượng kênh và cuối cùng là dung lượng hệ thống phụ thuộc rất nhiều vào kỹ thuật lập lịch để chọn ra tập thuê bao tối ưu.

Kiến trúc của lớp vô tuyến trong mạng 5G cũng có sự thay đổi lớn như hình vẽ

1.1. Remote Radio Head - Khối vô tuyến từ xa (RRH) trong mạng 4G thường kết nối với ăng ten ngoài nay được tích hợp trong một Active Antenna Unit - Đơn vị ăng ten chủ động (AAU). Baseband Unit - Đơn vị băng cơ sở (BBU) vốn là phần cứng theo từng hãng chuyên biệt cũng sẽ được tách ra thành Distributed Unit - Đơn vị phân tán (DU) và Central Unit - Đơn vị tập trung (CU) và có thể được ảo hóa đặt tại các tài nguyên tính toán của mạng.



Hình 1.1: Các thành phần mạng vô tuyến 5G.

1.2 Yêu cầu kỹ thuật của hệ thống thông tin di động 5G

Tháng 9 năm 2015, ITU đã định nghĩa các yêu cầu kỹ thuật cơ bản cho mạng 5G như sau [31]:

- Tốc độ dữ liệu đỉnh: tốc độ có thể đạt mức 10 Gbps so với 1 Gbps của mạng 4G.
- Tốc độ dữ liệu người dùng thực tế: trong khu vực đô thị và ngoại ô, người dùng có thể trải nghiệm tốc độ 100 Mbps so với 10 Mbps của mạng 4G.
- Hiệu quả phổ tần được tăng gấp 3 lần so với mạng 4G.
- Tính di động: mạng 5G cho phép người dùng di chuyển với tốc độ 500 km/giờ mà vẫn đạt được yêu cầu chất lượng dịch vụ QoS ở mức chấp nhận được trong khi ở 4G là 350 km/giờ.

- Độ trễ: giảm từ 10 ms của mạng 4G xuống còn 1 ms ở 5G để cung cấp được các dịch vụ yêu cầu độ trễ thấp như lái ô tô tự động.
- Mật độ kết nối: mạng 5G kỳ vọng sẽ có mật độ kết nối gấp 10 lần mạng 4G để phục vụ cho kết nối giữa các thiết bị với nhau.
- Hiệu quả sử dụng năng lượng: độ tiêu thụ năng lượng sẽ không cao hơn mạng 4G nhưng có hiệu quả cao hơn, kỳ vọng sẽ gấp 100 lần hiệu quả sử dụng năng lượng của mạng 4G.
- Dung lượng dữ liệu: mạng 5G kỳ vọng sẽ có dung lượng 10 $Mbps/m^2$, tức là cao hơn mạng 4G 100 lần.

Tùy từng hoàn cảnh cụ thể của người dùng mà mỗi yêu cầu cụ thể sẽ được đánh giá. Ví dụ khi người dùng ít di chuyển, ngồi trong các tòa nhà thì tốc độ dữ liệu người dùng thực tế là quan trọng, còn khi người dùng di chuyển tốc độ cao như tàu cao tốc thì tính di động của người dùng cần được đảm bảo.

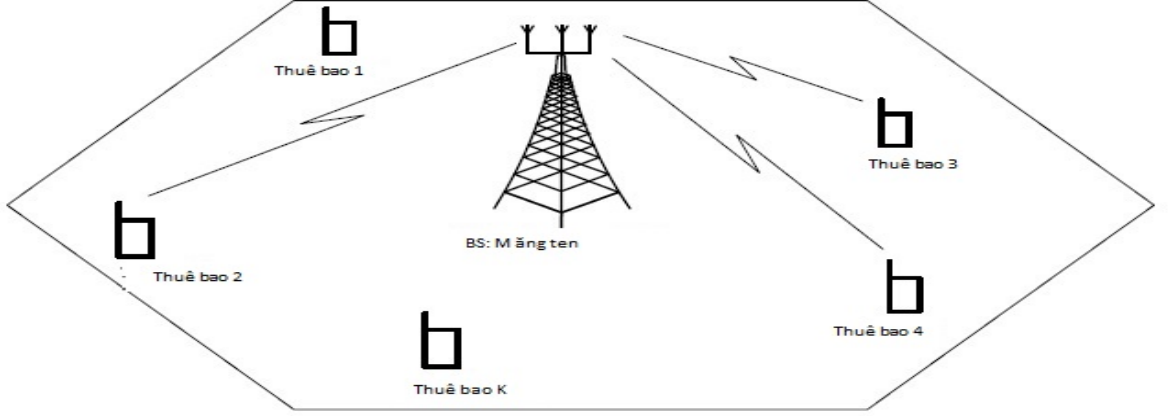
1.3 Hệ thống thông tin di động nhiều ăng ten đa người dùng MU-MIMO

Ở đây xem xét một hệ thống nhiều ăng ten đa người dùng bao gồm một trạm gốc và K thuê bao, mỗi thuê bao chỉ có một ăng ten như trên hình vẽ 1.2 [29]. Trạm gốc có M ăng ten, với $M \geq K$. Tất cả thuê bao sử dụng chung tài nguyên tần số và thời gian. Hệ thống hoạt động theo chế độ Time Division Duplex - ghép kênh phân chia theo thời gian (TDD). Kí hiệu T là độ dài một khung, trong mỗi khung thì τ_p tín hiệu đầu tiên được sử dụng cho việc ước lượng kênh, và $(T - \tau_p)$ tín hiệu còn lại sẽ sử dụng cho việc truyền tín hiệu theo chiều xuống.

Ký hiệu $\mathbf{h}_k[n]$ là véc tơ kênh truyền cỡ $M \times 1$ của thuê bao k , đây là một cột

của ma trận kênh truyền $\mathbf{H}[n]$ có công thức như sau [29]:

$$\mathbf{h}_k[n] = \mathbf{g}_k[n] \sqrt{\beta_k}. \quad (1.2)$$



Hình 1.2: Mô hình hệ thống thông tin nhiều ăng ten đa người dùng

1.3.1 Ước lượng kênh truyền theo chiều lên

Trong một hệ thống TDD, truyền dẫn tín hiệu đường lên và đường xuống dùng chung kênh vô tuyến chỉ khác nhau ở khe thời gian với giả thuyết là các kênh đường lên và đường xuống có tính đối xứng, giống nhau về biên độ và ngược nhau về pha. Để thực hiện việc ước lượng kênh, mỗi thuê bao sẽ truyền một chuỗi tín hiệu hoa tiêu có độ dài τ_p . Chuỗi tín hiệu hoa tiêu của thuê bao k được ký hiệu là $\mathbf{x}_k \in \mathbb{C}^{\tau_p \times 1}$. Các phần tử của chuỗi có độ lớn bằng 1, tức là $\|\mathbf{x}_k\|^2 = \mathbf{x}_k^H \mathbf{x}_k = \tau_p$. Công suất phát đường lên của thuê bao k là ρ_k .

Tín hiệu thu được tại BTS j là [29]

$$\mathbf{y} = \sum_{k=1}^K \sqrt{\rho_k} \mathbf{h}_k \mathbf{x}_k + \mathbf{n}, \quad (1.3)$$

trong đó, nhiễu tại trạm là $\mathbf{n} \approx \mathcal{CN}(0, \sigma_r^2)$ với $\mathbf{n} \in \mathbb{C}^{M \times 1}$.

Hệ thống có thể đạt được dung lượng chiều lên theo công thức 1.4 bằng cách sử dụng kỹ thuật triệt nhiễu, tín hiệu của từng thuê bao được giải mã và được trừ

đi khỏi tổng trước khi giải mã tín hiệu thuê bao tiếp theo như sau [29]:

$$\mathbf{C}_r = \log_2 \det(\mathbf{I}_K + \rho \mathbf{H}^H \mathbf{H}). \quad (1.4)$$

1.3.2 Truyền dữ liệu chiều xuống

Ký hiệu $\mathbf{x} \in \mathbb{C}^{M \times 1}$ là véc tơ tín hiệu cho K thuê bao, và $E\{||x||^2\} = 1$.

Véc tơ tín hiệu nhận được của K thuê bao được viết lại thành

$$\mathbf{Y} = \mathbf{H}^T \mathbf{x} + \mathbf{N}. \quad (1.5)$$

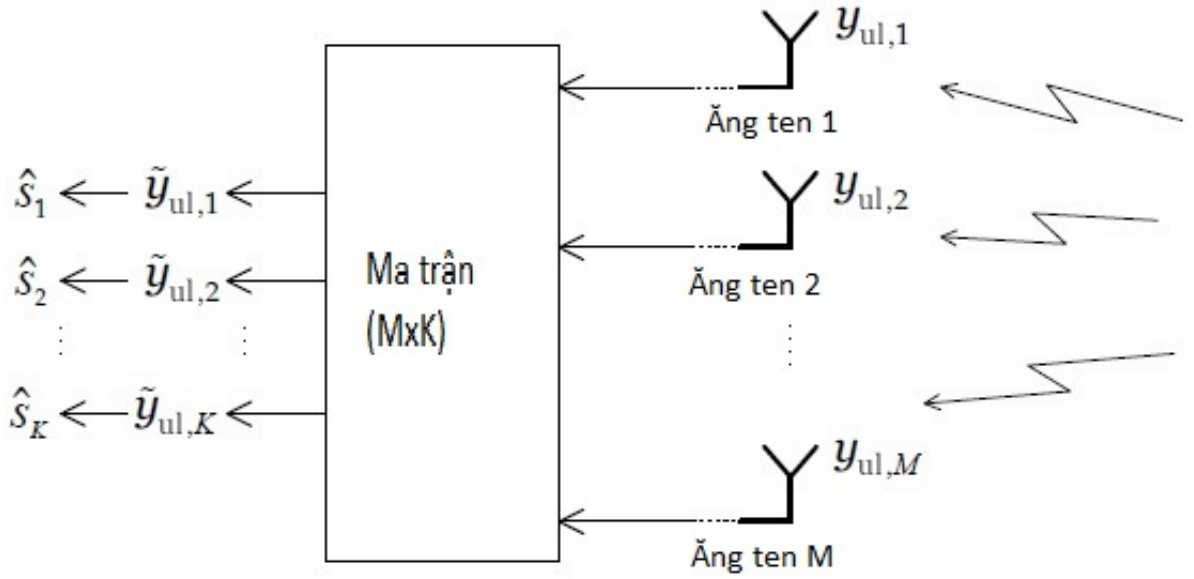
Tốc độ của toàn hệ thống là [29]

$$\mathbf{C}_f = \operatorname{argmax} \log_2 \det(\mathbf{I}_M + \rho \mathbf{H}^* \mathbf{D}_q \mathbf{H}^T). \quad (1.6)$$

Tốc độ này có thể đạt được nhờ sử dụng Dirty Paper Coding - mã hóa thuật toán giấy bẩn (DPC) [29].

1.3.3 Xử lý tuyến tính

Để đạt được hiệu quả tối ưu thì xét việc truyền tín hiệu theo chiều xuống, thuê bao nhận được tín hiệu phải tìm kiếm kết quả trong tất cả các khả năng có thể mà véc tơ \mathbf{x} đã truyền đi. Độ phức tạp của phép toán này tỉ lệ với số ăng ten M , gây khó khăn cho việc giải mã tín hiệu như minh họa tại hình 1.3 [29]. Trạng gốc có thể sử dụng kỹ thuật xử lý tuyến tính (tiền mã hóa trong chiều xuống) để giảm độ phức tạp. Kết quả không đạt được hiệu năng tối ưu tuy nhiên khi số ăng ten tăng lên vô cùng lớn thì kết quả đạt được sẽ tiệm cận với kết quả tối ưu.



Hình 1.3: Mô hình xử lý tuyến tính tại trạm gốc

Tiền mã hóa tuyến tính trong chiều xuống

Ký hiệu $\mathbf{v} \in \mathbb{C}^{K \times 1}$ là véc tơ tín hiệu cho K thuê bao với công suất khác nhau cho từng thuê bao. Tín hiệu của thuê bao k là $v_k = \sqrt{p_k} \times x_k$, và $E\{|x|^2\} = 1$.

Theo kỹ thuật tiền mã hóa chiều xuống, tín hiệu truyền từ M ăng ten (\mathbf{q}) được tiền mã hóa bởi một ma trận $\mathbf{W} \in \mathbb{C}^{M \times K}$ được tính toán từ giá trị kênh truyền $\hat{\mathbf{H}}$ ước lượng như sau:

$$\mathbf{q} = \mathbf{W}\mathbf{v}. \quad (1.7)$$

Công suất được phân bố cho K thuê bao với điều kiện tổng công suất bị giới hạn $\sum_{k=1}^K |\mathbf{w}_k|^2 p_k \leq P$. Véc tơ tín hiệu nhận được của K thuê bao được viết lại thành

$$\mathbf{Y} = \mathbf{H}^T \mathbf{W}\mathbf{v} + \mathbf{N}. \quad (1.8)$$

Tín hiệu nhận được tại thuê bao k là [29]:

$$\begin{aligned}
\mathbf{y}_k &= \mathbf{h}_k^T \mathbf{W} \mathbf{v} + \mathbf{n}_k \\
&= \mathbf{h}_k^T \mathbf{w}_k \mathbf{v}_k + \sum_{l=1, l \neq k}^K \mathbf{h}_k^T \mathbf{w}_l \mathbf{n}_l + \mathbf{n}_k, \\
&= \mathbf{h}_k^T \mathbf{w}_k \sqrt{p_k} x_k + \sum_{l=1, l \neq k}^K \mathbf{h}_k^T \mathbf{w}_l \sqrt{p_l} x_l + \mathbf{n}_k
\end{aligned} \tag{1.9}$$

trong đó \mathbf{h}_k là véc tơ kênh truyền của thuê bao k , và \mathbf{w}_k là cột thứ k ma trận \mathbf{W} .

Tỷ lệ tín hiệu trên nhiễu SINR của thuê bao k được viết là [29]

$$\gamma_k = \frac{|\mathbf{h}_k^T \mathbf{w}_k \sqrt{p_k}|^2}{\sum_{l=1, l \neq k}^K |\mathbf{h}_k^T \mathbf{w}_l \sqrt{p_l}|^2 + \sigma^2}. \tag{1.10}$$

Tốc độ có thể đạt được của thuê bao k là

$$\mathbf{R}_k = \log_2(1 + \gamma_k). \tag{1.11}$$

Tốc độ của toàn hệ thống là

$$\mathbf{R}_{sum} = \sum_{l=1}^K \log_2(1 + \gamma_k). \tag{1.12}$$

1.4 Hệ thống thông tin nhiều ăng ten cỡ rất lớn Massive MIMO

Hệ thống thông tin nhiều ăng ten cỡ rất lớn là một dạng của MU-MIMO trong đó số ăng ten của trạm gốc và số thuê bao phục vụ đều rất lớn. Thông thường, hàng trăm ăng ten sẽ phục vụ vài chục thuê bao. Một số đặc điểm của hệ thống thông tin nhiều ăng ten cỡ rất lớn:

- Hoạt động trong chế độ TDD: trạm gốc cần thông tin về Channel State Information - thông tin trạng thái kênh truyền (CSI) để thực hiện việc tiền mã hóa tín hiệu trước khi truyền cho thuê bao. Nếu hoạt động theo chế độ FDD một trạm gốc có M ăng ten sẽ phải truyền M chuỗi tín hiệu trực giao

cho các thuê bao của mình. Còn nếu hoạt động theo chế độ TDD thì độ dài chuỗi tín hiệu hoa tiêu tỉ lệ với số thuê bao phục vụ K . Vì vậy, hoạt động theo chế độ TDD thường được lựa chọn hơn trong các nghiên cứu gần đây.

- Xử lý tuyến tính: do số lượng ăng ten và thuê bao đều rất lớn nên các phép toán đều thực hiện với ma trận và véc tơ có kích thước lớn cần các kỹ thuật xử lý đơn giản. Vì vậy xử lý tuyến tính ví dụ như tiền mã hóa theo chiều xuống thường được sử dụng.
- Hệ thống thông tin nhiều ăng ten cỡ rất lớn nhưng kích thước vật lý của ăng ten không phải quá lớn, ở tần số 2,6 GHz thì khoảng cách tối thiểu giữa hai ăng ten là nửa bước sóng cỡ 6 cm. Kích thước ăng ten cỡ 28 cm x 29 cm.

1.4.1 Ưu điểm của hệ thống thông tin nhiều ăng ten cỡ rất lớn

Nhu cầu về thông lượng cũng như mật độ thuê bao ngày càng cao trong liên lạc không dây. Hệ thống thông tin nhiều ăng ten cỡ rất lớn có khả năng đáp ứng được yêu cầu này. Xét thông lượng hệ thống theo chiều lên ở điều kiện kênh truyền tối ưu, ta có công thức (1.4) trở thành [29]:

$$\mathbf{C} = \log_2 \det(\mathbf{I}_K + \rho M \mathbf{I}_K) = K \log_2 \det(1 + \rho M), \quad (1.13)$$

trong đó, K là độ lợi về ghép kênh, M là độ lợi mảng ăng ten, ta thấy thông lượng hệ thống có thể tăng lên vô cùng khi K và M tăng lên vô cùng mà không cần thay đổi về công suất. Và trong hệ thống thông tin nhiều ăng ten cỡ rất lớn thì dung lượng hệ thống sẽ đạt gần tối ưu khi áp dụng xử lý tuyến tính và K và M tăng lên vô cùng.

1.4.2 Thách thức của hệ thống thông tin nhiều ăng ten cỡ rất lớn

Bên cạnh các ưu điểm của hệ thống thông tin nhiều ăng ten cỡ rất lớn thì cũng tồn tại rất nhiều thách thức:

- **Can nhiễu trên kênh hoa tiêu**

Do sự giới hạn của tài nguyên thời gian và tần số mà có nhiều tế bào dùng chung tần số và thời gian. Và cũng không thể tạo ra tín hiệu hoa tiêu trực giao cho tất cả các thuê bao trong tất cả tế bào, mà các tín hiệu hoa tiêu trực giao cũng phải được sử dụng lại giữa các tế bào. Vì vậy việc ước lượng kênh truyền của một tế bào có thể bị can nhiễu bởi các tín hiệu hoa tiêu của các thuê bao khác của các tế bào khác. Hiện tượng này được gọi là can nhiễu tín hiệu hoa tiêu, gây suy giảm hiệu năng hệ thống. Hiện tượng can nhiễu kênh hoa tiêu không bị triệt tiêu kể cả khi số lượng ăng ten tiến đến vô cùng.

- **Điều kiện kênh truyền không lý tưởng**

Một ví dụ về điều kiện không lý tưởng của kênh truyền là khi số lượng tín hiệu tán xạ nhỏ hơn số thuê bao thì kênh truyền từ các thuê bao khác nhau sẽ có chung một số đặc điểm về tán xạ, không độc lập hoàn toàn so với nhau. Đây không phải là điều kiện tối ưu của kênh truyền trong hệ thống thông tin nhiều ăng ten cỡ rất lớn. Để giải quyết vấn đề này thì ăng ten của trạm gốc cần được phân tán trong một khu vực diện tích lớn.

- **Tối ưu việc ước lượng kênh**

Do trạm gốc trong mạng thông tin di động 5G có số lượng ăng ten rất lớn và có thể phục vụ hàng chục thuê bao đồng thời nên độ dài chuỗi tín hiệu hoa tiêu và tín hiệu phản hồi từ BTS về thuê bao cũng rất lớn. Cần có các kỹ thuật ước lượng kênh mới để tối ưu việc ước lượng kênh truyền.

1.5 Hệ thống thông tin nhiều ăng ten cỡ rất lớn mmWave Massive MIMO

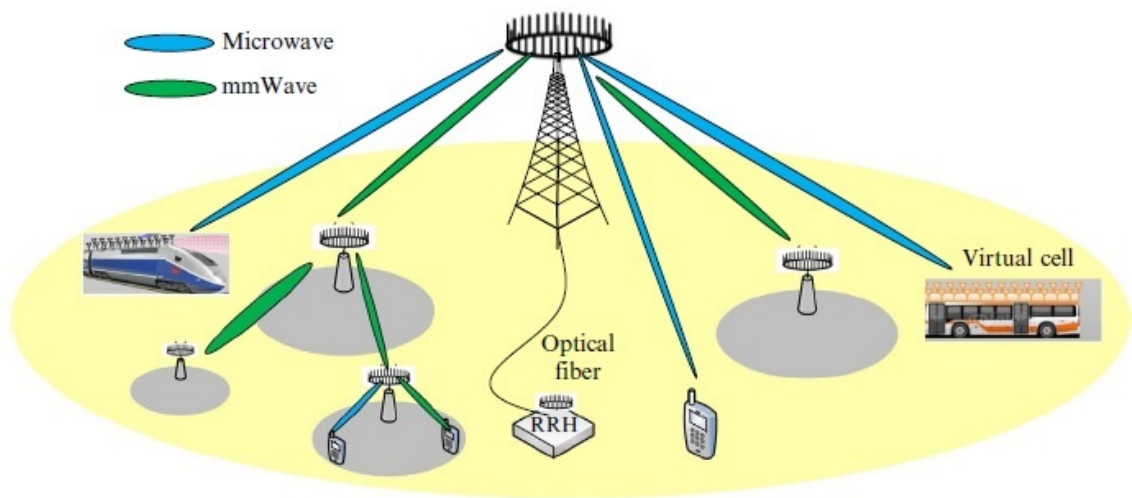
Phần lớn các hệ thống liên lạc hiện nay hoạt động dưới tần số 6 GHz. Tuy nhiên yêu cầu về dung lượng đã khiến cho dải tần số cơ bước sóng cỡ mm tức là từ 30-300 GHz đang được tích cực nghiên cứu. Các đo lường ở tần số 28 và 38 GHz đã cho các kết quả rất tích cực như sau:

- Hiện tượng truyền sóng bằng phương pháp phản xạ có thể thực hiện được nhưng mô hình suy hao đường truyền bị thay đổi.
- Ở cự ly dưới 1km thì suy hao do mưa là nhỏ.
- Phù hợp với truyền thông cự ly ngắn.
- Do bước sóng ngắn nên số lượng ăng ten cỡ rất lớn triển khai được tại trạm gốc và người dùng.

Thực tế là công nghệ bước sóng cỡ mm đã được triển khai trong mạng 802.11ad và mặc dù công nghệ bước sóng cỡ mm rất phù hợp cho truyền điểm tới điểm trong không gian tự do, bước sóng cỡ mm cũng có những điểm hạn chế khi dùng để phủ sóng tế bào như sau:

- Có độ suy hao đường truyền lớn.
- Suy hao khi thâm nhập vật cản lớn.
- Có hiện tượng fading do trời mưa.

Kiến trúc mạng trong hình 1.4 sử dụng cả tần số vi ba và bước sóng mm, và do đặc tính về tần số mà sóng vi ba sẽ được sử dụng ở cự ly xa còn sóng cỡ mm sẽ được dùng trong các phạm vi ngắn [55].



Hình 1.4: Kiến trúc của mạng 5G dựa trên bước sóng cỡ mm

1.5.1 Các thách thức của bước sóng cỡ mmWave Massive MIMO

Bên cạnh các ưu điểm của bước sóng cỡ mm khiến nó tối ưu trong các ứng dụng picocell và femtocell, vẫn còn các rào cản khi chuyển từ công nghệ vi ba sang bước sóng cỡ mm [55] :

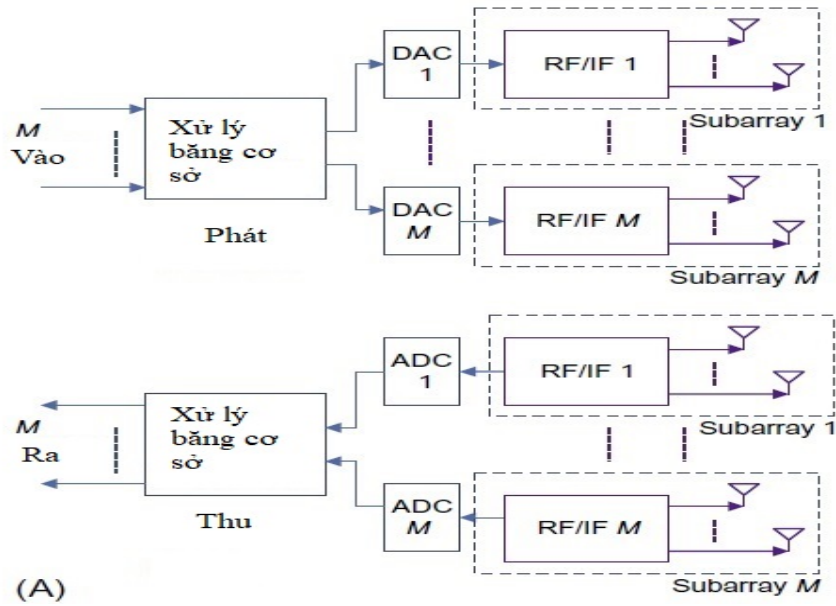
- Đối với sóng vi ba thì nhiễu chủ yếu do các nguồn ở gần và tổng các nguồn nhiễu gọi là nhiễu nền. Với bước sóng cỡ mm thì ảnh hưởng của nhiễu còn do bố trí mẫu ăng ten, không phụ thuộc nhiều vào khoảng cách.
- Tỷ lệ tín hiệu trên nhiễu SINR của sóng vi ba thay đổi chậm từ trung tâm vùng phủ ra đến vùng biên, còn ở sóng cỡ mm thì SINR thặng giáng mạnh phụ thuộc vào có nhận được bức sóng hay không hoặc có bị vật cản hay không.
- Độ phức tạp của xử lý: với bước sóng viba thì số ăng ten thay đổi từ 1-8 ăng ten còn với bước sóng cỡ mm thì dao động từ 8-256 ăng ten vì vậy ước lượng kênh truyền là thách thức rất lớn.
- Chuyển giao tại sóng vi ba thường được thực hiện tại biên giới vùng phủ theo

công suất tín hiệu còn với bước sóng mm thì việc chuyển giao ít liên quan đến vùng phủ mà xảy ra khi có vật chắn, hoặc mật độ kết nối.

- Vấn đề sức khỏe cần được nghiên cứu với bước sóng cỡ mm. Các nghiên cứu cho thấy có hiện tượng nóng lên của mắt và da khi hấp thụ sóng cỡ mm. Tuy nhiên hiện nay các quy định giới hạn tần số trên 100GHz cũng chưa được đề xuất.

1.5.2 Mảng ăng ten cho mmWave Massive MIMO

Hình 1.5 biểu diễn kiến trúc lai giữa số và tương tự cho mảng ăng ten cỡ sóng mm. Trong đó mảng ăng ten có nhiều mảng tương tự con, mà mỗi mảng con bao gồm một bộ RF và N ăng ten [55]. Mỗi mạng con này được nối đến một bộ xử lý tín hiệu bằng cơ sở thông qua bộ chuyển đổi ADC/DAC. Tín hiệu trong mỗi mảng con có thể thay đổi cả về biên độ và pha tùy nhiên để cho đơn giản nhất thì chỉ có pha tín hiệu được thay đổi. Ví dụ với 3 bit lượng tử hóa thì pha có thể có 8 mức trong phạm vi $[-\pi, \pi]$ [55] .

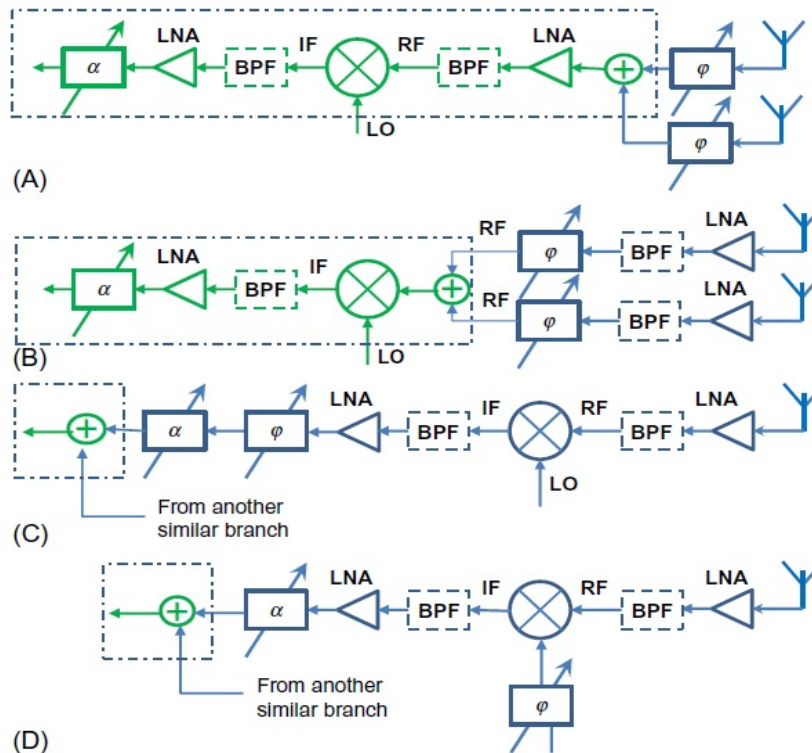


Hình 1.5: Kiến trúc của mảng ăng ten lai bước sóng cỡ mm

1.5.3 Kiến trúc bộ RF

Cách đơn giản nhất để tạo búp sóng tương tự là thông qua một mảng pha, mà trong đó độ trễ pha của tín hiệu mỗi ăng ten được điều khiển hoàn toàn. Cụ thể hơn là với các hệ thống tín hiệu băng hẹp dải tần dưới 40 GHz thì triển khai các bộ trễ đóng vai trò là các bộ dịch pha ở tần số trung tâm là một giải pháp có thể sử dụng. Với các tần số cao hơn thì hiện nay chưa có giải pháp cụ thể do lỗi lớn ở tần số cao. Hình 1.6 minh họa một số kiến trúc của bộ RF. Trong đó khối φ và α là các khối dịch pha và suy hao biên độ. Khối bao quanh bởi nét đứt là khối có thể dùng chung trong mảng con. Các Bandpass Filter - bộ lọc thông dải (BPF) có thể được sử dụng để giới hạn băng tần IF sau khi biến đổi tần số. Hình 1.6 A biểu diễn cấu trúc đơn giản nhất với chỉ có ăng ten và bộ dịch pha là các thành phần độc lập, còn các thành phần còn lại đều dùng chung với các phần tử khác trong mảng con. Hình 1.6 B cải thiện hơn với một Low Noise Amplifier - bộ khuếch đại nhiễu thấp (LNA) được dành riêng cho mỗi ăng ten giúp cải thiện độ nhạy tín

hiệu thu được. Hình 1.6 C và D thì có ý nghĩa thực tế hơn cả với việc triển khai bộ dịch pha tại bộ phận IF và LO [55] .



Hình 1.6: Kiến trúc của bộ RF

1.5.4 Ước lượng kênh truyền trong hệ thống mmWave Massive MIMO

Giải pháp tiền mã hóa (còn gọi là tạo búp sóng) truyền thống sử dụng kỹ thuật tương tự trong bước sóng cỡ mm được thực hiện bằng một mạng tương tự ví dụ như là mạng dịch pha với một số lượng nhỏ các bộ Radio Frequency - tần số vô tuyến điện (RF) có cùng biên độ. Cách này đơn giản để triển khai nhưng hiệu quả thu được khá thấp. Ngược lại, nếu chỉ sử dụng tiền mã hóa kỹ thuật số ở băng cơ sở như sử dụng mã hóa DPC thì có thể điều khiển được cả biên độ và pha, từ đó đạt được hiệu quả tối ưu. Tuy nhiên nếu sử dụng tiền mã hóa kỹ thuật số lại đòi hỏi mỗi ăng ten có một bộ RF riêng. Do số lượng ăng ten lớn nên dẫn đến chi phí phần cứng và công suất tiêu thụ điện sẽ rất lớn. Vì vậy, tiền mã hóa sử dụng cả kĩ

thuật tương tự và số được nghiên cứu và đề xuất. Trong đó, một lượng nhỏ tiền mã hóa kỹ thuật số (chỉ đòi hỏi một ít bộ RF) được sử dụng để triệt tiêu xuyên nhiễu. Sau đó một mạng tương tự dịch pha cỡ lớn (không đòi hỏi các bộ RF) được sử dụng để tăng độ lợi của ăng ten.

1.5.5 Mô hình kênh truyền

Các thực nghiệm mô tả kênh truyền mmWave Massive MIMO thể hiện một sự thừa thớt trong cả miền không gian và góc. Đây là do sự tổn hao lớn của các kênh Non-Line-of-Sight - truyền sóng không trong tầm nhìn thẳng (N-LOS), nên sẽ chỉ còn một số ít các đường (thường từ 3 đến 5 đường) thể hiện sự đa đường của mmWave Massive MIMO. Ở đây ta xem xét ăng ten là Uniform Linear Array - mảng tuyến tính đồng nhất (ULA), kênh truyền điểm tới điểm sẽ được biểu diễn như sau [7]:

$$\mathbf{H} = \sqrt{\frac{M_T M_R}{L}} \sum_{l=1}^L \alpha_l a_R(\theta_l) a_T^H(\phi_l) \quad (1.14)$$

$$= \sqrt{\frac{M_T M_R}{L}} \mathbf{A}_R \mathbf{H}_a \mathbf{A}_T^H, \quad (1.15)$$

$$\mathbf{A}_R = [a_R(\theta_1), a_R(\theta_2), \dots, a_R(\theta_L)], \quad (1.16)$$

$$\mathbf{A}_T = [a_T(\phi_1), a_T(\phi_2), \dots, a_T(\phi_L)], \quad (1.17)$$

$$\mathbf{H} = \text{diag}\{\alpha_1, \alpha_2, \dots, \alpha_L\}, \quad (1.18)$$

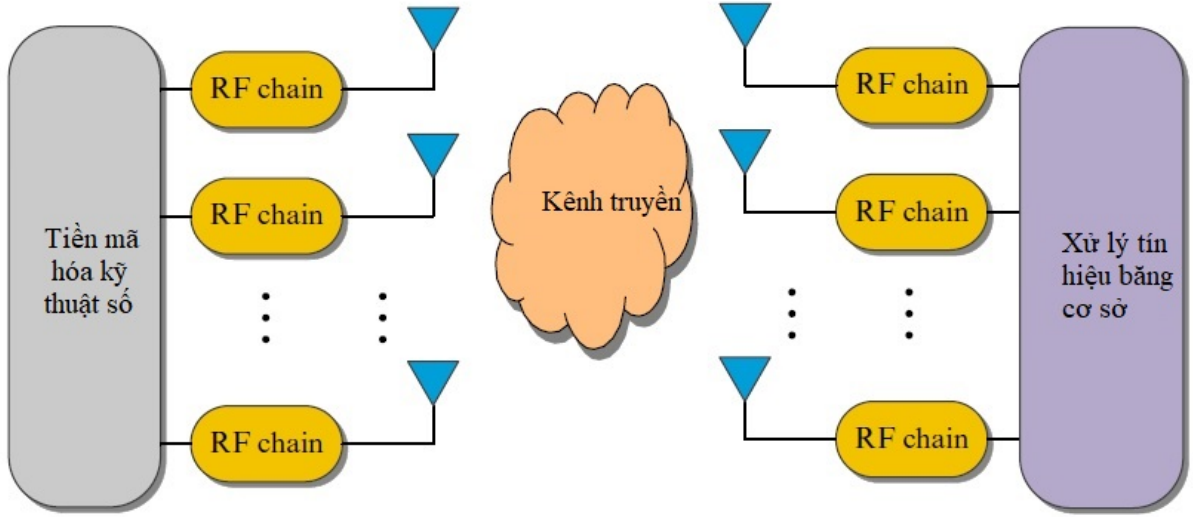
trong đó, N_T và N_R là số lượng ăng ten phát và thu, L là số lượng đa đường, α_l là độ lợi của đường thứ l , θ_l và ϕ_l là góc phương vị của góc tới và góc phản xạ.

1.5.6 Các giải pháp tiền mã hóa

- Tiền mã hóa kỹ thuật số

Ở đây, xem xét hệ thống bước sóng cỡ mm với một thuê bao sử dụng tiền

mã hóa kỹ thuật số như hình 1.7 [55]. Trong đó trạm gốc với M_T ăng ten, truyền đồng thời M_R luồng dữ liệu đồng thời tới M_R ăng ten của thuê bao, với $M_T \geq M_R$.



Hình 1.7: Cấu trúc tiền mã hóa kỹ thuật số cho một thuê bao của hệ thống mmWave

Tín hiệu nhận được tại thuê bao [55]:

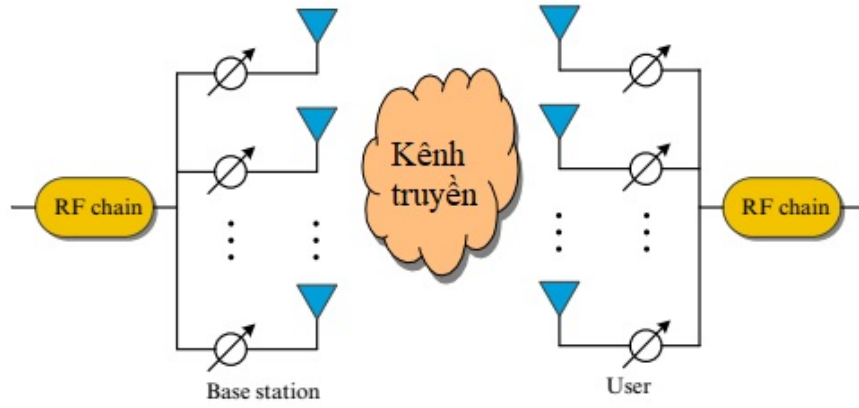
$$\mathbf{Y} = \sqrt{\rho}\mathbf{H}\mathbf{D}\mathbf{s} + \mathbf{n}, \quad (1.19)$$

trong đó, $\mathbf{H} \in \mathbb{C}^{M_R \times M_T}$ là ma trận kênh truyền, $\mathbf{D} \in \mathbb{C}^{M_T \times M_R}$ là ma trận tiền mã hóa.

- **Tạo búp sóng kỹ thuật tương tự**

Ở đây cả phía phát và phía thu chỉ sử dụng một bộ RF, mục tiêu của hệ thống là thiết kế các véc tơ mã hóa sao cho tỷ lệ SNR là cao nhất, như trên hình 1.8 [55]

$$(\mathbf{w}^{opt}, \mathbf{f}^{opt}) = \operatorname{argmax} \|\mathbf{w}^H \mathbf{H} \mathbf{f}\|^2. \quad (1.20)$$



Hình 1.8: Cấu trúc tạo búp sóng tương tự

- Tạo búp sóng lai giữa tương tự và số

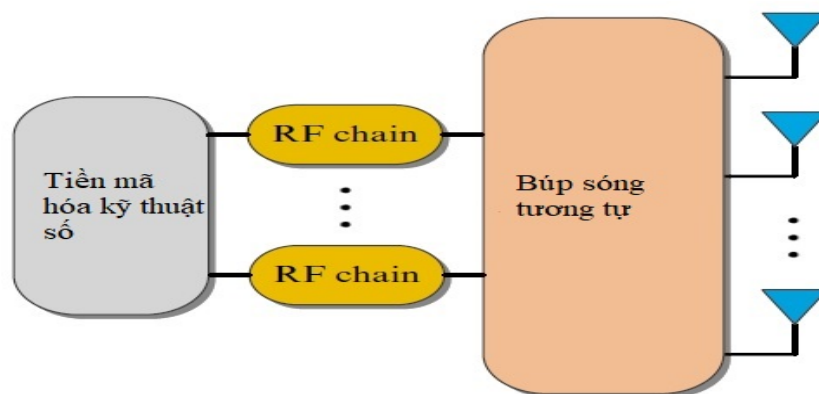
Ở đây, hệ thống bước sóng cỡ mm với một thuê bao sử dụng tiền hóa bao gồm cả kỹ thuật tương tự và kỹ thuật số như trên hình 1.9 [55]. Trong đó trạm gốc với M_T ăng ten và có M_T^{RF} bộ RF, truyền M_S luồng dữ liệu đồng thời tới M_R ăng ten của thuê bao, với $M_T \geq M_T^{RF} \geq M_R$. Tín hiệu truyền đi là [55]

$$\mathbf{X} = \mathbf{A}\mathbf{D}\mathbf{s}. \quad (1.21)$$

Tín hiệu thu được tại thuê bao được biểu diễn bởi

$$\mathbf{Y} = \sqrt{\rho}\mathbf{H}\mathbf{A}\mathbf{D}\mathbf{s} + \mathbf{n}. \quad (1.22)$$

Ở đây giả thiết việc ước lượng kênh truyền là hoàn hảo.

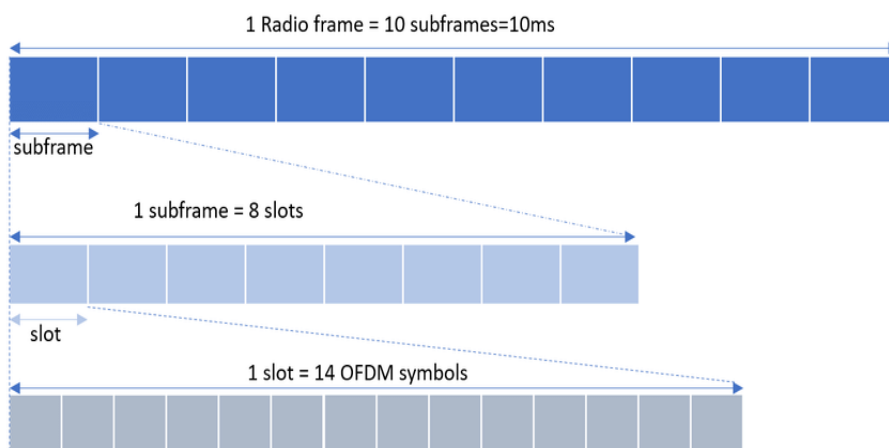


Hình 1.9: Cấu trúc tạo búp sóng lai cho hệ thống một thuê bao

1.6 Kỹ thuật lập lịch

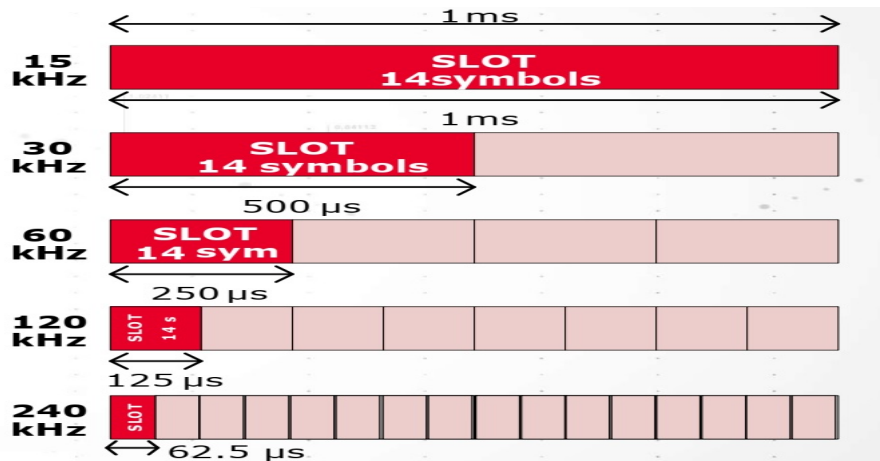
1.6.1 Cấu trúc khung của mạng vô tuyến

Một chu kỳ khung của mạng New Radio - Mạng vô tuyến thế hệ mới (NR) kéo dài 10ms, chia làm 10 khung con như hình 1.10. Khoảng cách giữa các sóng mang của mạng 5G thay đổi từ 15kHz, 30 kHz,... đến 480 kHz.



Hình 1.10: Cấu trúc khung của mạng vô tuyến NR

Tùy vào khoảng cách sóng mang mà một khung con có thể có một hoặc nhiều slot như hình 1.11. Mỗi slot bao gồm 14 tín hiệu Orthogonal Frequency Division Multiplexing - ghép kênh phân chia theo tần số trực giao (OFDM).



Hình 1.11: Cấu trúc khung con mạng vô tuyến NR

1.6.2 Kỹ thuật lập lịch ở mạng vô tuyến

Xét mạng thông tin di động 5G gồm một trạm gốc gNB và các thuê bao trong vùng phủ của nó. Trong mỗi khung thời gian mà chất lượng kênh hầu như không đổi, gNB sẽ chọn ra một tập con thuê bao để gửi tín hiệu hoa tiêu, và sau đó tùy theo thông tin về CSI, một tập con thuê bao sẽ được gNB phục vụ. Toàn bộ quá trình này được gọi là lập lịch. Có rất nhiều kỹ thuật được xây dựng dựa trên các tiêu chí khác nhau như độ trễ, sự công bằng giữa các thuê bao, hoặc tối ưu tốc độ toàn hệ thống... Các kỹ thuật với mục tiêu khác nhau dẫn đến sự khác nhau đáng kể về hiệu suất của hệ thống và khác biệt về chất lượng dịch vụ của thuê bao. Vì vậy, vấn đề lập lịch luôn là một vấn đề quan trọng thu hút được nhiều sự quan tâm và nghiên cứu ở cả trong và ngoài nước.

1.6.3 Thách thức

Các thách thức của quá trình lập lịch tại gNB:

- **Số lượng thuê bao rất lớn**

Trong hệ thống TDD Massive MIMO, thì dung lượng kênh phụ thuộc vào độ chính xác về thông tin trạng thái kênh truyền. Điều này yêu cầu độ dài chuỗi

hoa tiêu bằng với số lượng thuê bao. Vì vậy, nếu tăng số lượng thuê bao phục vụ sẽ làm tăng thời gian ước lượng kênh và làm giảm thời gian truyền dữ liệu tức là giảm dung lượng hệ thống. Như vậy việc ước lượng kênh cho số lượng nhiều thuê bao trong Massive MIMO là một thách thức cần phải đặt ra.

- **Đảm bảo chất lượng dịch vụ QoS**

Các thuê bao của mạng 5G sẽ có các yêu cầu về QoS khác nhau và kỹ thuật lập lịch cần phải đáp ứng tối đa các yêu cầu của từng thuê bao. Các yêu cầu này thậm chí xung đột với mục tiêu tối đa dung lượng hệ thống mà kỹ thuật lập lịch cần phải hướng tới. Ví dụ, việc phục vụ các thuê bao mong muốn độ trễ thấp nhưng có độ lợi kênh truyền thấp sẽ phải trả giá bằng dung lượng hệ thống bị giảm đi. Vì vậy kỹ thuật lập lịch được lựa chọn triển khai trên thực tế cần phải đáp ứng nhiều yêu cầu khác nhau về QoS, và dung hòa tốt nhất các sự xung đột về lợi ích nếu có.

- **Độ phức tạp của kỹ thuật lập lịch**

Cuối cùng là do số lượng ăng ten rất lớn nên bài toán tối ưu thường có độ phức tạp cao cần có các thuật toán tối ưu giảm thiểu độ phức tạp, từ đó trở nên khả thi khi áp dụng vào thực tế triển khai.

- **Tương thích với kỹ thuật tiền mã hóa**

Kỹ thuật lập lịch cần phải phù hợp với kỹ thuật tiền mã hóa. Khi hệ thống sử dụng kỹ thuật tiền mã hóa MRT thì nên chọn các thuê bao có độ lợi kênh truyền cao nhất. Ngược lại khi hệ thống sử dụng tiền mã hóa ZF thì cần lựa chọn tập thuê bao trực giao để ít can nhiễu lẫn nhau. Từ đó hệ thống sẽ đạt được dung lượng tối ưu nhất.

1.6.4 Một số kỹ thuật lập lịch

- **Kỹ thuật Proportional Fair - công bằng giữa các thuê bao (PF)**

Kỹ thuật lập lịch này hướng tới mục tiêu đảm bảo tốc độ sử dụng dịch vụ của các thuê bao là gần bằng nhau. Thuật toán dựa trên một đại lượng đo M , và luôn chọn các thuê bao có giá trị M cao nhất [3].

$$M = \operatorname{argmax} \frac{R_i(t)}{\bar{R}_i(t)}, \quad (1.23)$$

trong đó, $R_i(t)$ là tốc độ có thể đạt được hiện thời của thuê bao còn $\bar{R}_i(t)$ là tốc độ dữ liệu trung bình đã sử dụng cho đến thời điểm t .

Mục tiêu của kỹ thuật này là tìm ra thuê bao đã được phục vụ ít nhất và có tốc độ có thể đạt được hiện thời cao nhất. Hạn chế của kỹ thuật này là dung lượng tổng hệ thống đạt được khá thấp so với các kỹ thuật lập lịch khác.

- **Kỹ thuật Maximum Rate - tối đa tốc độ (MR)**

Mục tiêu của kỹ thuật này là tăng tốc độ của cả hệ thống bằng cách phát công suất $p_k(m, f)$ lớn cho các thuê bao có chất lượng kênh truyền cao [57], [59]:

$$p_k(m, f) = \begin{cases} \left[\frac{1}{\lambda_k} - \frac{\mathbf{N}_0}{\|\mathbf{H}_k(m, f)\|^2} \right]^+, & \text{if } \|\mathbf{H}_k(m, f)\|^2 \geq \frac{\lambda_k}{\lambda_l} \|\mathbf{H}_l(m, f)\|^2 \\ 0 & \text{cho các trường hợp khác,} \end{cases} \quad (1.24)$$

trong đó $[x]^+ = \max(0, x)$, $\mathbf{H}_k(m, f)$ là chất lượng kênh của thuê bao k tại RB m của chu kỳ khung f và λ_k là hệ số đảm bảo sự giới hạn về công suất

$$\sum_{k=1}^K p_k \leq P. \quad (1.25)$$

Hạn chế của kỹ thuật lập lịch này khó áp dụng trong các hệ thống ăng ten cỡ rất lớn do mất nhiều thời gian tìm ra thuê bao có dung lượng cao nhất để phục vụ.

- **Kỹ thuật lập lịch đảm bảo Quality of Service - chất lượng dịch vụ (QoS)**

Trong nghiên cứu [8] các tác giả quan tâm đến việc đảm bảo độ trễ tối thiểu cho các thuê bao sử dụng dịch vụ nhạy cảm với trễ như dịch vụ thoại.

$$P_k[n, s] = (1 + f(w_k)) \frac{R_k[n, s]}{[r_k[n]]^\alpha}, \quad (1.26)$$

trong đó, $f(w_k) = \frac{1}{1 + e^{-a_k(w_k - D)}}$,

α là tham số để điều chỉnh độ công bằng,

w_k thể hiện độ trễ của thuê bao k ,

Tham số a_k điều chỉnh sườn dốc của hàm sigmoid,

Tham số c xác định cận trên của độ trễ.

Hạn chế của kỹ thuật này là chưa quan tâm đến các tham số khác và thậm chí là quan trọng hơn của QoS như độ ưu tiên, hay tốc độ tối thiểu của thuê bao.

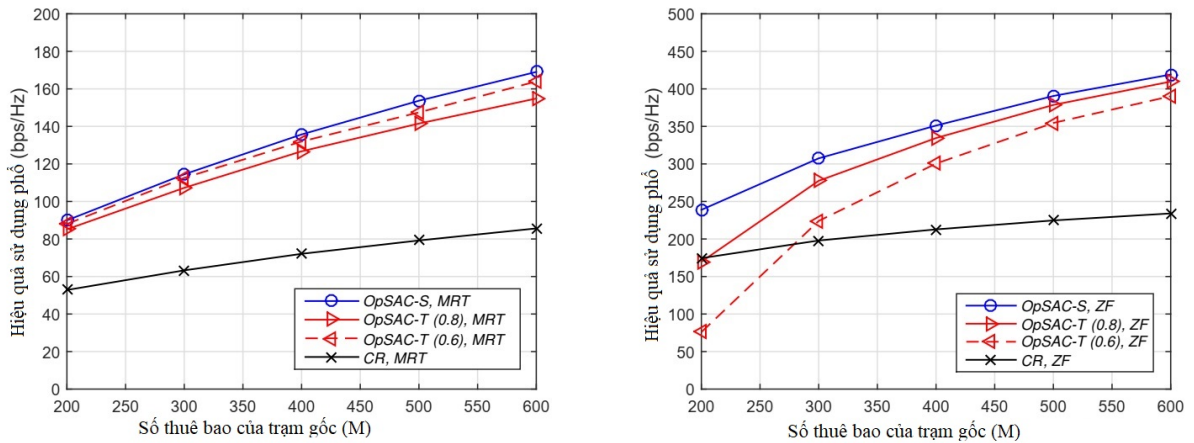
- **Kỹ thuật lập lịch sử dụng thông tin lão hóa kênh truyền**

Tác giả trong bài báo [40] đã đề xuất một kỹ thuật lập lịch sử dụng thông tin lão hóa kênh truyền để tăng độ lợi về ghép kênh không gian bằng cách phục vụ thêm nhiều thuê bao mà không cần tiêu tốn tài nguyên để ước lượng kênh. Kênh truyền thuê bao tại chu kỳ khung hiện tại có thể được ước lượng từ trạng thái của chu kỳ khung trước đó:

$$\mathbf{h}_k[n] = \alpha_k \mathbf{h}_k[n-1] + \mathbf{e}_k[n], \quad (1.27)$$

trong đó α_k là hệ số tương quan tạm thời của thuê bao k , $\mathbf{h}_k[n-1]$ là véc tơ hệ số kênh truyền của chu kỳ khung trước đó của thuê bao k , và $\mathbf{e}_k[n]$ là hệ số kênh truyền không tương quan.

Hình 1.12 so sánh hiệu quả sử dụng phổ khi sử dụng các ngưỡng về độ lão hóa kênh truyền khác nhau với kỹ thuật lập lịch truyền thống là Round Robin [40]. Kết quả mô phỏng cho thấy kỹ thuật lập lịch sử dụng thông tin lão hóa kênh truyền có hiệu quả sử dụng phổ cao hơn kỹ thuật lập lịch Round Robin trong cả hai phương pháp tiền mã hóa MRT và ZF. Điểm chú ý là hiệu quả sử dụng của kỹ thuật đề xuất sẽ kém hơn kỹ thuật truyền thống khi số lượng ăng ten sử dụng thấp. Khi đó sai số do sự sai lệch của thông tin ước lượng kênh dẫn đến hiệu quả sử dụng của kỹ thuật đề xuất thấp.



Hình 1.12: So sánh hiệu quả sử dụng phổ

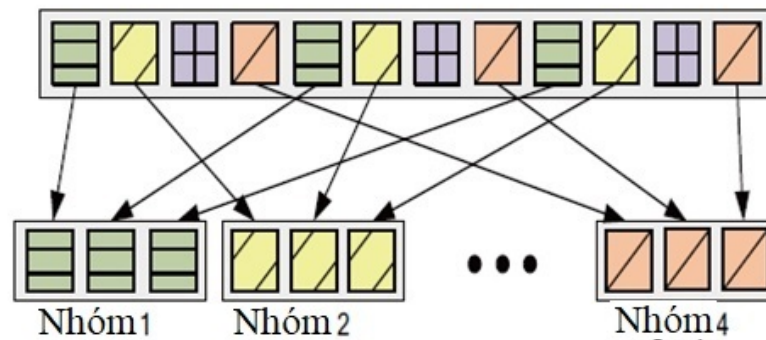
Hạn chế của kỹ thuật lập lịch này chưa xác định được số lượng thuê bao tối ưu khi lập lịch nên sẽ khó đảm bảo chất lượng dịch vụ cho thuê bao.

- **Kỹ thuật lập lịch phân nhóm ăng ten [37]**

Tại bài báo này tác giả nghiên cứu hệ thống FDD Massive MIMO với đặc tính thời gian ước lượng kênh bằng số lượng ăng ten tại trạm gốc. Hệ thống FDD Massive MIMO có một số ưu việt hơn hệ thống TDD như: độ trễ nhỏ, ước lượng kênh liên tục và có thể thích hợp với các hệ thống trước đây. Do số lượng ăng ten trong hệ thống Massive MIMO là rất lớn, cần thiết phải giảm

độ dài của chuỗi tín hiệu hoa tiêu. Và một ràng buộc nữa của bài báo là hệ thống sử dụng hồi đáp giới hạn (limited-feedback) tức là thông tin về trạng thái kênh truyền sẽ được lượng hóa bằng một từ mã trước khi truyền về trạm gốc.

Nhóm nghiên cứu đã đề xuất mỗi thuê bao sẽ được phục vụ bởi một nhóm các ăng ten. Như vậy bên cạnh thông tin truyền về là chỉ số của từ mã thì cần truyền thêm chỉ số của nhóm ăng ten. Nhóm nghiên cứu đề xuất kỹ thuật lập lịch có tên là AGS kết hợp giữa lựa chọn ăng ten và lập lịch để giảm lượng thông tin hồi đáp. Theo thuật toán này, đầu tiên các nhóm ăng ten sẽ được định nghĩa trước như trong hình 1.13 và mỗi thuê bao chọn ra nhóm tối ưu nhất cho mình dựa trên độ lợi kênh của cả nhóm [37]. Sau đó tiến hành lượng hóa véc tơ thông tin kênh truyền cho nhóm ăng ten đã chọn. Thông tin bao gồm từ mã và chỉ số nhóm ăng ten sẽ được gửi về trạm gốc. Dựa trên thông tin này thì trạm gốc sẽ ước lượng được tốc độ của mỗi nhóm ăng ten từ đó chọn ra các nhóm được sử dụng để tối ưu tốc độ.



Hình 1.13: Nhóm các ăng ten sử dụng

Kết quả mô phỏng cho thấy kỹ thuật AGS có dung lượng tổng cao hơn cách truyền thống khi không nhóm các ăng ten để tối ưu thông tin phản hồi.

- **Kỹ thuật lập lịch phân nhóm thuê bao**

Trong bài báo [63] thì tác giả Xu và cộng sự cũng nghiên cứu hệ thống FDD Massive MIMO dưới tác động của hiện tượng fading. Nhóm đề xuất kỹ thuật lập lịch chia làm hai bước: bước một tiến hành nhóm các thuê bao có cùng điều kiện kênh truyền, bước hai tiến hành mã hóa dựa theo chất lượng kênh truyền của nhóm. Nhóm nghiên cứu sử dụng mô hình một trạm gốc có M ăng ten phục vụ K thuê bao, mỗi thuê bao sử dụng một ăng ten. Trạm gốc sử dụng mảng ăng ten hai chiều ULA. Kết quả cho thấy khi số lượng thuê bao càng nhiều thì cách nhóm thuê bao càng hiệu quả [63].

- **Kỹ thuật lập lịch sử dụng tính trực giao của kênh truyền**

Trong bài báo [23], các tác giả đã nghiên cứu ảnh hưởng của việc sử dụng bộ ADC phân giải thấp lên hệ thống mmWave theo đường lên. Việc này sẽ gây sai số cho quá trình lượng tử hóa tín hiệu ở đầu ra của bộ tạo búp sóng tương tự. Tác giả đã tính toán độ trực giao giữa các kênh truyền của thuê bao xét tới yếu tố sử dụng bộ ADC phân giải thấp. Từ đó đề xuất kỹ thuật lập lịch nhằm loại bỏ các thuê bao trong tập ứng cử viên mà không trực giao với tập thuê bao lập lịch đã chọn. Kết quả mô phỏng chỉ rõ, dung lượng tổng hệ thống có thể tiệm cận với dung lượng tổng của kỹ thuật MR [23]. Đây là kết quả rất quan trọng nhằm thúc đẩy các nghiên cứu để xây dựng kỹ thuật lập lịch trong các hệ thống khác sử dụng tính trực giao của kênh truyền.

1.7 Kết luận chương 1

Trong chương này, luận án đã trình bày những cơ sở lý thuyết nền tảng của hệ thống thông tin di động 5G nói chung và một số kỹ thuật lập lịch tiêu biểu hiện nay nói riêng cùng với hạn chế của các kỹ thuật này. Trên cơ sở đó, luận án đề ra hướng nghiên cứu về kỹ thuật lập lịch gồm ba phần: (i) thứ nhất, luận án nghiên

cứ kỹ thuật lập lịch sử dụng thông tin lão hóa kênh truyền để làm tăng dung lượng toàn hệ thống và đảm bảo mọi thuê bao thỏa mãn yêu cầu về tốc độ tối thiểu, (ii) thứ hai, luận án phát triển kỹ thuật lập lịch có thể đảm bảo cho các thuê bao có thể có nhiều tốc độ tối thiểu khác nhau, cũng như sử dụng các tham số về độ ưu tiên hoặc độ rút gói, thay vì chỉ dừng lại ở tham số về độ trễ như các nghiên cứu hiện nay, (iii) thứ ba là luận án nghiên cứu kỹ thuật lập lịch sử dụng đặc tính trực giao kênh truyền của mạng thông tin di động cỡ rất lớn để giảm thời gian xác định tập thuê bao phục vụ.

Chương 2

Kỹ thuật lập lịch đảm bảo tốc độ tối thiểu của thuê bao dựa trên thông tin lão hóa kênh truyền

Trong chương này luận án sẽ trình bày mô hình hệ thống về vấn đề lão hóa kênh truyền, kỹ thuật ước lượng kênh truyền lão hóa và tốc độ có thể đạt được khi sử dụng hai phương pháp tiền mã hóa là MRT và ZF. Từ đó đề xuất kỹ thuật lập lịch sao cho việc sử dụng thông tin lão hóa kênh truyền vẫn đảm bảo được tốc độ tối thiểu cho thuê bao, hơn nữa là không ảnh hưởng đến chất lượng dịch vụ của các thuê bao khác. Đây là điểm khác biệt so với kỹ thuật lập lịch sử dụng thông tin lão hóa kênh truyền ở [40].

2.1 Ý tưởng kỹ thuật lập lịch đảm bảo tốc độ tối thiểu cho thuê bao khi sử dụng thông tin lão hóa kênh truyền

Hệ thống thông tin di động Multiple Input Multiple Output - nhiều ăng ten thu nhiều ăng ten phát (MIMO) cỡ rất lớn là một kỹ thuật mang tính quyết định để đạt được dung lượng lớn cho các mạng thông tin di động 5G. Ý tưởng của kỹ thuật này là sử dụng số lượng lớn ăng ten tại trạm gốc để phục vụ nhiều thuê bao đồng thời, mà trong đó mỗi thuê bao chỉ có vài ăng ten. Thông thường, mỗi trạm

gốc sẽ có hàng trăm ăng ten để cung cấp kết nối cho vài chục thuê bao với mỗi thuê bao chỉ sử dụng một ăng ten. Dung lượng hệ thống thông tin di động nhiều ăng ten cỡ rất lớn tỉ lệ thuận với số lượng ăng ten tại trạm gốc và tại thuê bao [53]. Dung lượng tối ưu có thể đạt được khi kỹ thuật mã hóa DPC được sử dụng [25]. Hơn nữa sử dụng các kỹ thuật tiền mã hóa tuyến tính đơn giản như MRT hoặc ZF cũng có thể đạt được hiệu năng tối ưu khi số ăng ten của trạm gốc rất lớn [30, 34]. Mặc dù các ưu điểm của hệ thống MIMO cỡ rất lớn là rất rõ ràng nhưng nó vẫn còn các thách thức phải xử lý nếu muốn đạt hiệu suất tối ưu.

Để đạt được dung lượng lớn của hệ thống đòi hỏi trạm gốc phải có được CSI chính xác để thực hiện tiền mã hóa trước khi truyền dữ liệu. Ở chế độ Frequency Division Duplex - ghép kênh phân chia theo tần số (FDD) thì lượng dữ liệu hoa tiêu tỉ lệ với số lượng ăng ten tại trạm gốc, trong khi đó ở chế độ TDD thì lượng dữ liệu hoa tiêu tỉ lệ với số lượng thuê bao. Vì vậy nhiều nghiên cứu đã sử dụng chế độ hoạt động TDD trong nghiên cứu của mình [46, 43, 64, 36]. Trong chế độ TDD, thông tin trạng thái kênh truyền CSI được ước lượng trong giai đoạn truyền dẫn đường lên. Trạm gốc sẽ lựa chọn một tập thuê bao có mặt trong tế bào để phục vụ trong giai đoạn truyền dẫn đường xuống. Quá trình này gọi là lập lịch và hiện nay có rất nhiều kỹ thuật lập lịch khác nhau hướng đến các mục tiêu khác nhau.

Để tăng tốc độ của thuê bao cũng như tốc độ tổng của toàn hệ thống, quá trình lập lịch sẽ lựa chọn các thuê bao có độ lợi kênh truyền tốt nhất và yêu cầu công suất nhỏ nhất trong chu kỳ khung để phục vụ [57, 17, 5]. Kỹ thuật này gọi là tối đa hóa tốc độ. Nhược điểm của kỹ thuật này là các thuê bao có độ lợi kênh truyền kém sẽ gần như không được phục vụ và điều này thì rất khó chấp nhận khi nhìn từ góc độ của người dùng. Một kỹ thuật khác để tối ưu tốc độ hệ thống là sử dụng

điều khiển công suất, theo đó thì công suất phát cho các thuê bao có độ lợi kênh truyền kém sẽ được tăng lên và công suất của các thuê bao có độ lợi kênh truyền tốt sẽ giảm đi [65]. Một bài báo khác thì quan tâm về tổng công suất sử dụng khi nghiên cứu về hiệu quả sử dụng năng lượng và phổ tần số hơn là tối ưu công suất của từng thuê bao cụ thể [47]. Để giải quyết nhược điểm của phương pháp tối đa hóa tốc độ thì phương pháp công bằng theo tỷ lệ tiến hành xem xét lưu lượng đã sử dụng trong quá khứ của thuê bao để cố gắng tạo ra một tốc độ bằng nhau giữa các thuê bao. Mục tiêu của phương pháp này là duy trì một sự công bằng giữa các thuê bao [3].

Bởi vì số lượng thuê bao được ước lượng kênh là có giới hạn khi so sánh với số lượng ăng ten để tối ưu hiệu quả sử dụng phổ tần [42, 15, 12] nên số lượng thuê bao được phục vụ cũng bị hạn chế. Hơn nữa, việc nhóm các thuê bao và nhóm các ăng ten cũng được cân nhắc để cải thiện hiệu quả sử dụng phổ tần [37, 63, 13]. Các tác giả trong [66, 58] thì tìm kiếm việc tiết kiệm tài nguyên nhờ sử dụng lại các chuỗi tín hiệu hoa tiêu. Một cách khác để tăng dung lượng của hệ thống là phục vụ thêm các thuê bao mà không cần tốn thêm tài nguyên cho việc ước lượng kênh của các thuê bao này. Thông tin CSI hiện thời của các thuê bao này được ước lượng bằng lượng kênh truyền biến thiên so với thông tin CSI được lưu giữ trước đó. Hiện tượng này được gọi là lão hóa kênh truyền và đã thu hút được rất nhiều sự quan tâm của giới nghiên cứu [60, 49, 33, 35].

Trong [40], tác giả đề xuất một kỹ thuật lập lịch sử dụng thông tin lão hóa kênh để tăng độ lợi ghép kênh không gian bằng cách phục vụ thêm nhiều thuê bao mà không cần tăng thông tin ước lượng kênh. Bên cạnh đó, lớp vật lý của mạng vô tuyến 5G sẽ hỗ trợ QoS để đảm bảo chất lượng dịch vụ từ mạng lõi tới tận thuê bao [1]. Đặc biệt, trạm gốc gNB sẽ phải đảm bảo tốc độ tối thiểu cho các luồng

dữ liệu có Guaranteed Flow Bit Rate - tốc độ luồng bit được đảm bảo (GFBR) để phục vụ các kết nối có Guaranteed Bit Rate- tốc độ bit được đảm bảo (GBR) [2].

Từ các yếu tố này, luận án đề xuất một kỹ thuật lập lịch sử dụng thông tin lão hóa kênh truyền, gọi là QoS-Aware, kỹ thuật khai thác thông tin lão hóa kênh không chỉ nhằm mục tiêu tăng dung lượng hệ thống mà còn hỗ trợ QoS bằng việc đảm bảo rằng mọi thuê bao sẽ được đảm bảo một tốc độ tối thiểu GFBR. Hiệu năng của QoS-Aware được đánh giá về mặt dung lượng hữu ích (tổng dung lượng của các thuê bao thỏa mãn yêu cầu tốc độ tối thiểu) và dung lượng không hữu ích (tổng dung lượng của các thuê bao không được thỏa mãn tốc độ tối thiểu), số lượng thuê bao có dung lượng hữu ích và không hữu ích, và được so sánh với kỹ thuật áp dụng tại [40]. Kết quả mô phỏng cho thấy kỹ thuật đề xuất có thể đảm bảo được tốc độ tối thiểu cho thuê bao và có dung lượng hữu ích cao hơn dung lượng hữu ích của kỹ thuật trong [40].

2.2 Mô hình hệ thống

Mô hình hệ thống được nghiên cứu là một tế bào đa người dùng Massive MIMO bao gồm một trạm gốc BS và nhiều thuê bao, với mỗi thuê bao có một ăng ten. Ký hiệu $\mathbb{K}_a = \{1, 2, \dots, K_a\}$ là tập các thuê bao này, trong đó K_a là tổng số thuê bao. Trạm gốc BS được trang bị M ăng ten, với $M \geq K_a$. Để cho đơn giản, giả thiết rằng BS biết mọi thuê bao đều được mong muốn phục vụ trong các chu kỳ khung được xem xét.

Hệ thống hoạt động ở chế độ TDD và tính thuận nghịch của kênh truyền là hoàn hảo. Ký hiệu T là khoảng thời gian của một chung kỳ khung. Trong mỗi khung τ_p tín hiệu đầu tiên được sử dụng để ước lượng kênh và $(T - \tau_p)$ tín hiệu còn lại được sử dụng để truyền dữ liệu chiều xuống.

Luận án giả thiết rằng kênh truyền là phẳng trong miền tần số và là hằng số trong mỗi chu kỳ khung. Ký hiệu $\mathbf{g}_k[n] \in \mathbb{C}^{M \times 1}$ là hệ số kênh truyền fa đình nhanh, giá trị này thay đổi giữa các khung thời gian. Để phân tích thì luận án sử dụng mô hình kênh Rayleigh fa đình với giá trị của $\mathbf{g}_k[n] \in \mathbb{C}^{M \times 1}$ là phân bố độc lập (i.i.d.) tuân theo phân bố Gauss với trung bình không và phương sai đơn vị. Ký hiệu β_k là hệ số fa đình chậm của kênh truyền của thuê bao $k \in \mathbb{K}_a$ đối với trạm gốc BS, giá trị này không thay đổi trong các chu kỳ khung được xem xét. Hệ số kênh đường lên của thuê bao $k \in \mathbb{K}_a$ được xác định bởi $\mathbf{h}_k[n] = \sqrt{\beta_k} \mathbf{g}_k[n] \in \mathbb{C}^{M \times 1}$.

Ở đây giả thiết rằng véc tơ hệ số kênh truyền thay đổi giữa các chu kỳ khung do hiện tượng lão hóa kênh truyền. Để tiện tính toán thì mối quan hệ giữa hai véc tơ hệ số kênh truyền giữa hai chu kỳ khung liên tiếp được thể hiện theo [40]

$$\mathbf{h}_k[n] = \alpha_k \mathbf{h}_k[n-1] + \mathbf{e}_k[n], \quad (2.1)$$

trong đó α_k là hệ số tự tương quan tạm thời của thuê bao k , $\mathbf{h}_k[n-1]$ là véc tơ hệ số kênh truyền trong khung trước đó của thuê bao k , và $\mathbf{e}_k[n]$ là sự biến đổi hệ số kênh truyền không tương quan do hiện tượng lão hóa kênh truyền. Về nguyên tắc, α_k phụ thuộc vào môi trường địa lý, tốc độ di chuyển của thuê bao và đặc tính của ăng ten [32]. Để cho đơn giản, giả thiết rằng α_k không thay đổi trong các khung được xem xét với mọi thuê bao $k \in \mathbb{K}_a$.

2.3 Ước lượng kênh truyền

Ở giai đoạn ước lượng kênh truyền của khung n , trạm gốc BS lựa chọn ngẫu nhiên một lượng cố định K_p thuê bao từ tập K_a thuê bao theo cách tuần tự cho mục đích ước lượng kênh truyền. Ký hiệu $\mathbb{K}_p[n] \in \mathbb{K}_a$ là tập thuê bao được lựa chọn tại khung n . Sau khi được thông báo, các thuê bao này sẽ truyền đồng thời các chuỗi hoa tiêu trực giao từng đôi xác định trước có chiều dài $\tau_p \geq K_p$ với cùng một công

suất p_p . Ký hiệu $\mathbf{v}_k^H[n] \in \mathbb{C}^{1 \times \tau_p}$ là chuỗi hoa tiêu gửi bởi thuê bao $k \in \mathbb{K}_p[n]$. Tín hiệu hoa tiêu nhận được tại trạm gốc BS là

$$\mathbf{Y}_r[n] = \sum_{k \in \mathbb{K}_p} \sqrt{\tau_p p_p} \mathbf{h}_k[n] \mathbf{v}_k^H[n] + \mathbf{Z}[n], \quad (2.2)$$

trong đó p_p là công suất phát trung bình tại mỗi thuê bao, $\mathbf{Z}[n] \in \mathbb{C}^{M \times \tau_p}$ là ma trận nhiễu trắng Gauss với các hệ số có phân bố độc lập $\mathcal{CN}(0, \sigma_r^2 \mathbf{I}_M)$. Giá trị ước lượng MMSE của $\mathbf{h}_k[n]$ được tính bằng [27]:

$$\hat{\mathbf{h}}_k[n] = \frac{\sqrt{\tau_p p_p}}{\sigma_r^2 + \tau_p p_p} \mathbf{Y}_r[n] \mathbf{v}_k[n]. \quad (2.3)$$

Hơn nữa, vì nguyên tắc trực giao của ước lượng MMSE, $\mathbf{h}_k[n]$ có thể được tách ra thành hai thành phần không tương quan như sau [40]

$$\mathbf{h}_k[n] = \hat{\mathbf{h}}_k[n] + \tilde{\mathbf{h}}_k[n], \quad (2.4)$$

trong đó $\tilde{\mathbf{h}}_k[n]$ là véc tơ lỗi ước lượng không tương quan. $\tilde{\mathbf{h}}_k[n]$ là một véc tơ có các giá trị phân bố độc lập $\mathcal{CN}(0, \xi_k \mathbf{I}_M)$, $\xi_k = \frac{\tau_p p_p \beta_k^2}{\tau_p p_p \beta_k + \sigma_r^2}$.

2.4 Truyền dữ liệu chiều xuống

Sau giai đoạn ước lượng kênh truyền, trạm gốc BS lựa chọn tập thuê lập lịch $\mathbb{K}_s (K_s \geq K_p)$ để truyền dữ liệu. Ký hiệu $\mathbf{x}[n] \in \mathbb{C}^{K_s \times 1}$ là véc tơ tín hiệu cho K_s thuê bao và $\mathbb{E} \{ \|\mathbf{x}[n]\|^2 \} = 1$. Trạm gốc BS sử dụng một ma trận tiền mã hóa tuyến tính $\mathbf{F} \in \mathbb{C}^{M \times K_s}$, đó là một hàm được tính toán từ kênh truyền $\hat{\mathbf{H}}[n]$ để gán dữ liệu cần truyền $\mathbf{x}[n]$ vào các ăng ten phát. Công suất được sử dụng cho thuê bao k là $p_k[n]$ với điều kiện giới hạn về công suất $\sum_{k=1}^{K_s} |\mathbf{f}_k[n]|^2 p_k[n] \leq P$. Tín

hiệu nhận được tại thuê bao k có thể được viết lại thành:

$$\begin{aligned}
\mathbf{y}_k[n] &= \mathbf{h}_k^T[n] \mathbf{f}_k[n] \sqrt{p_k[n]} x_k[n] \\
&+ \sum_{l=1, l \neq k}^{K_s} \mathbf{h}_k^T[n] \mathbf{f}_l[n] \sqrt{p_l[n]} x_l[n] + \mathbf{n}_k \\
&= \mathbb{E}\{\mathbf{h}_k^T[n] \mathbf{f}_k[n]\} \sqrt{p_k[n]} x_k[n] \\
&+ \sum_{l=1}^{K_s} \mathbf{h}_k^T[n] \mathbf{f}_l[n] \sqrt{p_l[n]} x_l[n] \\
&- \mathbb{E}\{\mathbf{h}_k^T[n] \mathbf{f}_k[n]\} \sqrt{p_k[n]} x_k[n] + \mathbf{n}_k,
\end{aligned} \tag{2.5}$$

trong đó $\mathbf{h}_k[n]$ là véc tơ kênh truyền của thuê bao k , và $\mathbf{f}_k[n]$ là cột thứ k của ma trận $\mathbf{F}[n]$.

Tỷ lệ tín hiệu trên nhiễu SINR của thuê bao k có thể được viết lại thành:

$$\gamma_k[n] = \frac{p_k[n] |\phi_k|^2}{\sum_{l=1}^{K_s} p_l[n] \mathbb{E}\{|\mathbf{h}_k^T[n] \mathbf{f}_l[n]|^2\} - p_k[n] |\phi_k|^2 + \sigma^2}, \tag{2.6}$$

trong đó $\phi_k = \mathbb{E}\{\mathbf{h}_k^T[n] \mathbf{f}_k[n]\}$.

Tốc độ có thể đạt được của thuê bao k là

$$\mathbf{R}_k[n] = \log_2(1 + \gamma_k[n]). \tag{2.7}$$

Tốc độ tổng của hệ thống là

$$\mathbf{R}_{sum}[n] = \sum_{l=1}^{K_s} \log_2(1 + \gamma_l[n]). \tag{2.8}$$

2.5 Lão hóa kênh truyền

Sự biến thiên về kênh truyền xảy ra đối với tập thuê bao còn lại $\mathbb{K}_r[n] = \mathbb{K}_a \setminus \mathbb{K}_p[n]$ do sự khác nhau về mặt thời gian giữa thời điểm ước lượng kênh và thời điểm sử dụng kênh truyền để truyền dữ liệu chiều xuống. Luận án định nghĩa hệ số biến

đổi kênh truyền $\delta_k[n]$ dùng để đo lường sự biến đổi kênh truyền giữa khung thời gian cuối cùng l_k mà kênh truyền của thuê bao k được ước lượng và khung thời gian hiện tại n [40]

$$\delta_k[n] = \alpha_k^{n-l_k} \quad (2.9)$$

$$\mathbf{h}_k[n] = \delta_k[n]\hat{\mathbf{h}}_k[l_k] + \tilde{\mathbf{e}}_k[n], \quad (2.10)$$

trong đó, $\tilde{\mathbf{e}}_k[n] = \delta_k[n]\tilde{\mathbf{h}}_k[l_k] + \mathbf{e}_k[n]$ là phân bố độc lập đều với trung bình không và phương sai $(\beta_k - \delta_k^2[n]\xi_k)$.

2.6 Tốc độ tổng có thể đạt được với hiện tượng lão hóa kênh truyền

Luận án nghiên cứu tốc độ tổng chiều xuống có thể đạt được khi sử dụng thông tin lão hóa kênh truyền CSI để lập lịch. Giá trị ước lượng hiện thời CSI $\hat{\mathbf{h}}_k[n]$ của thuê bao k được tính từ giá trị ước lượng kênh lần cuối $\hat{\mathbf{h}}_k[l_k]$ tại khung thời gian l_k .

$$\hat{\mathbf{h}}_k[n] = \delta_k\hat{\mathbf{h}}_k[l_k]. \quad (2.11)$$

Tín hiệu nhận được của thuê bao k là

$$\begin{aligned} \mathbf{y}_k[n] &= \hat{\mathbf{h}}_k^T[n]\mathbf{F}[n]\mathbf{P}\mathbf{x}[n] + \tilde{\mathbf{e}}_k[n]\mathbf{F}[n]\mathbf{P}\mathbf{x}[n] + \mathbf{n}_k[n] \\ &= \hat{\mathbf{h}}_k^T[n]\mathbf{f}_k[n]\sqrt{p_k}x_k[n] \\ &\quad + \sum_{l=1, l \neq k}^{K_s} \hat{\mathbf{h}}_k^T[n]\mathbf{f}_l[n]\sqrt{p_l}x_l[n] \\ &\quad + \tilde{\mathbf{e}}_k[n]\mathbf{F}[n]\mathbf{P}\mathbf{x}[n] + \mathbf{n}_k \\ &= \mathbb{E}\{\hat{\mathbf{h}}_k^T[n]\mathbf{f}_k[n]\}\sqrt{p_k}x_k[n] + \epsilon_k[n], \end{aligned} \quad (2.12)$$

trong đó $\epsilon_k[n] = \sum_{l=1}^{K_s} \hat{\mathbf{h}}_k^T[n]\mathbf{f}_l[n]\sqrt{p_l}x_l[n] - \mathbb{E}\{\hat{\mathbf{h}}_k^T[n]\mathbf{f}_k[n]\}\sqrt{p_k}x_k[n] + \tilde{\mathbf{e}}_k[n]\mathbf{F}[n]\mathbf{P}\mathbf{x}[n] + \mathbf{n}_k[n]$.

2.6.1 Tiền mã hóa MRT

Ma trận tiền mã hóa được xây dựng bằng

$$\mathbf{F}[n] = \hat{\mathbf{H}}_S^*[n]. \quad (2.13)$$

$$\mathbb{E}\{\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_k^*[n]\} = \xi_k\delta_k^2[n]M. \quad (2.14)$$

$$\mathbb{E}\{|\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_k^*[n]|^2\} = \xi_k^2\delta_k^4[n](M^2 + M). \quad (2.15)$$

$$\mathbb{E}\{|\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_l^*[n]|^2\} = \xi_k\delta_k^2[n]\xi_l\delta_l^2[n]M. \quad (2.16)$$

$$\mathbb{E}\{|\tilde{\mathbf{e}}_k[n]\hat{\mathbf{H}}_S^*[n]\mathbf{P}\mathbf{x}[n]|^2\} = (\beta_k - \delta_k^2[n]\xi_k). \quad (2.17)$$

Kết quả là,

$$\gamma_k^{MRT}[n] = \frac{p_k\xi_k^2\delta_k^4[n]M^2}{\sum_{l=1}^{K_s} \xi_k\delta_k^2[n]p_l\xi_l\delta_l^2[n]M + (\beta_k - \delta_k^2[n]\xi_k + \sigma^2)}. \quad (2.18)$$

2.6.2 Tiền mã hóa ZF

Đối với tiền mã hóa ZF, ma trận tiền mã hóa được xây dựng bằng

$$\mathbf{F}[n] = \hat{\mathbf{H}}_S^*[n](\hat{\mathbf{H}}_S^T[n]\hat{\mathbf{H}}_S^*[n])^{-1}. \quad (2.19)$$

$$\mathbb{E}\{\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_k^*[n](\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_k^*[n])^{-1}\} = 1. \quad (2.20)$$

$$\begin{aligned} \mathbb{E}\{|\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_k^*[n](\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_k^*[n])^{-1}|^2\} &= 1 \\ &+ \frac{(\beta_k - \delta_k^2\xi_k)}{\delta_k^2\xi_k(M - K_s)}. \end{aligned} \quad (2.21)$$

$$\mathbb{E}\{|\hat{\mathbf{h}}_k^T[n]\hat{\mathbf{h}}_l^*[n](\hat{\mathbf{h}}_l^T[n]\hat{\mathbf{h}}_l^*[n])^{-1}|^2\} = \frac{(\beta_k - \delta_k^2\xi_k)}{\delta_l^2\xi_l(M - K_s)}, \quad (2.22)$$

$$l \neq k$$

$$\mathbb{E}\{|\tilde{\mathbf{e}}_k[n]\mathbf{F}[n]\mathbf{P}\mathbf{x}[n]|^2\} = (\beta_k - \delta_k^2[n]\xi_k). \quad (2.23)$$

Kết quả là

$$\gamma_k^{zf}[n] = \frac{p_k(M - K_s)}{\sum_{l=1}^{K_s} \frac{p_l(\beta_k - \delta_k^2[n]\xi_k)}{\delta_l^2[n]\xi_l} + \varphi(M - K_s)},$$

$$\varphi = (\beta_k - \delta_k^2[n]\xi_k + \sigma^2). \quad (2.24)$$

2.7 Kỹ thuật lập lịch đề xuất QoS-Aware

Sử dụng thông tin lão hóa kênh truyền CSI có thể làm tăng độ lợi về ghép kênh nhưng nó có thể dẫn tới tốc độ trung bình của thuê bao bị giảm dần và có thể không đảm bảo yêu cầu về tốc độ tối thiểu của thuê bao theo các cam kết về chất lượng dịch vụ QoS. Để giải quyết vấn đề này, luận án đề xuất một kỹ thuật lập lịch không những phục vụ được nhiều thuê bao hơn để tăng độ lợi về ghép kênh mà còn đảm bảo tốc độ tối thiểu của thuê bao nhằm thỏa mãn các yêu cầu về chất lượng dịch vụ QoS. Ý tưởng chính của kỹ thuật đề xuất là kiểm soát tốc độ có thể đạt được của các thuê bao được phục vụ và các thuê bao đang được xem xét phục vụ để đảm bảo việc thêm các thuê bao mới này vào nhóm lập lịch vẫn đảm bảo yêu cầu về tốc độ tối thiểu của mỗi thuê bao.

Bước tiếp theo là xác định đối tượng của việc lập lịch và các ràng buộc của nó. Trạm gốc BS thu thập thông tin trạng thái kênh truyền $\hat{\mathbf{H}}$, tổng công suất phát P và yêu cầu về tốc độ tối thiểu T của các thuê bao được phục vụ theo các yêu cầu về QoS. Dựa trên thông tin thu thập được, kỹ thuật lập lịch tối đa hóa tốc độ tổng của toàn hệ thống bằng cách lựa chọn ra tập thuê bao tối ưu nhất $\mathbb{K}_s[n]$ từ tập thuê bao được ước lượng kênh $\mathbb{K}_p[n]$ và thông tin lão hóa kênh truyền trong

mỗi chu kỳ khung.

$$\begin{aligned}
& \max_{\mathbb{K}_s[n] \in \mathbb{K}_a} \sum_{k=1}^{K_s} \log_2(1 + \gamma_k[n]) \\
& s.t \quad \sum_{k=1}^{K_s} |\mathbf{f}_k|^2 p_k \leq P \\
& \log_2(1 + \gamma_k[n]) \geq T
\end{aligned} \tag{2.25}$$

Như thuật toán 1, kỹ thuật lập lịch bao gồm bốn phần: *Ước lượng hệ số biến đổi kênh truyền* giúp BS ước lượng sự biến đổi về kênh truyền của mỗi thuê bao tính từ lần ước lượng CSI cuối cùng, *Lựa chọn tập thuê bao hoa tiêu* lựa chọn một tập thuê bao $\mathbb{K}_p[n]$ để ước lượng kênh, *Ước lượng kênh truyền* sẽ thực hiện việc ước lượng kênh cho tập thuê bao $\mathbb{K}_p[n]$, và *Lựa chọn tập thuê bao phù hợp* sẽ cho phép thêm nhiều thuê bao được phục vụ dựa trên thông tin lão hóa kênh truyền để tăng độ lợi về ghép kênh nhưng vẫn đảm bảo tốc độ tối thiểu của thuê bao.

Algorithm 1 Kỹ thuật lập lịch QoS-Aware

- 1: Ước lượng hệ số biến đổi kênh truyền
 $\delta_k[n] = \alpha_k^{n-l_k}, \forall k \in \mathbb{K}_a$
 - 2: Lựa chọn tập thuê bao hoa tiêu
 $\mathbb{K}_p[n]$ từ tập \mathbb{K}_a
 - 3: Ước lượng kênh truyền
Cập nhật $(l_k, \hat{\mathbf{h}}_k[l_k], n, \hat{\mathbf{h}}_k[n])$ cho thuê bao $k \in \mathbb{K}_p[n]$
 - 4: Lựa chọn tập thuê bao phù hợp $\mathbb{K}_s[n]$
 - 5: Truyền dữ liệu chiều xuống cho tập $\mathbb{K}_s[n]$
-

Sau mỗi thủ tục ước lượng kênh và cập nhật thông tin trạng thái kênh truyền cho tập thuê bao $\mathbb{K}_p[n]$, trạm gốc BS sẽ chọn ra tập thuê bao phù hợp để tăng được dung lượng hệ thống nhưng vẫn thỏa mãn được tốc độ tối thiểu cho từng thuê bao. Ở đây ký hiệu $\mathbb{K}_s[n]$ là tập thuê bao được lập lịch và $\mathbb{K}_c[n]$ là tập thuê bao ứng cử viên. Để xác định xem thuê bao k có được lập lịch hay không, tại khung thời gian n tốc độ R_l của thuê bao $l \in \{k \cup \mathbb{K}_s[n]\}$ phải cao hơn tốc độ tối thiểu T . Nếu tất cả thuê bao đều thỏa mãn yêu cầu về QoS thì chúng ta sẽ kiểm tra xem liệu việc thêm thuê bao k có giúp làm tăng dung lượng hệ thống

hay không $R^{pc}(k \cup \mathbb{K}_s[n]) > R_{sum}$. Cuối cùng, nếu tồn tại thuê bao phù hợp nhất k trong tập $\mathbb{K}_c[n]$ thỏa mãn cả hai điều kiện trên thì thuê bao k sẽ được chuyển sang tập $\mathbb{K}_s[n]$:

$$\begin{aligned}\mathbb{K}_s[n] &= k_{best} \cup \mathbb{K}_s[n] \\ \mathbb{K}_c[n] &= \mathbb{K}_c[n] \setminus \{k_{best}\}\end{aligned}$$

Trạm gốc sẽ lặp lại những thủ tục này cho đến khi không còn thuê bao nào thỏa mãn được điều kiện. Chi tiết quá trình *Lựa chọn tập thuê bao phù hợp* này được miêu tả chi tiết trong thuật toán 2.

Algorithm 2 Lựa chọn tập thuê bao phù hợp

```

1:  $\mathbb{K}_s[n] = \emptyset, \mathbb{K}_c[n] = \mathbb{K}_a, R_{sum} = 0$ 
2:  $morevaliduser = 1$ 
3: while  $morevaliduser$  do
4:    $k_{best} \leftarrow 0$ 
5:   for user  $k \in \mathbb{K}_c[n]$  do
6:      $QoScheck = 1$ 
7:     for user  $l \in \{k \cup \mathbb{K}_s[n]\}$  do
8:       Calculate  $\gamma_l[n]$  follow MRT or ZF
9:        $R_l = \log_2(1 + \gamma_l[n])$ 
10:      if  $R_l < T$  then
11:         $QoScheck = 0$ 
12:      end if
13:    end for
14:    if  $(R^{pc}(k \cup \mathbb{K}_s[n]) > R_{sum})$  and  $(QoScheck)$  then
15:       $k_{best} \leftarrow k$ 
16:       $R_{sum} \leftarrow R^{pc}(k \cup \mathbb{K}_s[n])$ 
17:    end if
18:  end for
19:  if  $k_{best} \neq 0$  then
20:     $\mathbb{K}_s[n] = k_{best} \cup \mathbb{K}_s[n]$ 
21:     $\mathbb{K}_c[n] = \mathbb{K}_c[n] \setminus \{k_{best}\}$ 
22:  else
23:     $morevaliduser = 0$ 
24:  end if
25: end while

```

2.8 Kết quả mô phỏng

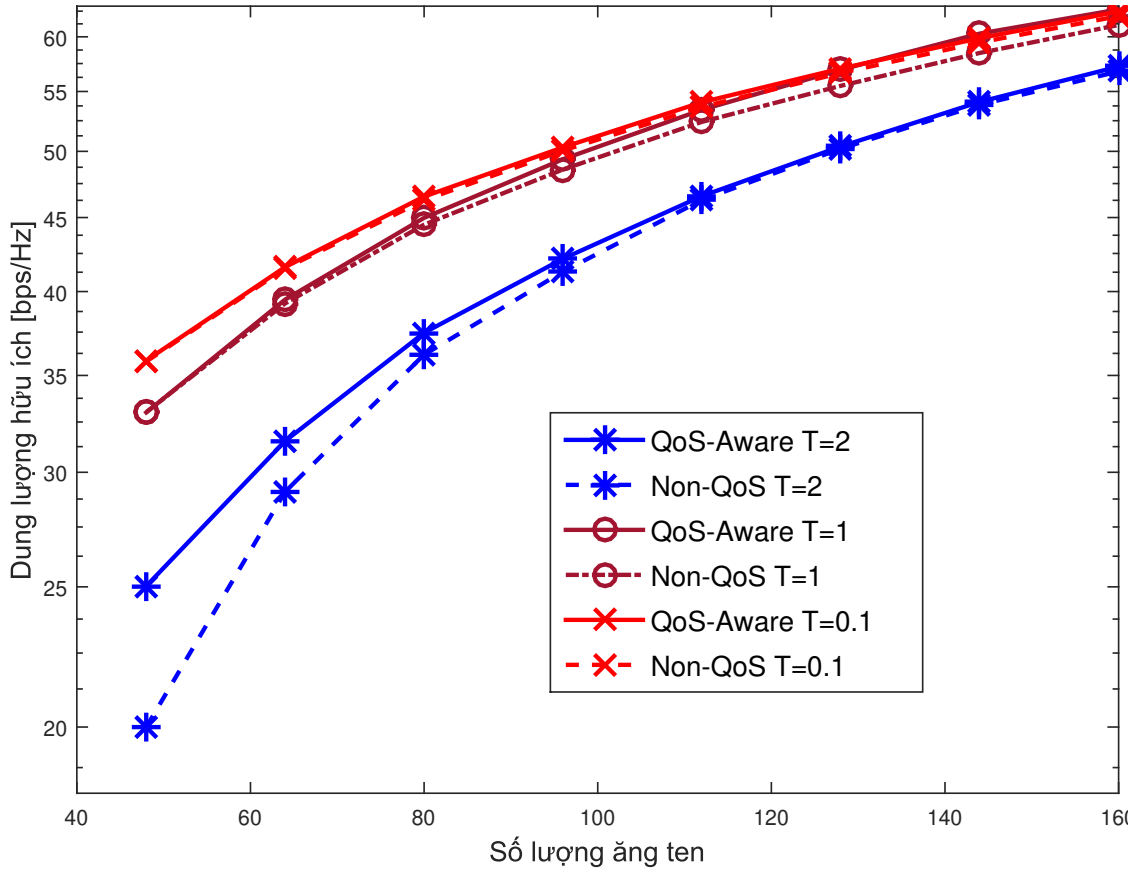
Để đo lường hiệu quả của kỹ thuật QoS-Aware, nhiều trường hợp mô phỏng đã được thực hiện trên nền hệ thống Massive MIMO bằng phần mềm Matlab để so sánh giữa các kỹ thuật lập lịch sau:

- Kỹ thuật Non-QoS (OpSac trong [40]).
- Kỹ thuật toán QoS-Aware.

Luận án tập trung so sánh dung lượng hữu ích, đây là tổng dung lượng của các thuê bao trong đó tốc độ của thuê bao cao hơn tốc độ yêu cầu tối thiểu, và ngược lại thì gọi là dung lượng không hữu ích. Trong các tình huống mô phỏng, các tham số mô phỏng được mặc định như sau: độ dài chuỗi tín hiệu hoa tiêu $\tau_p = K_p$, và tỷ lệ tín hiệu trên nhiễu của đường lên và đường xuống là $SINR_r = 20$ dB, $SINR = 30$ dB. Kịch bản mô phỏng là tăng dần số lượng ăng ten để đánh giá hiệu quả của kỹ thuật đề xuất đối với dung lượng hữu ích, dung lượng không hữu ích và số lượng thuê bao được phục vụ trong cả hai kỹ thuật tiền mã hóa ZF và MRT.

Hình 2.1 so sánh dung lượng hữu ích giữa QoS-Aware và Non-QoS khi số lượng ăng ten của trạm gốc BS tăng dần và $K_a = 40$ cho các trường hợp tốc độ tối thiểu khác nhau là $T = 0.1, 1$ và 2 (bit/s/Hz). Cả hai kỹ thuật đều sử dụng tiền mã hóa MRT. Có thể thấy rằng dung lượng hữu ích của hệ thống tăng dần khi số lượng ăng ten tại trạm gốc M tăng lên. Hơn nữa, dung lượng hữu ích của kỹ thuật QoS-Aware luôn cao hơn dung lượng hữu ích của kỹ thuật Non-QoS. Nếu như yêu cầu về tốc độ tối thiểu của thuê bao T giảm dần thì sự khác biệt giữa dung lượng hữu ích của hai kỹ thuật cũng giảm dần. Đặc biệt là với $T = 0.1$ thì dung lượng hữu ích của QoS-Aware và Non-Qos gần như giống nhau. Cuối cùng,

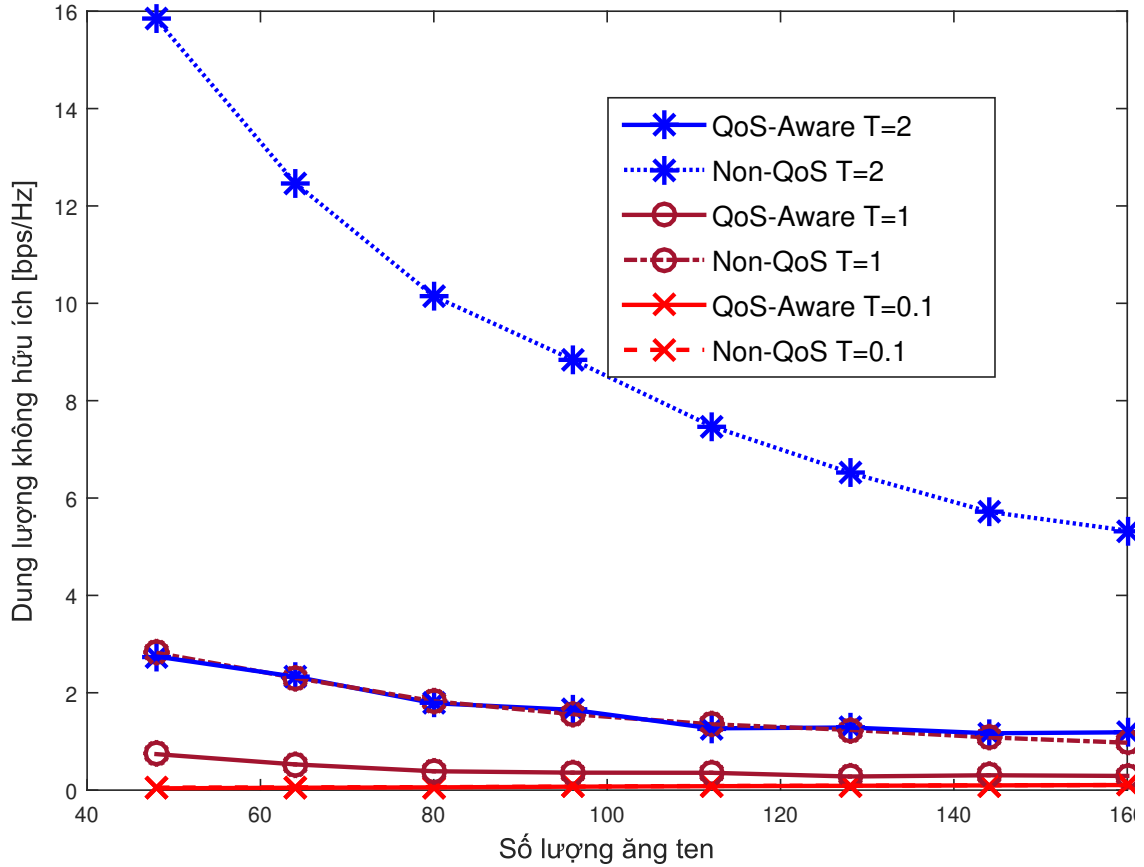
khi T giảm thì dung lượng hữu ích của cả hai kỹ thuật đều tăng dần.



Hình 2.1: So sánh dung lượng hữu ích khi sử dụng MRT

Hình 2.2 cho thấy dung lượng không hữu ích của kỹ thuật QoS-Aware là ít hơn so với dung lượng không hữu ích của kỹ thuật Non-QoS với cùng giá trị T . Yêu cầu về tốc độ tối thiểu T càng cao thì dung lượng không hữu ích càng lớn. Trường hợp xấu nhất của dung lượng không hữu ích là khi $T=2$ khi sử dụng kỹ thuật Non-QoS. Điều đó cho thấy rõ rằng kỹ thuật Non-QoS không nên được áp dụng triển khai tại những mạng vô tuyến đòi hỏi chất lượng dịch vụ với tốc độ cao.

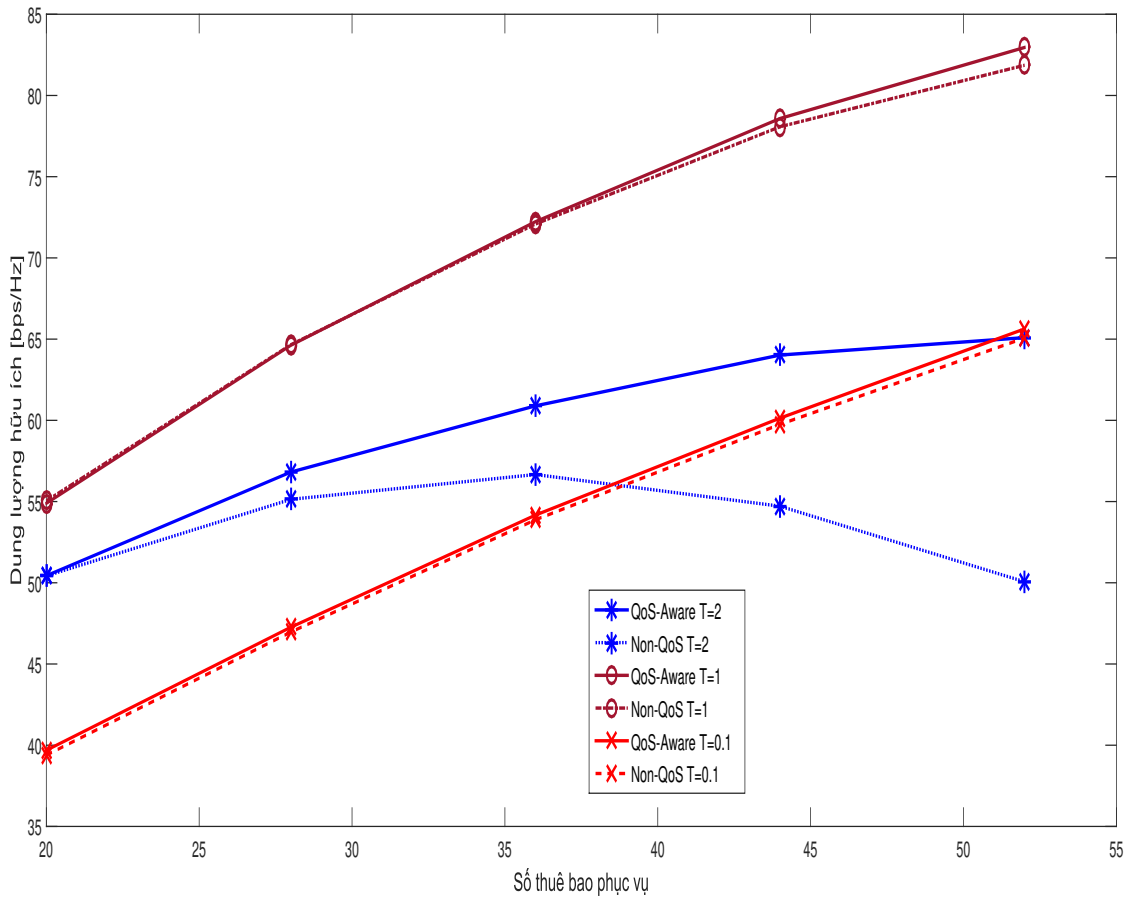
Hình 2.3 biểu diễn dung lượng hữu ích của hệ thống khi số thuê bao K_a tăng dần và cố định số ăng ten $M = 128$. Bình thường khi thêm thuê bao sẽ làm tăng dung lượng hữu ích của hệ thống. Tuy nhiên, nếu sử dụng kỹ thuật Non-QoS và T cao thì dung lượng hữu ích sẽ giảm xuống ví dụ như trường hợp $T = 2$. Điều



Hình 2.2: So sánh dung lượng không hữu ích khi sử dụng MRT

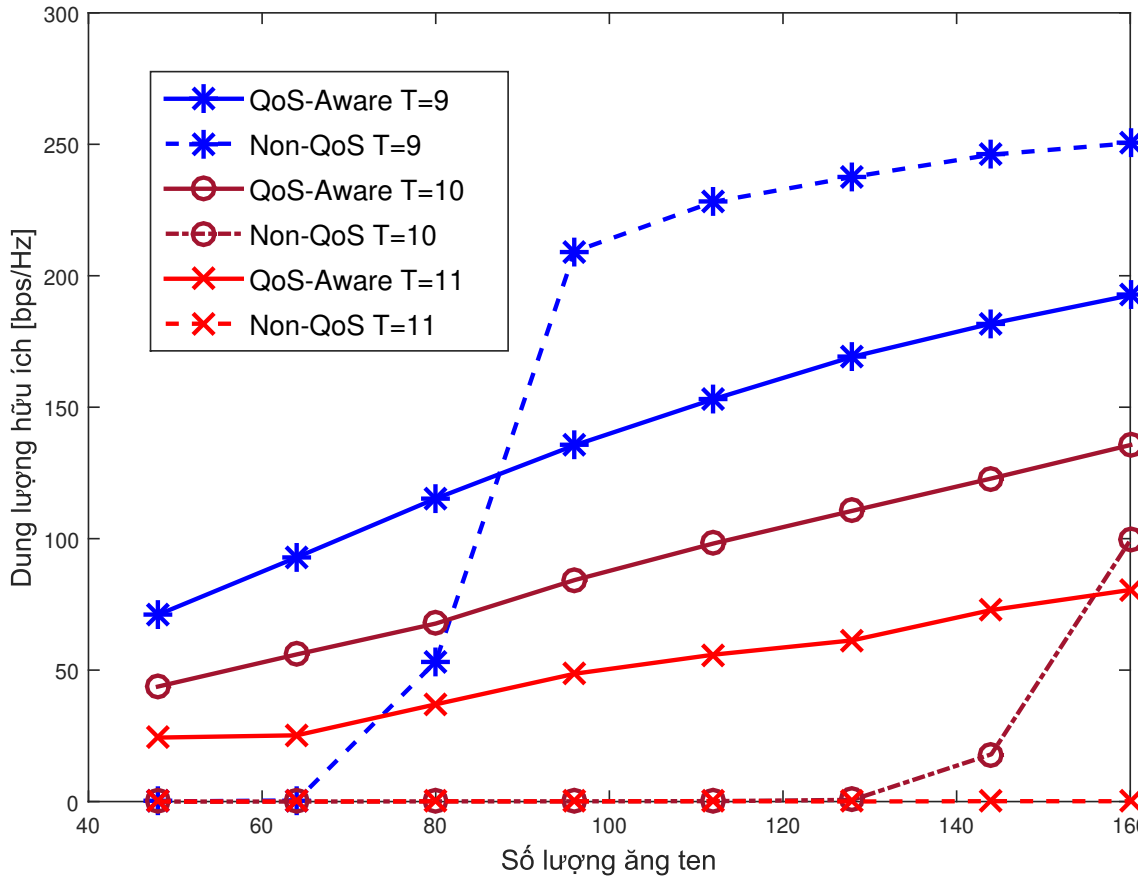
đó có nghĩa là nếu sử dụng kỹ thuật Non-QoS thì khi phục vụ thêm thuê bao sẽ dẫn tới có nhiều thuê bao có dung lượng không hữu ích. Hiện tượng này xảy ra nhiều hơn khi tốc độ trung bình của thuê bao nhỏ hơn tốc độ yêu cầu tối thiểu T . Tuy nhiên, với $T = 2$ thì khi sử dụng kỹ thuật QoS-Aware chúng ta thấy rằng dung lượng hữu ích vẫn tăng lên khi thêm vào thuê bao mới bởi vì kỹ thuật chỉ lựa chọn các thuê bao có chất lượng kênh truyền tốt nhất và vẫn luôn đảm bảo rằng việc thêm thuê bao vẫn thỏa mãn điều kiện về QoS.

Hình 2.4 thể hiện dung lượng hữu ích của thuật toán QoS-Aware và Non-QoS với tiền mã hóa ZF khi $K_a = 40$ và M tăng dần cho trường hợp tốc độ tối thiểu $T = 9, 10$, và 11 . Có thể thấy rằng bình thường dung lượng hữu ích của QoS-Aware luôn cao hơn dung lượng hữu ích của Non-QoS. Tuy nhiên, nếu như T nhỏ



Hình 2.3: Dung lượng hữu ích của hệ thống khi sử dụng MRT và $M = 128$

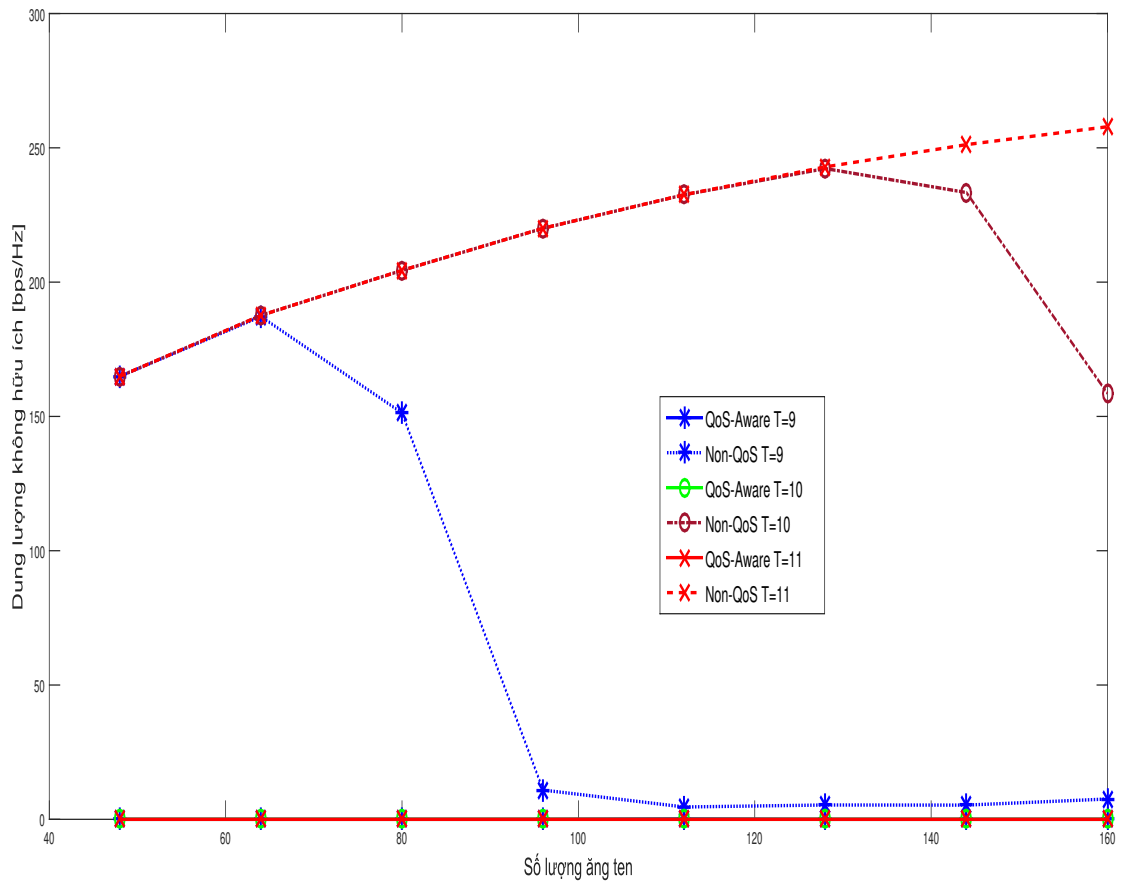
hơn tốc độ trung bình của thuê bao thì dung lượng hữu ích của QoS-Aware sẽ nhỏ hơn dung lượng hữu ích của Non-QoS, ví dụ với $T = 9$. Điều đó có nghĩa là nếu tốc độ tối thiểu T quá thấp thì kỹ thuật QoS-Aware sẽ không còn ý nghĩa. Cuối cùng, khi số lượng ăng ten tại trạm gốc tăng lên thì dung lượng hữu ích của cả hai kỹ thuật cùng tăng lên.



Hình 2.4: So sánh dung lượng hữu ích khi sử dụng tiền mã hóa ZF và $K_a = 40$

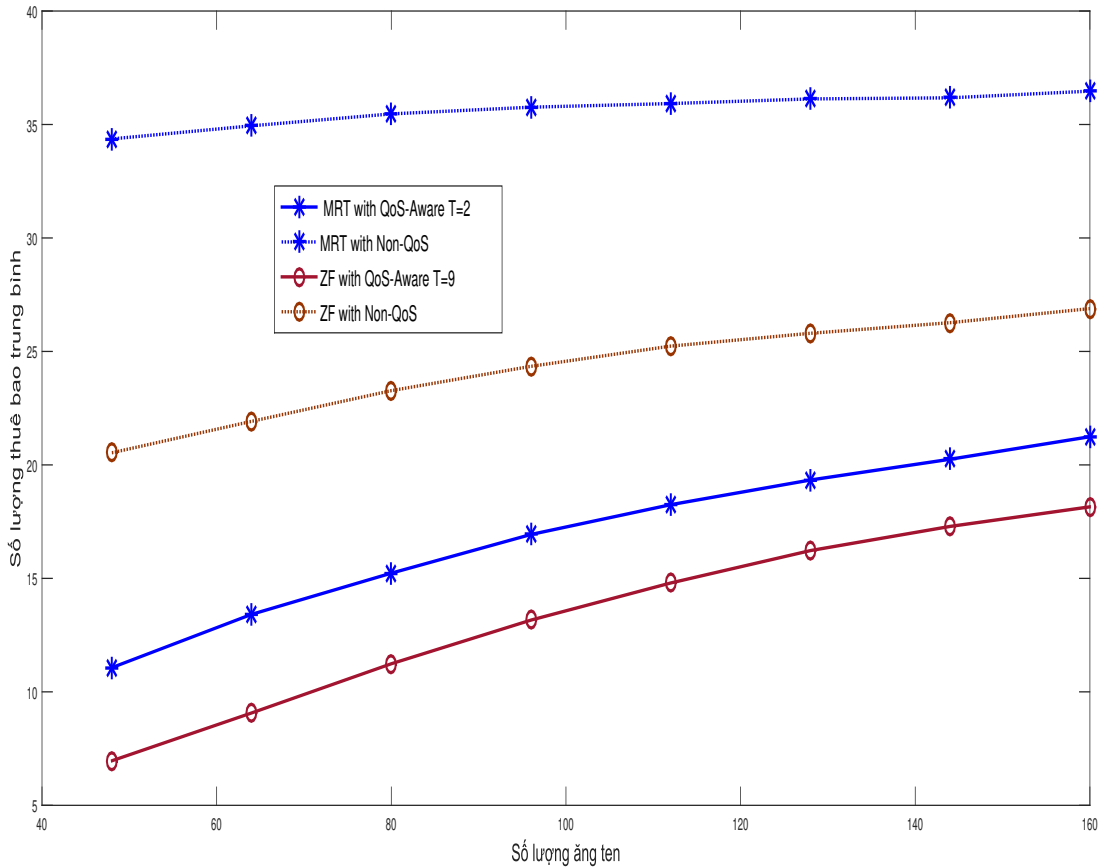
Hình 2.5 thể hiện dung lượng không hữu ích của cả hai kỹ thuật khi sử dụng tiền mã hóa ZF. QoS-Aware gần như không có dung lượng không hữu ích vì vấn đề can nhiễu bị triệt tiêu khi M tăng lên. Với kỹ thuật Non-QoS, nếu như T thấp hơn tốc độ trung bình của thuê bao thì dung lượng không hữu ích cũng tiến về không, ví dụ với $T = 9$. Một lần nữa khẳng định rằng nếu T rất thấp thì không cần thiết sử dụng kỹ thuật QoS-Aware.

Hình 2.6 thể hiện số lượng thuê bao được phục vụ khi sử dụng tiền mã hóa MRT và ZF. Như thường lệ, kỹ thuật lập lịch Non-QoS với MRT sẽ phục vụ số thuê bao lớn nhất. Kỹ thuật QoS-Aware sẽ phục vụ ít thuê bao hơn nếu ngưỡng chất lượng dịch vụ T tăng lên. Hơn nữa, MRT sẽ luôn phục vụ nhiều thuê bao hơn ZF với cùng ngưỡng tốc độ tối thiểu T . Bởi vì với Non-QoS sử dụng MRT, hầu



Hình 2.5: So sánh lưu lượng không hữu ích khi sử dụng tiền mã hóa ZF với $K_a = 40$

hết thuê bao sẽ được phục vụ kể cả khi thuê bao ở trong điều kiện kênh truyền xấu và chỉ đạt được dung lượng không hữu ích. Trong khi đó, Non-QoS sử dụng ZF sẽ chỉ lựa các thuê bao với kênh truyền ở điều kiện tốt.



Hình 2.6: Số thuê bao được phục vụ khi sử dụng ZF và MRT

2.9 Kết luận chương 2

Trong chương 2 của luận án đã đề xuất kỹ thuật lập lịch có thể đảm bảo được tốc độ tối thiểu cho các thuê bao khi sử dụng thông tin lão hóa kênh truyền. Kết quả mô phỏng đã xác nhận dung lượng hữu ích của kỹ thuật đề xuất vượt trội so với kỹ thuật lập lịch ở [40]. Dung lượng không hữu ích tức là không đáp ứng yêu cầu tốc độ của khách hàng không đáng kể. Dung lượng hữu ích của kỹ thuật đề xuất không bị giảm đi khi số thuê bao có mặt trong trạm phát sóng tăng lên nhờ vào việc kỹ thuật sẽ giới hạn số thuê bao phục vụ để đảm bảo rằng tất cả thuê bao được chọn phục vụ đều được đảm bảo về tốc độ tối thiểu. Tính khả thi của kỹ thuật đề xuất được minh chứng là nếu số ăng ten tăng lên vô hạn thì dung lượng

không hữu ích sẽ tiến tới không, mọi thuê bao đều chỉ có dung lượng hữu ích. Hạn chế của kỹ thuật đề xuất là nếu ngưỡng tốc độ tối thiểu là quá thấp so với tốc độ trung bình của thuê bao thì việc áp dụng kỹ thuật QoS-Aware là không cần thiết.

Chương 3

Kỹ thuật lập lịch cho hệ thống ăng ten cỡ rất lớn đảm bảo đa tốc độ tối thiểu

Chương này của luận án đề xuất kỹ thuật lập lịch hướng tới cho phép thuê bao có thể yêu cầu tốc độ tối thiểu khác nhau. Kỹ thuật cũng xây dựng công thức tính toán chính xác công suất tối thiểu mà thuê bao cần có để đạt được tốc độ tối thiểu tùy ý. Hơn nữa, kỹ thuật cũng đề xuất sử dụng tham số về độ ưu tiên của dung lượng để đảm bảo rằng trong các điều kiện như nhau thì dung lượng có độ ưu tiên cao hơn sẽ được phục vụ trước.

3.1 Ý tưởng về kỹ thuật lập lịch đảm bảo đa tốc độ tối thiểu cho người dùng

Sự bùng nổ của internet đã ảnh hưởng đến sự phát triển của mạng di động. Các dịch vụ trên nền IP được truyền tải thông qua mạng di động. Vì thế các mạng không dây sẽ phải hỗ trợ các đặc tính của lưu lượng IP. Lưu lượng IP được chia thành các lớp dịch vụ với các yêu cầu khác nhau về QoS cho mỗi lớp. QoS xác định các thông tin như độ ưu tiên, độ trễ và tốc độ của mỗi lớp. Mạng 5G sẽ hỗ trợ QoS từ mạng lõi đến tận thuê bao đầu cuối [1].

Trong hướng nghiên cứu về kỹ thuật lập lịch đảm bảo QoS, việc đảm bảo độ trễ của gói tin nhận được nhiều sự quan tâm nghiên cứu. Độ trễ của lưu lượng được cân nhắc trong kỹ thuật M-LWDF [10]. Độ ưu tiên và đảm bảo tốc độ được nghiên cứu trong [8]. Tuy nhiên kỹ thuật về độ ưu tiên của họ lại phụ thuộc vào độ trễ, còn theo quan điểm của nghiên cứu sinh thì độ ưu tiên của lưu lượng phụ thuộc loại dịch vụ của lưu lượng đó chứ không chỉ phụ thuộc vào độ trễ. Hơn nữa, đề xuất của họ sẽ truyền đi toàn bộ lưu lượng GBR theo yêu cầu của khách hàng nhằm đảm bảo không có sự rớt gói nào cho lưu lượng GBR này. Điều này sẽ dẫn tới việc chỉ một lượng ít thuê bao GBR sẽ sử dụng gần hết băng thông của các thuê bao không đăng ký đảm bảo tốc độ. Điều khiển công suất cũng được thực hiện trong [20] để xác định lượng dữ liệu cần truyền đi phù hợp để cải thiện hiệu năng về độ trễ. Sự kết hợp giữa kỹ thuật lập lịch và điều khiển công suất cũng được nghiên cứu trong [23]. Tốc độ tối thiểu để hỗ trợ QoS trong các hệ thống ăng ten cỡ rất lớn được nghiên cứu trong [4]. Để đảm bảo được tốc độ tối thiểu cho thuê bao, một tập con của thuê bao và tập con của ăng ten được lựa chọn dựa trên hàm lỗi.

Với hiểu biết của nghiên cứu sinh, các nghiên cứu hiện nay về kỹ thuật lập lịch hỗ trợ QoS cho các hệ thống ăng ten cỡ rất lớn chủ yếu nghiên cứu về đảm bảo độ trễ hoặc chỉ đảm bảo một tốc độ tối thiểu cho thuê bao. Vấn đề đảm bảo nhiều tốc độ tối thiểu cho các thuê bao khác nhau vẫn chưa được xem xét thấu đáo. Vì vậy ở mục này nghiên cứu sinh đề xuất một kỹ thuật lập lịch với tên gọi QoS-Assurance, trong đó xác định công suất tối ưu cho mỗi thuê bao để đạt được tốc độ tối thiểu theo yêu cầu. Việc tối ưu công suất trên từng thuê bao sẽ giúp tiết kiệm công suất sử dụng để có thể phục vụ được nhiều thuê bao hơn. Hiệu quả của kỹ thuật QoS-Assurance được nghiên cứu dưới khía cạnh dung lượng toàn hệ

thống, tốc độ trung bình của thuê bao và dung lượng hữu ích. Kết quả mô phỏng cho thấy dung lượng hệ thống cải thiện hơn nhiều so với các thuật toán tập trung vào độ trễ như [8] trong khi lại có thể đảm bảo được đa tốc độ tối thiểu cho người dùng và có thể phân biệt được độ ưu tiên của lưu lượng.

3.2 Mô hình hệ thống

Mô hình hệ thống được nghiên cứu là hệ thống ăng ten cỡ rất lớn gồm một trạm gốc BTS và K_a thuê bao. Trạm gốc có M ăng ten và mỗi thuê bao chỉ có một ăng ten với điều kiện $M \geq K_a$. Tất cả thuê bao chia sẻ chung tài nguyên về thời gian và tần số. Hệ thống hoạt động trong chế độ TDD và tính thuận nghịch của kênh truyền là hoàn hảo. Ký hiệu T là chu kỳ khung. Trong mỗi chu kỳ khung, τ_p tín hiệu đầu tiên sẽ được dùng để ước lượng kênh truyền CSI và $(T - \tau_p)$ tín hiệu còn lại sẽ được dùng để truyền dữ liệu.

3.3 Ước lượng kênh truyền

Để ước lượng ma trận kênh truyền giữa BS và thuê bao, trạm gốc BS sẽ gửi chuỗi tín hiệu hoa tiêu trực giao $\mathbf{V} \in \mathbb{C}^{K_t \times \tau_p}$ có chiều dài τ_p tín hiệu xuống K_t thuê bao ($\tau_p \geq K_t$) được lựa chọn từ tập K_a . Ký hiệu $\mathbf{H} \in \mathbb{C}^{M \times K_t}$ là ma trận kênh truyền giữa trạm gốc và K_t thuê bao. Mô hình kênh truyền fa đình được sử dụng với các hệ số kênh giữ không đổi trong mỗi chu kỳ khung. Ký hiệu \mathbf{h}_k là véc tơ kênh truyền kích thước $M \times 1$ của thuê bao k -th, đây chính là một cột của ma trận \mathbf{H} và được biểu diễn bởi

$$\mathbf{h}_k = \mathbf{g}_k \sqrt{\beta_k}, \quad (3.1)$$

trong đó thành phần của \mathbf{g}_k là các phân bố Gauss với trị trung bình bằng không và phương sai đơn vị, β_k là hệ số fa đình của kênh truyền do che chắn của địa

hình và thường ổn định qua nhiều chu kỳ khung. Trạm gốc BS sẽ nhận được ma trận $M \times \tau_p$:

$$\mathbf{Y}_r = \sqrt{\tau_p p_p} \mathbf{H} \mathbf{V} + \mathbf{N}, \quad (3.2)$$

trong đó p_p là công suất phát của mỗi thuê bao, \mathbf{N} là ma trận nhiễu Gauss với các giá trị phân bố giống nhau và độc lập (i.i.d) $\mathcal{CN}(0, \sigma^2 \mathbf{I}_M)$. Vì thế, ước lượng lỗi trung bình nhỏ nhất (MMSE) của \mathbf{H} được tính bằng [41]:

$$\hat{\mathbf{H}} = \frac{\sqrt{\tau_p p_p}}{\sigma^2 + \tau_p p_p} \mathbf{Y}_r \mathbf{V}^H. \quad (3.3)$$

$$\hat{\mathbf{h}}_k = \frac{\sqrt{\tau_p p_p} \beta_k}{\sigma^2 + \tau_p p_p \beta_k} \hat{\mathbf{y}}_{r,k}. \quad (3.4)$$

$$\mathbb{E}\{(\mathbf{h}_k - \hat{\mathbf{h}}_k)(\mathbf{h}_k - \hat{\mathbf{h}}_k)^H\} = \left(\beta_k - \frac{p_p \tau_p \beta_k^2}{p_p \tau_p \beta_k + \sigma^2}\right) \mathbf{I}_M. \quad (3.5)$$

$$\xi_k = \frac{p_p \tau_p \beta_k^2}{p_p \tau_p \beta_k + \sigma^2}. \quad (3.6)$$

3.4 Truyền tín hiệu chiều xuống

Sau giai đoạn ước lượng kênh, trạm gốc BS chọn một tập thuê bao phục vụ $\mathbb{K}_s = \{1, 2, \dots, K_s\}$ từ tập thuê bao đã được ước lượng kênh $\mathbb{K}_p = \{1, 2, \dots, K_p\}$ với $K_s \leq K_p$ để phục vụ trong giai đoạn truyền tín hiệu chiều xuống. Ký hiệu $\mathbf{x} \in \mathbb{C}^{K_s \times 1}$ là dữ liệu của K_s thuê bao với giới hạn $\mathbb{E}\{\|\mathbf{x}\|^2\} = 1$. Trạm gốc BS tính toán ma trận tiền mã hóa tuyến tính $\mathbf{F} \in \mathbb{C}^{M \times K_s}$ từ ước lượng kênh $\hat{\mathbf{H}}$ để ghép dữ liệu \mathbf{x} đến các ăng ten. Công suất của thuê bao k là p_k chịu điều kiện giới hạn về công suất tổng $\sum_{k=1}^{K_s} |\mathbf{f}_k|^2 p_k \leq P$. Tín hiệu nhận được tại thuê bao k có thể

được viết lại như sau:

$$\begin{aligned}
\mathbf{y}_k &= \mathbf{h}_k^T \mathbf{f}_k \sqrt{p_k} x_k + \sum_{l=1, l \neq k}^{K_s} \mathbf{h}_k^T \mathbf{f}_l \sqrt{p_l} x_l + \mathbf{n}_k \\
&= \mathbb{E}\{\mathbf{h}_k^T \mathbf{f}_k\} \sqrt{p_k} x_k \\
&\quad + \sum_{l=1}^{K_s} \mathbf{h}_k^T \mathbf{f}_l \sqrt{p_l} x_l - \mathbb{E}\{\mathbf{h}_k^T \mathbf{f}_k\} \sqrt{p_k} x_k + \mathbf{n}_k,
\end{aligned} \tag{3.7}$$

trong đó \mathbf{h}_k là véc tơ kênh truyền của thuê bao k , và \mathbf{f}_k là cột thứ k của ma trận \mathbf{F} .

Tỷ lệ tín hiệu trên nhiễu tức thời (SINR) của thuê bao k có thể được viết như sau:

$$\gamma_k = \frac{p_k |\mathbb{E}\{\mathbf{h}_k^T \mathbf{f}_k\}|^2}{\sum_{l=1}^{K_s} p_l \mathbb{E}\{|\mathbf{h}_k^T \mathbf{f}_l|^2\} - p_k |\mathbb{E}\{\mathbf{h}_k^T \mathbf{f}_k\}|^2 + \sigma^2}. \tag{3.8}$$

Tốc độ có thể đạt được của thuê bao k là

$$\mathbf{R}_k = \log_2(1 + \gamma_k). \tag{3.9}$$

Tốc độ tổng của hệ thống là

$$\mathbf{R}_{sum} = \sum_{l=1}^{K_s} \log_2(1 + \gamma_k). \tag{3.10}$$

3.5 Bài toán tối ưu của kỹ thuật lập lịch

3.5.1 Các tiêu chí tối ưu

- **Tối đa hóa tốc độ:** kỹ thuật kinh điển có mục tiêu tối đa hóa dung lượng tổng của hệ thống được gọi là MR. Dung lượng tổng của K thuê bao trong hệ thống là $\max \sum_{k=1}^K R_k$, trong đó R_k là tốc độ của thuê bao k . Kỹ thuật MR

đạt được với công suất phát cho từng thuê bao được tính theo [57], [59]:

$$p_k(m, f) = \begin{cases} [\frac{1}{\lambda_k} - \frac{N_0}{\|\mathbf{H}_k(m, f)\|^2}]^+, \\ \text{if } \|\mathbf{H}_k(m, f)\|^2 \geq \frac{\lambda_k}{\lambda_l} \|\mathbf{H}_l(m, f)\|^2 \\ 0 \quad \text{trong các trường hợp còn lại,} \end{cases} \quad (3.11)$$

trong đó $[x]^+ = \max(0, x)$, $\mathbf{H}_k(m, f)$ là kênh truyền của thuê bao k tại khối tài nguyên (RB) m trong chu kỳ khung f và λ_k là hằng số được chọn để thỏa mãn yêu cầu về giới hạn công suất:

$$\sum_{k=1}^K p_k \leq P. \quad (3.12)$$

Kết quả tại công thức (3.11) chỉ ra rằng chỉ có các thuê bao với kênh truyền tốt nhất mới được phục vụ. Một biến thể của chiến lược quản lý tài nguyên này được gọi là kỹ thuật lập lịch "Tối đa hóa tốc độ với công suất không đổi" mà trong đó chỉ có các thuê bao với kênh truyền tốt nhất được phục vụ nhưng công suất phát là như nhau cho mọi thuê bao được phục vụ [57]:

$$p_k = \frac{P}{K}. \quad (3.13)$$

- **Công bằng theo tỷ lệ:** Từ quan điểm khách hàng kỹ thuật lập lịch MR là không chấp nhận được vì các thuê bao có kênh truyền kém gần như không bao giờ được phục vụ. Công bằng theo tỷ lệ PF giải quyết nhược điểm này bằng cách quan tâm đến lưu lượng đã sử dụng của thuê bao để cung cấp một tốc độ như nhau cho mọi thuê bao. Kỹ thuật lựa chọn một thuê bao có tiêu chí M cao nhất. Tiêu chí của thuê bao i -th là M_i được tính theo công thức sau [3]:

$$M_i = \arg \max \frac{R_i(t)}{\bar{R}_i(t)}. \quad (3.14)$$

$$\bar{R}_i(t) = (1 - \frac{1}{t_c}) * \bar{R}_i(t-1) + \frac{1}{t_c} * R_i(t-1), \quad (3.15)$$

trong đó,

$R_i(t)$ là tốc độ tức thời có thể đạt được,

$\bar{R}_i(t)$ là tốc độ trung bình của thuê bao i -th ở thời điểm t ,

t_c là kích thước của chu kỳ cập nhật,

$\bar{R}_i(t-1) = 0$ nếu như thuê bao không được chọn để phục vụ ở thời điểm $t-1$.

Có thể nhận thấy rằng kỹ thuật PF cung cấp sự ưu tiên cho không chỉ thuê bao với kênh truyền tốt mà cả với thuê bao có tốc độ sử dụng dịch vụ thấp.

• QoS-Scheduler

Trong [8], Ameigeiras và các tác giả đề xuất thuật toán với tên gọi QoS-Scheduler đã xem xét các yếu tố về độ ưu tiên, độ trễ của lưu lượng cũng như là đảm bảo yêu cầu tốc độ dịch vụ tối thiểu cho lưu lượng GBR. Kỹ thuật cải thiện hiệu năng của hệ thống bằng cách theo dõi độ trễ của thuê bao đạt đến một ngưỡng nhất định D bởi một hàm toán học có biến số là độ trễ:

$$P_k[n, s] = (1 + f(w_k)) \frac{R_k[n, s]}{[r_k[n]]^\alpha}, \quad (3.16)$$

trong đó, $f(w_k) = \frac{1}{1 + e^{-a_k(w_k - D)}}$,

α là hệ số điều khiển độ công bằng,

w_k thể hiện thời gian chờ của thuê bao k .

Tham số a_k điều khiển độ dốc của hàm số.

Tham số D thiết lập ngưỡng trên của độ trễ.

Sau đó, nó nhân với độ ưu tiên của lưu lượng $F_k^{QCI_m}$:

$$P_k^{QCI_m}[n, s] = P_k[n, s] F_k^{QCI_m}. \quad (3.17)$$

Kỹ thuật để xác định $F_k^{QCI_m}$ phụ thuộc vào chỉ báo hiệu năng chất lượng Q_k của mỗi sóng mang vô tuyến. Q_k là một hàm của độ trễ d_k đối với thuê bao k thuộc lớp đảm bảo tốc độ dịch vụ GBR class, và là một hàm của lưu lượng đã phát của thuê bao k thuộc lớp không đảm bảo dịch vụ non-GBR.

$$Q_k[n] = \begin{cases} d_k[n] = (1 - \rho_d)d_k[n - 1] + \rho_d \frac{q_k[n]}{\lambda_k} \\ QCI_m = 1, 2, 3, 4 \\ r_k[n] = (1 - \rho_r)r_k[n - 1] + \rho_r r_k[n] \\ QCI_m = 5, 6, 7, 8, 9, \end{cases} \quad (3.18)$$

trong đó, $q_k[n]$ ký hiệu số lượng bit đang ở trong hàng đợi của sóng mang k tại khung thời gian n .

λ_k là ước lượng tốc độ trung bình của sóng mang k .

$r_k[n]$ đại diện cho tốc độ dữ liệu đã truyền trong khung thời gian n của sóng mang k .

ρ_d và ρ_r là hằng số.

Hàm số trong (3.18) không kiểm tra tốc độ yêu cầu tối thiểu của thuê bao GBR, điều này có thể dẫn đến một thuê bao GBR có lưu lượng lớn sẽ sử dụng hết băng thông trong khi nên chia sẻ với các thuê bao non-GBR.

3.5.2 Xây dựng hàm mục tiêu

Trong chồng giao thức của mặt phẳng thuê bao của 5G, một lớp mới có tên là Service Data Adaptation Protocol - giao thức thích nghi dữ liệu dịch vụ (SDAP) được thêm vào trên lớp PDCP để ghép các luồng QoS và các sóng mang vô tuyến [1]. Đây là lần đầu tiên, QoS được quan tâm đầy đủ ở lớp vật lý là các sóng mang vô tuyến. Tốc độ kết nối là một trong các yếu tố hàng đầu của yêu cầu về chất

lượng dịch vụ QoS. Kỹ thuật lập lịch đóng vai trò quan trọng trong việc đảm bảo QoS cho các ứng dụng của mạng 5G.

Thông tin trạng thái kênh truyền $\hat{\mathbf{H}}$, tổng công suất phát P và yêu cầu về tốc độ T_k của thuê bao k được thu thập tại trạm gốc BS. Từ những thông tin này, tập thuê bao con tối ưu nhất \mathbb{K}_s được chọn từ tập thuê bao được ước lượng kênh \mathbb{K}_p sao cho tối đa hóa được tốc độ tổng của hệ thống trong mỗi chu kỳ khung.

Mục tiêu của kỹ thuật đề xuất là xác định công suất phát tối thiểu mà vẫn đáp ứng được yêu cầu về tốc độ tối thiểu của mỗi thuê bao cần được xác định. Nhờ đó, có thể dành phần công suất còn lại để truyền dữ liệu cho các thuê bao khác, tức là sẽ làm tăng dung lượng tổng của hệ thống:

$$\begin{aligned} & \max_{\{K_s\}} \sum_{k=1}^{K_s} \log_2(1 + \gamma_k) \\ & s.t \quad \sum_{k=1}^{K_s} |\mathbf{f}_k|^2 p_k \leq P \\ & \quad \log_2(1 + \gamma_k) \geq T_k \end{aligned} \quad (3.19)$$

3.6 Giải pháp đề xuất

Đầu tiên, công suất vừa đủ đáp ứng yêu cầu về tốc độ của mỗi khách hàng sẽ được xác định. Ở đây, giải pháp tiền mã hóa MRT sẽ được sử dụng:

$$R_k = \log_2(1 + \gamma_k) \geq T_k. \quad (3.20)$$

$$\gamma_k \geq 2^{T_k} - 1. \quad (3.21)$$

$$\frac{p_k |\mathbb{E}\{\mathbf{h}_k^T \mathbf{f}_k\}|^2}{\sum_{l=1}^{K_s} p_l \mathbb{E}\{|\mathbf{h}_k^T \mathbf{f}_l|^2\} - p_k |\mathbb{E}\{\mathbf{h}_k^T \mathbf{f}_k\}|^2 + \sigma^2} \geq 2^{T_k} - 1. \quad (3.22)$$

Công suất p_k trong (3.19) được tính theo công thức nghịch đảo kênh truyền [14]:

$$p_k = \frac{P}{K_s |\mathbf{f}_k|^2}. \quad (3.23)$$

Ma trận tiền mã hóa MRT là

$$\mathbf{f}_k^{MRT} = \hat{\mathbf{h}}_k^*. \quad (3.24)$$

Từ đó, ta có

$$\mathbb{E}\{\mathbf{h}_k^T \mathbf{f}_k\} = \xi_k M. \quad (3.25)$$

$$\mathbb{E}\{\|\mathbf{h}_k^T \mathbf{f}_k\|^2\} = \xi_k^2 M^2 + \xi_k \beta_k M. \quad (3.26)$$

$$\mathbb{E}\{\|\mathbf{h}_k^T \mathbf{f}_l\|^2\} = \xi_l \beta_k M. \quad (3.27)$$

Từ (3.22) đến (3.27), ta có

$$\gamma_k = \frac{p_k \xi_k^2 M^2}{\sum_{l=1}^{K_s} p_l \xi_l \beta_k M + \sigma^2} \quad (3.28)$$

$$\geq \frac{p_k \xi_k^2 M^2}{\sum_{l=1}^{K_s} p_l \beta_l \beta_k M + \sigma^2} \quad (3.29)$$

$$= \frac{M^2 \left(\frac{\tau_p p_p \beta_k}{\sigma^2 + \tau_p p_p \beta_k}\right) \beta_k \frac{P}{K_s}}{\sum_{l=1}^{K_s} M \beta_k \frac{P}{K_s} + \sigma^2} \quad (3.30)$$

$$= \frac{\left(\frac{\tau_p p_p \beta_k}{\sigma^2 + \tau_p p_p \beta_k}\right) \beta_k M^2 P}{K_s (\beta_k M P + \sigma^2)} \geq 2^{T_k} - 1 \quad (3.31)$$

$$\frac{\left(\frac{\tau_p p_p \beta_k}{\sigma^2 + \tau_p p_p \beta_k}\right) \beta_k M^2 P}{(2^{T_k} - 1)(\beta_k M P + \sigma^2)} \geq K_s. \quad (3.32)$$

Rõ ràng rằng tốc độ tối thiểu của thuê bao k chỉ phụ thuộc vào các tham số β_k , K_s và P , không phụ thuộc vào các thuê bao khác. Vì thế nghiên cứu sinh có thể mở rộng kết quả (3.28) cho trường hợp các thuê bao có tốc độ tối thiểu khác nhau. Ở đây, nghiên cứu sinh có thể tính công suất p_k của thuê bao k bằng cách điều chỉnh tham số K_s trong công thức (3.23) để đạt được tốc độ T_k .

Tập thuê bao K_s được chọn bằng cách lựa chọn các thuê bao có tiêu chí cao nhất. Tiêu chí của thuê bao k là một hàm số của phụ thuộc vào công suất sử dụng

của thuê bao p_k , chất lượng kênh truyền h_k , và độ ưu tiên của lưu lượng ψ_k tại khung n :

$$\rho_k[n] = (1 + f_w(k)) \frac{|\mathbf{h}_k|^2 T_k[n]}{\psi_k^\alpha p_k \bar{R}_k[n]}, \quad (3.33)$$

trong đó, α là một hệ số dùng để điều chỉnh độ ưu tiên của lớp.

$T_k[n]$ là tốc độ truyền dẫn mong muốn của thuê bao tại khung n .

$\bar{R}_k[n]$ là dữ liệu trung bình đã truyền trong quá khứ của thuê bao k tại khung n .

Ví dụ, sau khi thuê bao k được chọn thì phần công suất còn lại sẽ bị giảm đi $f_k^2 p_k$. Quá trình này lặp đi lặp lại cho đến khi không còn đủ công suất cho thuê bao hoặc không còn thuê bao cần phục vụ. Chi tiết hoạt động của kỹ thuật được miêu tả chi tiết trong thuật toán 3.

Algorithm 3 Thuật toán QoS-Assurance

- 1: Trạm gốc BS khởi tạo $\mathcal{S}(1 : K_p) = 0, \mathbb{K}_p = \{1, 2, \dots, K_p\}, i = 1$
 - 2: Tính toán $f_w(k), K_s, p_k, \rho_k[n]$ for $\forall k \in \mathbb{K}_p$ in 3.16, 3.32, 3.23, 3.33
 - 3: Lựa chọn thuê bao thứ i :
 - $stop = 0;$
 - $\rho_i[n] = \operatorname{argmax} \rho_k[n], \forall k \in \mathbb{K}_p$
 - 4: **if** $P \geq \|\mathbf{f}_i\|^2 p_i$ **then**
 - $\mathbb{K}_p = \mathbb{K}_p \setminus i;$
 - $stop = 1;$
 - $P = P - \|\mathbf{f}_i\|^2 p_i;$
 - $\mathcal{S}(i) = 1;$
 - 5: **end if**
 - 6: **if** $stop = 1$ **then**
 - Quay lại bước 3
 - 7: **else**
 - Thuật toán kết thúc.
 - 8: **end if**
-

3.7 Kết quả mô phỏng

Để đánh giá hiệu quả của kỹ thuật lập lịch QoS-Assurance, một số trường hợp điển hình được mô phỏng bằng phần mềm Matlab để so sánh hiệu năng giữa các

kỹ thuật sau:

- Kỹ thuật MR.
- Kỹ thuật QoS Scheduler trong [8].
- Kỹ thuật QoS-Assurance.

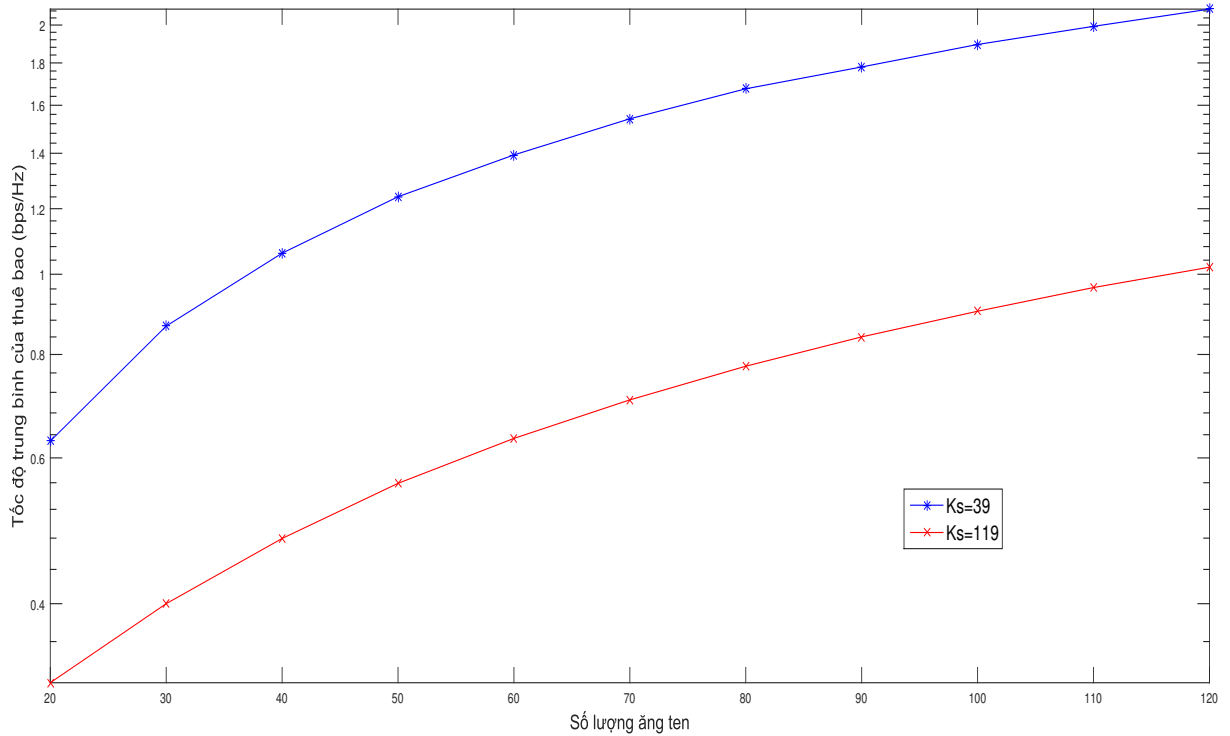
Bảng 3.1 liệt kê những tham số mô phỏng chính. Ở đây các thuê bao được chia thành hai lớp dịch vụ có mong muốn tốc độ tối thiểu khác nhau. Lớp thứ nhất với tốc độ 2 (bit/s/Hz) tương đương với tốc độ 200 Mbps và lớp thứ hai có tốc độ tối thiểu 1 (bit/s/Hz) tương đương với tốc độ 100 Mbps khi trạm BTS sử dụng băng thông 100 MHz. Các kịch bản mô phỏng là tăng dần số ăng ten để kiểm tra sự hội tụ về tốc độ mong muốn của thuê bao trong từng lớp, so sánh tốc độ tổng của hệ thống giữa hai lớp dịch vụ khác nhau, và so sánh dung lượng hữu ích của kỹ thuật đề xuất và kỹ thuật MR khi số lượng thuê bao phục vụ tăng dần.

Bảng 3.1: Tham số mô phỏng

Tham số mô phỏng	
Tham số	Giá trị
Số lượng lớp C	2
Tốc độ tối thiểu của lớp thuê bao lớp 1	2 (bit/s/Hz)
Tốc độ tối thiểu của lớp thuê bao lớp 2	1 (bit/s/Hz)

Hình 3.1 thể hiện tốc độ mỗi thuê bao khi sử dụng kỹ thuật QoS-Assurance khi số ăng ten của trạm gốc tăng dần. Trong trường hợp này, nghiên cứu sinh giả định rằng tất cả thuê bao trong tế bào đều được ước lượng kênh và phục vụ để xem tốc độ trung bình mà mỗi thuê bao sẽ đạt được. Nghiên cứu sinh nghiên cứu hai trường hợp: một là $K_p = K_s = 39$, hai là $K_p = K_s = 119$. Chúng ta có thể thấy từ kết quả, với $K_s = 39$ tốc độ trung bình mỗi thuê bao sẽ tiến tới 2 (bit/s/Hz). Với $K_s = 119$ tốc độ trung bình mỗi thuê bao sẽ tiến tới 1 (bit/s/Hz). Kết quả

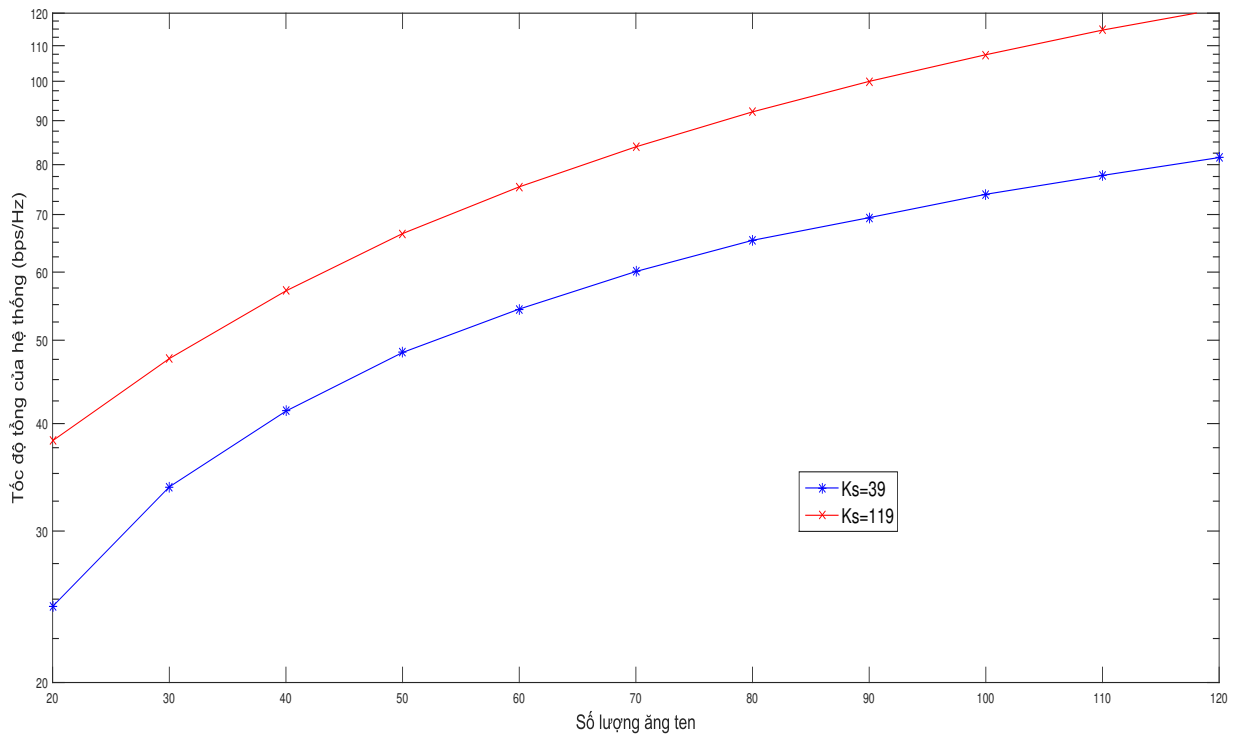
này khẳng định kỹ thuật QoS-Assurance có thể đảm bảo bất kỳ tốc độ tối thiểu nào cho thuê bao bằng cách điều khiển số lượng thuê bao được phục vụ K_s .



Hình 3.1: Tốc độ mỗi thuê bao khi sử dụng kỹ thuật QoS-Assurance

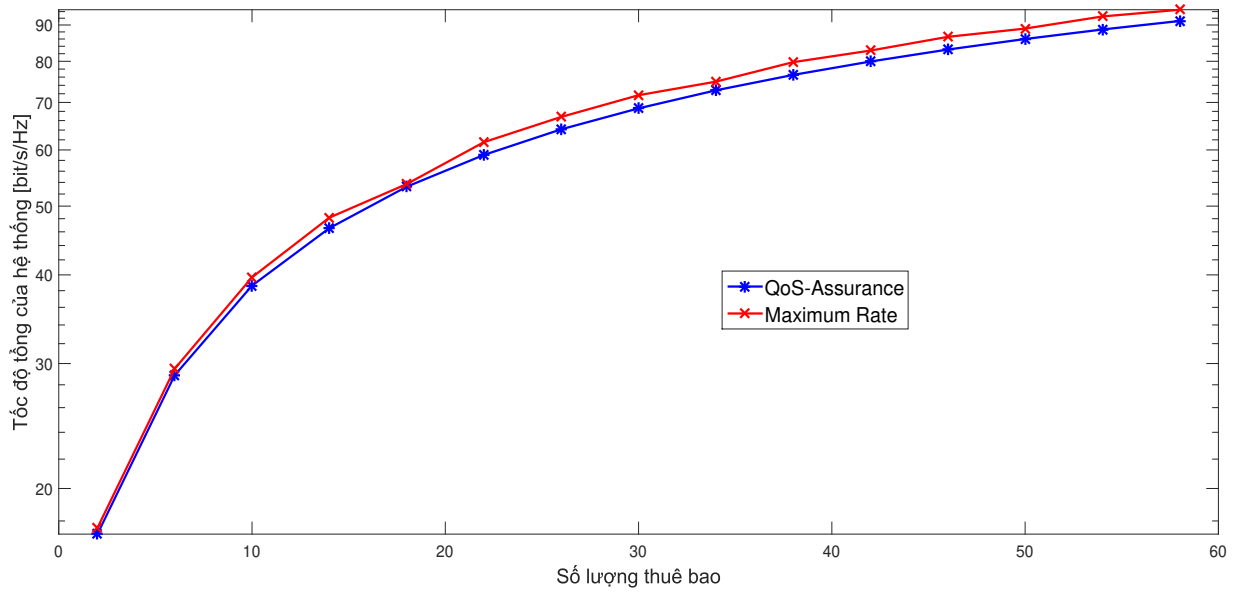
Hình 3.2 thể hiện tốc độ tổng của hệ thống khi sử dụng kỹ thuật QoS-Assurance. Ở đây có thể thấy từ hình vẽ rằng tốc độ tổng của hệ thống sẽ tăng dần khi tăng số lượng thuê bao được phục vụ K_s . Khi số lượng thuê bao được phục vụ $K_s = 119$, tốc độ tổng của hệ thống đạt được khoảng 120 (bit/s/Hz). Nó sẽ giảm khoảng 30% xuống khoảng 80 (bit/s/Hz) khi $K_s = 39$. Như vậy càng phục vụ nhiều thuê bao thì tốc độ tổng của hệ thống sẽ tăng lên nhưng tốc độ trung bình mỗi thuê bao sẽ giảm đi.

Từ lúc này, luận án sẽ nghiên cứu vấn đề đảm bảo tốc độ tối thiểu của hai thuật toán QoS-Assurance và MR. Ở đây, tổng số thuê bao $K_p = 58$ được chia thành hai lớp dịch vụ. Lớp thứ nhất bao gồm 29 thuê bao với yêu cầu tốc độ tối thiểu là 2 (bit/s/Hz) và thuê bao thuộc lớp thứ hai yêu cầu tốc độ tối thiểu



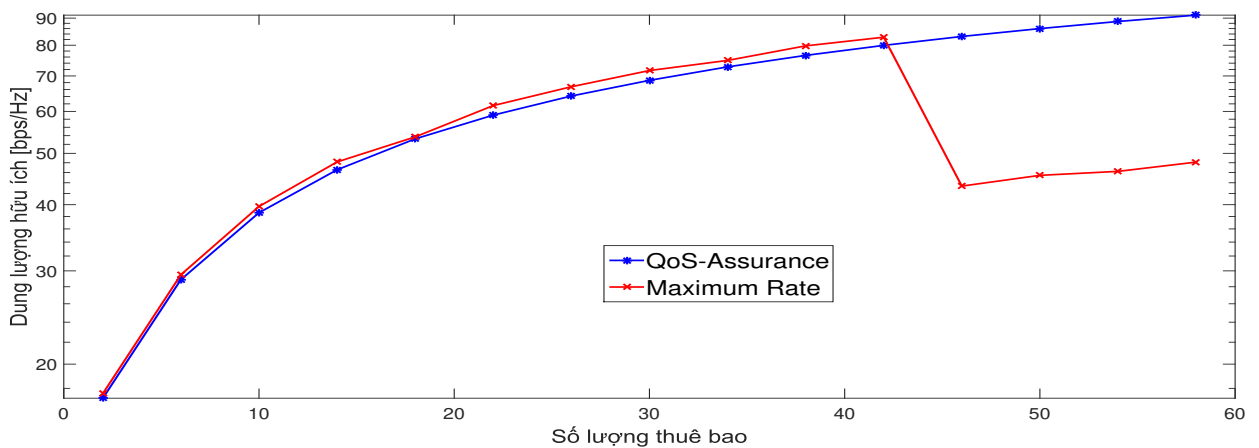
Hình 3.2: Tốc độ tổng của hệ thống

là 1 (bit/s/Hz). Kỹ thuật MR sử dụng công suất giống nhau cho mọi thuê bao $p_k = \frac{P}{K_t}$. Hình 3.3 thể hiện tốc độ tổng của kỹ thuật MR luôn cao hơn tổng tốc độ của kỹ thuật QoS-Assurance.



Hình 3.3: So sánh tổng tốc độ

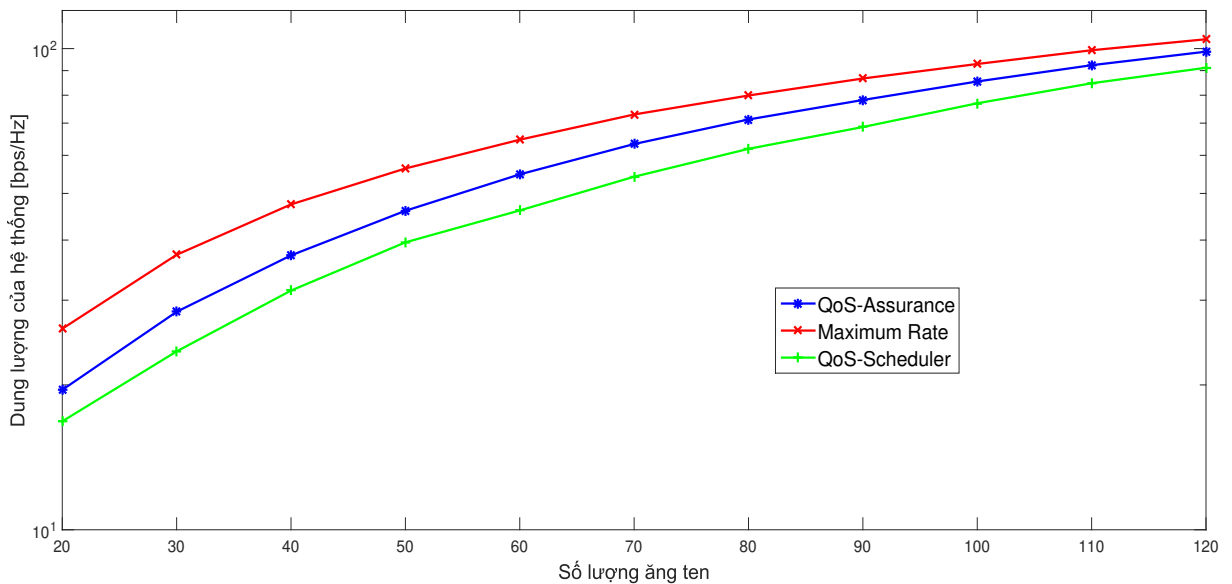
Hình 3.4 thể hiện dung lượng hữu ích của các thuê bao tức là chỉ có dung lượng mà đáp ứng được về tốc độ tối thiểu của các thuê bao được tính. Đối với kỹ thuật MR, có thể thấy rằng khi $M = 120, K_p \geq 40$, tổng dung lượng hệ thống là hữu hạn trong khi số thuê bao phục vụ đã tăng đủ lớn thì phần lớn số thuê bao sẽ nhận được tốc độ nhỏ hơn 2 (bit/s/Hz). Vì vậy hầu hết thuê bao được phục vụ của lớp thứ nhất sẽ không có dung lượng hữu ích. Kết quả là, dung lượng hữu ích của kỹ thuật MR sẽ giảm mạnh trước khi tăng nhẹ trở lại do số thuê bao thuộc lớp thứ hai được phục vụ tăng lên. Bên cạnh đó, các thuê bao thuộc lớp thứ nhất sẽ được thỏa mãn về tốc độ yêu cầu 2 (bit/s/Hz) và các thuê bao thuộc lớp thứ hai được thỏa mãn yêu cầu 1 (bit/s/Hz) khi sử dụng kỹ thuật QoS-Assurance. Vì vậy mọi dung lượng khi sử dụng kỹ thuật QoS-Assurance đều là dung lượng hữu ích. Điều đó cho thấy kỹ thuật MR cung cấp tốc độ tổng lớn nhất nhưng kỹ thuật QoS-Assurance có dung lượng hữu ích cao hơn, cải thiện được độ công bằng giữa các lớp dịch vụ, và đặc biệt là đảm bảo được các tốc độ tối thiểu khác nhau của thuê bao.



Hình 3.4: So sánh dung lượng hữu ích giữa hai kỹ thuật

Các trường hợp trên chủ yếu nhấn mạnh ưu điểm của kỹ thuật điều khiển công suất trong kỹ thuật đề xuất. Trường hợp cuối cùng thể hiện ảnh hưởng của các

tham số về độ ưu tiên lên dung lượng hệ thống. Hình 3.5 thể hiện so sánh về dung lượng hệ thống giữa ba kỹ thuật : QoS-Assurance, MR và QoS-Scheduler. Kỹ thuật điều khiển công suất được áp dụng cho cả ba kỹ thuật. Sự khác biệt ở đây chỉ còn do kỹ thuật đề xuất QoS-Assurance có độ ưu tiên về dung lượng. Từ tập hợp $K_p = 100$ thuê bao với một nửa số thuê bao thuộc lớp thứ nhất và nửa còn lại thuộc lớp thứ hai, một tập con gồm K_s thuê bao sẽ được chọn để phục vụ. Có thể thấy rằng thuật toán MR có dung lượng cao nhất, kỹ thuật QoS-Assurance đứng ở vị trí thứ hai và cuối cùng là QoS-Scheduler. Với kỹ thuật QoS-Assurance, sự đánh đổi giữa dung lượng và độ công bằng phụ thuộc vào tham số độ ưu tiên. Dung lượng kỹ thuật QoS-Assurance sẽ tiệm cận với dung lượng của kỹ thuật MR nếu hệ số điều chỉnh độ ưu tiên α tiến tới không với nhược điểm là dung lượng của các thuê bao thuộc lớp thứ nhất sẽ tiến tới không. Nói một cách khác là độ công bằng về dung lượng giữa các lớp sẽ bị giảm đi.



Hình 3.5: So sánh về dung lượng giữa ba kỹ thuật

3.8 Kết luận chương 3

Trong chương này, luận án đề xuất kỹ thuật lập lịch cho phép thuê bao có thể đạt được tốc độ tối thiểu tùy ý, và trạm gốc có thể phục vụ các thuê bao có tốc độ tối thiểu khác nhau trong cùng một chu kỳ khung. Đây là điểm mới so với các kỹ thuật lập lịch khác chỉ có thể phục vụ một tốc độ tối thiểu duy nhất. Tính khả thi của kỹ thuật lập lịch đề xuất được chứng minh bằng cách thay đổi công suất của từng thuê bao kỹ thuật có thể đảm bảo tốc độ tối thiểu cho thuê bao mà hoàn toàn không phụ thuộc vào tốc độ của các thuê bao khác. Và thực nghiệm cũng cho thấy để phục vụ các thuê bao yêu cầu tốc độ tối thiểu cao sẽ làm giảm dung lượng hệ thống. Nếu giảm dần sự khác nhau về độ ưu tiên thì thuật toán của hệ thống sẽ tiến tới dung lượng của thuật toán tối đa hóa tốc độ hệ thống. Hạn chế của kỹ thuật đề xuất là dung lượng tổng của hệ thống vẫn kém hơn so với kỹ thuật lập lịch MR.

Chương 4

Kỹ thuật lập lịch cho hệ thống ăng ten bước sóng mm sử dụng các bộ ADC phân giải thấp

Trong chương này luận án sẽ trình bày mô hình hệ thống ăng ten cỡ rất lớn bước sóng mm, mô hình lượng tử hóa, phương pháp tiền mã hóa hai giai đoạn và tốc độ có thể đạt được. Từ đó đề xuất kỹ thuật lập lịch sử dụng hệ số trực giao của kênh truyền giữa các thuê bao để lựa chọn thuê bao một cách nhanh và tối ưu hơn. Đây là điểm khác biệt so với kỹ thuật lập lịch truyền thống như MR phải sử dụng thuật toán vét cạn [57, 38].

4.1 Ý tưởng kỹ thuật lập lịch cho hệ thống ăng ten cỡ rất lớn bước sóng mm khi sử dụng các bộ ADC phân giải thấp

Hệ thống thông tin di động MIMO cỡ rất lớn có bước sóng cỡ mm là một kỹ thuật có nhiều lợi thế để đạt được dung lượng lớn cho các mạng thông tin di động 5G [50, 9, 16]. Ưu điểm về băng thông rất rộng trong miền tần số có bước sóng cỡ mm hấp dẫn các nghiên cứu giải quyết các thách thức của hệ thống mmWave để triển khai trong thực tế [51, 19, 28]. Truyền sóng vô tuyến trong băng sóng cỡ mm có tính định hướng rất cao cùng với sự suy hao rất lớn và có ít thành phần đa

đường. Để bù đắp cho sự tổn hao lớn thì cần tạo ra các búp sóng định hướng cao nhờ có số lượng ăng ten rất lớn. Đi cùng tốc độ bit cao là các bộ biến đổi tương tự số ADC tiêu thụ phần lớn năng lượng của máy thu. Vì thế, nhiều nghiên cứu tập trung vào cách sử dụng các bộ ADC có độ phân giải thấp trong các hệ thống mmWave [26].

Rất nhiều nghiên cứu đã chỉ ra rằng nhờ số lượng ăng ten cỡ rất lớn có thể sử dụng các bộ ADC phân giải thấp trong hệ thống mmWave. Việc ước lượng kênh truyền có thể thực hiện được trong các hệ thống mmWave sử dụng các bộ ADC độ phân giải chỉ vài bit [54, 44, 45]. Hệ thống mmWave với kiến trúc thay đổi số bit ADC có hiệu năng về công suất tối ưu hơn các hệ thống sử dụng số bit ADC cố định [21].

Hệ thống ăng ten cỡ rất lớn bước sóng mmWave khi sử dụng các bộ ADC phân giải thấp đặt ra một số vấn đề cho kỹ thuật lập lịch ví dụ như là sự quá tải và sự không chính xác về ước lượng CSI. Một trong những cách vượt qua gánh nặng về thông tin ước lượng kênh là kỹ thuật lập lịch sử dụng búp sóng ngẫu nhiên [38]. Nghiên cứu này đã chỉ ra rằng dung lượng hệ thống tỷ lệ với số ăng ten của trạm phát nếu số thuê bao cũng tăng theo tương ứng. Hướng nghiên cứu cũng được tiếp tục phát triển cho trường hợp số lượng thuê bao thừa thớt bằng cách lựa chọn búp sóng tối ưu trong bài [39]. Mặt khác, các kỹ thuật lập lịch truyền thống như MR hay PF không hoàn toàn phù hợp với các hệ thống mmWave do đây là các thuật toán vét cạn trong khi số lượng ăng ten và thuê bao là rất lớn [39]. Các kỹ thuật này cũng không khai thác các đặc tính về kênh truyền trong hệ thống mmWave. Hơn nữa, khi các bộ ADC phân giải thấp được sử dụng thì kỹ thuật lập lịch cần cân nhắc tới sai số của quá trình lượng tử hóa. Một kỹ thuật lập lịch cho chiều lên khai thác đặc tính kênh truyền để giảm thiểu thời gian tìm kiếm tập thuê bao

tối ưu trong hệ thống mmWave sử dụng các bộ ADC phân giải thấp được đề xuất trong [23]. Tuy nhiên kỹ thuật lập lịch cho chiều xuống trong hệ thống mmWave sử dụng các bộ ADC phân giải thấp vẫn còn là dấu hỏi.

Trong phần nghiên cứu này, luận án đề xuất kỹ thuật lập lịch NOUS cho chiều xuống của hệ thống mmWave hướng đến lựa chọn một tập thuê bao tối ưu bằng cách sử dụng đặc tính bán trực giao của kênh truyền và có thể tiệm cận với hiệu năng của thuật toán MR trong khi giảm thiểu được thời gian tìm kiếm. Số lượng ăng ten của thuê bao là $M_{MS} > 1$ mang nhiều ý nghĩa thực tế hơn trường hợp thuê bao có một ăng ten trong nghiên cứu [23]. Kỹ thuật đề xuất được chia làm ba giai đoạn: lựa chọn thuê bao tốt nhất trong tập thuê bao ứng cử viên, tính toán thành phần trực giao của thuê bao so với tập thuê bao đã được lựa chọn, bước cuối cùng là loại bỏ các thuê bao trong tập ứng cử viên mà không bán trực giao với thuê bao vừa được lựa chọn. Bước cuối cùng có ý nghĩa rất quan trọng trong việc giảm thiểu thời gian tìm kiếm tập thuê bao tối ưu do giảm nhanh kích thước của tập thuê bao được lựa chọn. Hơn nữa, kết quả mô phỏng cho thấy kỹ thuật NOUS có thể tiệm cận hiệu năng của kỹ thuật MR và đánh bại hiệu năng của kỹ thuật PF mà lại có độ phức tạp ít hơn.

4.2 Mô hình hệ thống

Luận án xem xét một hệ thống đơn tế bào đa người dùng có một trạm gốc và \mathbb{K}_a thuê bao. Trạm gốc BS được trang bị M_{BS} ăng ten và M_{RF} bộ RF. Mỗi thuê bao có M_{MS} ăng ten, và giả định rằng chỉ có một luồng dữ liệu với trạm gốc. Hơn nữa, ở đây giả thiết rằng trạm gốc BS chỉ lựa chọn tập thuê bao $\mathbb{K}_s = \{1, 2, \dots, K_s\}$ từ \mathbb{K}_a thuê bao để phục vụ. Vì thế tổng số luồng dữ liệu là M_s , và $M_s \leq M_{RF} \leq M_{BS}$.

4.2.1 Truyền dẫn chiều xuống

Trạm gốc BS sử dụng tiền mã hóa bằng cơ sở $\mathbf{F}_{\text{BB}} = [\mathbf{f}_1^{\text{BB}}, \mathbf{f}_2^{\text{BB}}, \dots, \mathbf{f}_{M_S}^{\text{BB}}]$ với kích thước $M_S \times M_S$ theo sau là tiền mã hóa RF $\mathbf{F}_{\text{RF}} = [\mathbf{f}_1^{\text{RF}}, \mathbf{f}_2^{\text{RF}}, \dots, \mathbf{f}_{M_S}^{\text{RF}}]$ với kích thước $M_{\text{BS}} \times M_S$. Vì vậy, tín hiệu truyền đi là

$$\mathbf{x} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s}, \quad (4.1)$$

trong đó $\mathbf{s} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{M_S}]^T$ là véc tơ của tín hiệu phát có $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{1}{M_S}\mathbf{I}_{M_S}$. Bởi vì bộ RF sử dụng các bộ dịch pha tương tự nên các giá trị của nó là hằng số thỏa mãn $|[\mathbf{F}_{\text{RF}}]_{m,n}|^2 = \frac{1}{M_{\text{BS}}}$. Hay nói một cách khác, $[\mathbf{F}_{\text{RF}}]_{m,n} = \frac{1}{\sqrt{M_{\text{BS}}}}e^{j\phi_{m,n}}$, trong đó $\phi_{m,n}$ là góc pha được lượng tử hóa. Để thỏa mãn giới hạn về công suất $|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}|^2 = M_S$.

Mô hình kênh truyền là fa đình khối hẹp, thuê bao u sẽ nhận được tín hiệu [11, 6, 18, 62, 7]

$$\mathbf{y}_u = \sqrt{\rho}\mathbf{w}_u^H\mathbf{H}_u\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{w}_u^H\mathbf{n}_u, \quad (4.2)$$

trong đó \mathbf{H}_u với kích thước $M_{\text{MS}} \times M_{\text{BS}}$ là ma trận kênh truyền giữa trạm gốc BS và thuê bao u , $\mathbf{n}_u \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ là nhiễu Gauss, ρ thể hiện công suất phát và $\mathbf{w}_u \in \mathbb{C}^{M_{\text{MS}} \times 1}$ là ma trận kết hợp RF. Luận án giả định rằng kênh truyền gồm L_u tín hiệu tán xạ cho thuê bao u [52, 56]. Vì thế, kênh truyền \mathbf{H}_u có thể được biểu diễn như sau [55]:

$$\mathbf{H}_u = \sqrt{\frac{M_{\text{BS}}M_{\text{MS}}}{L_u}} \sum_{l=1}^{L_u} \alpha_{u,l} a_{\text{MS}}(\theta_{u,l}) a_{\text{BS}}^H(\phi_{u,l}), \quad (4.3)$$

trong đó $\alpha_{u,l}$ là độ lợi của đường thứ l , $\theta_{u,l}$ và $\phi_{u,l} \in [0, 2\pi]$ là góc tới và góc phát (AoAs/AoDs) của đường thứ l , $a_{\text{MS}}(\theta_{u,l})$ và $a_{\text{BS}}(\phi_{u,l})$ là véc tơ đáp ứng mảng ăng ten của trạm gốc BS và thuê bao u . Ở đây giả định ăng ten mảng tuyến tính

(ULA) được sử dụng, véc tơ đáp ứng được định nghĩa như sau [55]

$$\mathbf{a}_{\text{MS}}(\theta_{u,l}) = \sqrt{\frac{1}{M_{\text{MS}}}} [1, e^{j2\pi\epsilon}, \dots, e^{j2\pi(N_{\text{MS}}-1)\epsilon}]^T, \quad (4.4)$$

$$\mathbf{a}_{\text{BS}}(\phi_{u,l}) = \sqrt{\frac{1}{M_{\text{BS}}}} [1, e^{j2\pi\varepsilon}, \dots, e^{j2\pi(N_{\text{BS}}-1)\varepsilon}]^T, \quad (4.5)$$

trong đó $\epsilon = \frac{d \sin(\theta_{u,l})}{\lambda}$, $\varepsilon = \frac{d \sin(\phi_{u,l})}{\lambda}$, λ là bước sóng của tín hiệu, và d là cự ly giữa các chấn tử của ăng ten.

Ký hiệu kênh truyền hiệu dụng:

$$\bar{\mathbf{h}}_u = \mathbf{w}_u^H \mathbf{H}_u \mathbf{F}_{\text{RF}} \quad (4.6)$$

$$\bar{\mathbf{H}} = [\bar{\mathbf{h}}_1^T, \bar{\mathbf{h}}_2^T, \dots, \bar{\mathbf{h}}_{M_S}^T]^T \quad (4.7)$$

4.2.2 Mô hình lượng tử hóa

Sau bộ kết hợp tương tự, từng thành phần thực và ảo của tín hiệu đầu ra \mathbf{Y} được lượng tử hóa ở bộ ADC. Luận án sử dụng mô hình nhiễu lượng tử hóa (AQNM) để biểu diễn quá trình lượng tử hóa theo mô hình tuyến tính. Mô hình này chính xác cả trong phạm vi SINR thấp và trung bình [48]. Tín hiệu lượng tử hóa \mathbf{Z} là

$$\mathbf{Z} = \vartheta(\mathbf{Y}) = \alpha \sqrt{\rho} \bar{\mathbf{H}} \mathbf{F}_{\text{BB}} \mathbf{s} + \alpha \mathbf{W}^H \mathbf{n} + \mathbf{q}, \quad (4.8)$$

trong đó $\vartheta(\cdot)$ là hàm lượng tử cho từng thành phần thực và ảo, $\alpha = 1 - \beta$ là độ lợi của quá trình lượng tử hóa, β được tính xấp xỉ như sau $\beta \approx \frac{\pi\sqrt{3}}{2} 2^{-2b}$ với số bit ADC $b > 5$, và các giá trị β cho $b \leq 5$ được liệt kê trong bảng 1 tại [22]. Nhiễu $\mathbf{W}^H \mathbf{n}$ tuân theo phân bố $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ bởi vì \mathbf{W} là ma trận đơn vị. Tín hiệu nhiễu không có tương quan với \mathbf{R} và $\mathbf{q} \sim \mathcal{CN}(0, \mathbf{R}_{\text{qq}})$ trong đó ma trận tương quan $\mathbf{R}_{\text{qq}} = \alpha\beta \text{diag}(\rho \bar{\mathbf{H}} \mathbf{F}_{\text{BB}} (\bar{\mathbf{H}} \mathbf{F}_{\text{BB}})^H + \mathbf{I})$.

4.2.3 Tiền mã hóa lai đa người dùng hai giai đoạn cho kênh truyền một đường

Trong phần này, luận án xem xét trường hợp $L_u = 1, u = 1, 2, \dots, U$. Vì thế ở đây lược bỏ ký hiệu l trong thông tin kênh truyền. Vì kênh truyền \mathbf{H}_u chỉ có một đường và giả định rằng có thể lái búp sóng một cách liên tục. Vì thế véc tơ tiền mã hóa và kết hợp tối ưu là

$$\mathbf{w}_u = a_{\text{MS}}(\theta_u). \quad (4.9)$$

$$\mathbf{f}_u^{\text{RF}} = a_{\text{BS}}(\phi_u). \quad (4.10)$$

Kênh truyền hiệu dụng của thuê bao u là

$$\bar{\mathbf{h}}_u = \mathbf{w}_u^H \mathbf{H}_u \mathbf{F}_{\text{RF}} \quad (4.11)$$

$$= \sqrt{M_{\text{BS}} M_{\text{MS}}} \alpha_u a_{\text{BS}}^H(\phi_u) \mathbf{F}_{\text{RF}}. \quad (4.12)$$

Vì vậy,

$$\bar{\mathbf{H}} = \mathbf{D} A_{\text{BS}}^H A_{\text{BS}}, \quad (4.13)$$

trong đó \mathbf{D} là ma trận có kích thước $M_{\text{S}} \times M_{\text{S}}$: $[\mathbf{D}]_{u,u} = \sqrt{M_{\text{BS}} M_{\text{MS}}} \alpha_u$.

Trạm gốc sử dụng tiền mã hóa bằng cơ sở

$$\mathbf{F}_{\text{BB}} = \bar{\mathbf{H}}^H (\bar{\mathbf{H}} \bar{\mathbf{H}}^H)^{-1} \Upsilon, \quad (4.14)$$

trong đó Υ là ma trận kích thước $M_{\text{S}} \times M_{\text{S}}$ với $[\Upsilon]_{u,u} = \sqrt{\frac{M_{\text{BS}} M_{\text{MS}}}{(A_{\text{BS}}^H A_{\text{BS}})_{u,u}^{-1}}} |\alpha_u|$, $u = 1, 2, \dots, K_{\text{a}}$. Khi đó tốc độ đạt được cho thuê bao u là

$$\mathcal{R}_u = \log_2 \left(1 + \frac{\alpha^2 \rho (\bar{\mathbf{h}}_u \mathbf{f}_u^{\text{BB}})^2}{\mathbf{R}_{\text{qq}}(u, u) + \alpha^2} \right) \quad (4.15)$$

$$= \log_2 \left(1 + \frac{\alpha^2 \rho M_{\text{BS}} M_{\text{MS}} \alpha_u^2}{\mathbf{R}_{\text{qq}}(u, u) + \alpha^2} \right). \quad (4.16)$$

Tốc độ tổng của hệ thống là

$$\mathcal{R} = \log_2 \left(\det \left(\mathbf{I}_{M_{\text{S}}} + \frac{\alpha^2 \rho (\bar{\mathbf{H}} \mathbf{F}^{\text{BB}})^2}{\mathbf{R}_{\text{qq}} + \alpha^2 \mathbf{I}_{M_{\text{S}}}} \right) \right). \quad (4.17)$$

4.3 Kỹ thuật lập lịch đề xuất

Sử dụng tốc độ đạt được trong các công thức (4.15) và (4.17), kỹ thuật lập lịch được xác định như sau

$$\mathcal{R} = \max_{\mathbb{K}_s \subset \{1,2,\dots,K_a\}} \sum_{u \in \mathbb{K}_s} \mathcal{R}_u. \quad (4.18)$$

Hiệu năng tối ưu của tiền mã hóa ZF trong \mathbf{F}_{BB} đạt được khi kỹ thuật lập lịch lựa chọn được các thuê bao thỏa mãn điều kiện không can nhiễu lẫn nhau $\bar{\mathbf{h}}_u \mathbf{f}_v^{\text{BB}} = 0$ cho $u \neq v$. Xác định tập tối ưu \mathbb{K}_s và \mathbf{F}^{BB} để đạt cả độ lợi ghép kênh và độ lợi đa dạng người dùng là rất khó, đặc biệt khi số lượng thuê bao \mathbb{K}_a là rất lớn [61]. Tuy nhiên, với số lượng lớn thuê bao chúng ta có thể lựa chọn được một tập thuê bao \mathbb{K}_s gần trực giao [65], tốc độ tổng của tập thuê bao này sẽ tiệm cận với hiệu năng tối ưu vì yếu tố đa dạng thuê bao.

Trong nghiên cứu này đề xuất một kỹ thuật lập lịch (NOUS) được miêu tả trong kỹ thuật 4 nhằm tối đa hóa tốc độ tổng ở công thức 4.17 với độ phức tạp ít hơn bằng cách loại bỏ dần các thuê bao theo điều kiện bán trực giao trong [65]. Như có thể thấy trong kỹ thuật 4, kỹ thuật lập lịch bao gồm năm bước: đầu tiên trạm gốc BS đặt toàn bộ thuê bao vào tập ứng cử viên \mathcal{C}_1 và khởi tạo một tập rỗng \mathbb{K}_s cho tập thuê bao được lựa chọn, mỗi vòng lặp thứ i trạm gốc BS sẽ tìm thuê bao tốt nhất thứ i trong tập ứng cử viên \mathcal{C}_i để phục vụ sao cho tối đa hóa tốc độ tổng trong công thức 4.17 và loại bỏ thuê bao được lựa chọn trong tập ứng cử viên \mathcal{C}_i tại bước 2. Ở bước 3, kỹ thuật tính thành phần trực giao $\mathbf{g}_{\mathbb{K}_s(i)}$ của thuê bao thứ i mới được lựa chọn đối với tập cơ sở $\{\mathbf{g}_{\mathbb{K}_s1}, \dots, \mathbf{g}_{\mathbb{K}_s(i-1)}\}$. Tại bước 4, kỹ thuật sẽ loại bỏ các thuê bao trong tập ứng cử viên không thỏa mãn điều kiện trực giao với thuê bao mới được lựa chọn $\mathbb{K}_s(i)$ và cuối cùng kỹ thuật sẽ quay lại bước 2 trừ khi kỹ thuật đã lựa chọn đủ thuê bao hoặc không tìm thấy ứng cử viên

nào phù hợp. Kỹ thuật 4 được trình bày như sau:

Algorithm 4 Kỹ thuật lập lịch bán trực giao NOUS

- 1: Trạm gốc BS khởi tạo $\mathcal{C}_1 = \{1, 2, \dots, \mathbb{K}_a\}$, $\mathbb{K}_s = \emptyset$, $i = 1$
- 2: Sử dụng tốc độ tổng trong 4.17, trạm gốc BS lựa chọn thuê bao thứ i như sau:

$$\mathbb{K}_s(i) = \underset{k \in \mathcal{C}_i}{\operatorname{argmax}} \mathcal{R}([\bar{\mathbf{H}}(\mathbb{K}_s), \bar{\mathbf{h}}_k])$$
 và cập nhật $\mathbb{K}_s = \mathbb{K}_s \cup \mathbb{K}_s(i)$, $\mathcal{C}_i = \mathcal{C}_i \setminus \mathbb{K}_s(i)$
- 3: Tính toán thành phần trực giao của $\bar{\mathbf{h}}_{\mathbb{K}_s(i)}$ đối với tập cơ sở $\{\mathbf{g}_{\mathbb{K}_s(1)}, \dots, \mathbf{g}_{\mathbb{K}_s(i-1)}\}$

$$\mathbf{g}_{\mathbb{K}_s(i)} = \bar{\mathbf{h}}_{\mathbb{K}_s(i)} \left(\mathbf{I} - \sum_{j=1}^{i-1} \frac{\mathbf{g}_{\mathbb{K}_s(j)}^H \mathbf{g}_{\mathbb{K}_s(j)}}{\|\mathbf{g}_{\mathbb{K}_s(j)}\|^2} \right)$$
- 4: Lựa chọn các thuê bao còn lại đáp ứng được yêu cầu trực giao với thuê bao $\mathcal{S}(i)$

$$\mathcal{C}_{i+1} = \{k \in \mathcal{C}_i \mid \frac{|\bar{\mathbf{h}}_k \mathbf{g}_{\mathcal{S}(i)}^H|}{\|\bar{\mathbf{h}}_k\| \|\mathbf{g}_{\mathcal{S}(i)}\|} \leq \xi\} \quad (4.19)$$

- 5: Nếu $i < M_S$ và $\mathcal{C}_{i+1} \neq \emptyset$, cập nhật $i = i + 1$. Ngược lại, kỹ thuật kết thúc.
-

Luận án so sánh kỹ thuật NOUS với MR và PF về phương diện tốc độ tổng của hệ thống. Kỹ thuật MR sẽ thực hiện kỹ thuật vét cạn để lựa chọn thuê bao có tốc độ đạt được cao nhất trong công thức 4.15 trong mỗi vòng lặp. Khi số lượng thuê bao rất lớn, kỹ thuật MR sẽ chịu gánh nặng về tính toán. Không giống như MR, kỹ thuật NOUS sử dụng điều kiện trực giao trong công thức 4.19 để giảm kích thước của tập ứng cử viên sau mỗi vòng lặp. Điều này giúp giảm tải về tính toán rất nhiều so với thuật toán MR. Hơn nữa, thuật toán NOUS ưu việt hơn thuật toán PF về phương diện tổng dung lượng bởi vì PF lựa chọn các thuê bao hiện có tốc độ thấp để phục vụ nhằm mục tiêu công bằng giữa các thuê bao.

4.4 Kết quả mô phỏng

Để đo lường hiệu quả của kỹ thuật NOUS, một số tình huống mô phỏng đã được nghiên cứu dựa trên hệ thống mmWave bằng phần mềm Matlab để so sánh các kỹ thuật lập lịch sau:

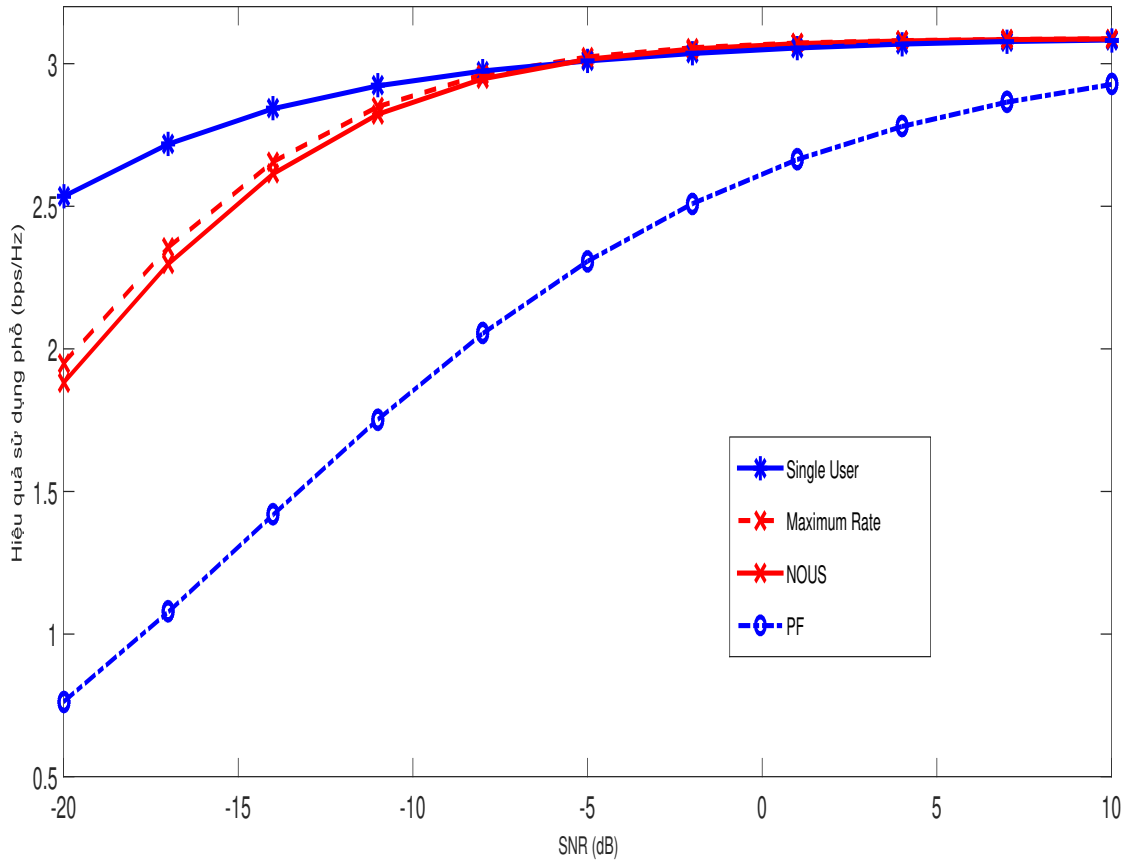
- NOUS.

- MR.
- PF.
- Một thuê bao.

Ở đây phần ước lượng kênh truyền được coi như là lý tưởng, mô phỏng không tính toán đến ước lượng kênh truyền ở chiều lên. Kịch bản mô phỏng là xem xét hiệu quả sử dụng phổ với tỷ lệ tín hiệu trên nhiễu SINR thấp phù hợp với thực tế khi sử dụng các bộ ADC phân giải thấp, so sánh giữa tình huống sử dụng bộ ADC phân giải thấp và bộ ADC hoàn hảo (No-ADC) để đánh giá hiệu quả kỹ thuật đề xuất cho các trường hợp khác nhau của độ phân giải ADC và so sánh hiệu quả sử dụng phổ giữa các kỹ thuật lập lịch khi số lượng thuê bao lập lịch tăng dần.

Trong hình 4.1 luận án so sánh hiệu quả sử dụng phổ theo tỷ lệ tín hiệu trên nhiễu SINR. Luận án xem xét với $M_{BS} = 256$ ăng ten ở trạm gốc, $M_{MS} = 16$ ăng ten tại thuê bao, $M_S = 20$ thuê bao sẽ được lựa chọn để phục vụ, $K_a = 160$ thuê bao trong tế bào, số bit ADC là $b = 2$ (bit) và SINR tăng từ -20 đến 10 dB. Kết quả nghiên cứu chỉ ra rằng cả hai kỹ thuật NOUS và MR có thể tiệm cận hiệu quả sử dụng phổ của trường hợp một thuê bao khi SINR có giá trị rất lớn. Hiệu quả sử dụng phổ của kỹ thuật NOUS gần như xấp xỉ với kết quả của kỹ thuật MR với thời gian tìm kiếm ngắn hơn. Kỹ thuật toán PF tập trung vào độ công bằng giữa các thuê bao nên không thể có hiệu quả sử dụng phổ cao. Khoảng cách giữa các kỹ thuật giảm xuống khi SINR tăng lên vì sai số của quá trình lượng tử hóa sẽ giảm đi khi SINR ở miền giá trị cao. Kỹ thuật được đề xuất NOUS có thể hoạt động tốt ngay cả miền SINR nhỏ vì sai số lượng tử tăng lên làm giảm độ trực giao của kênh truyền.

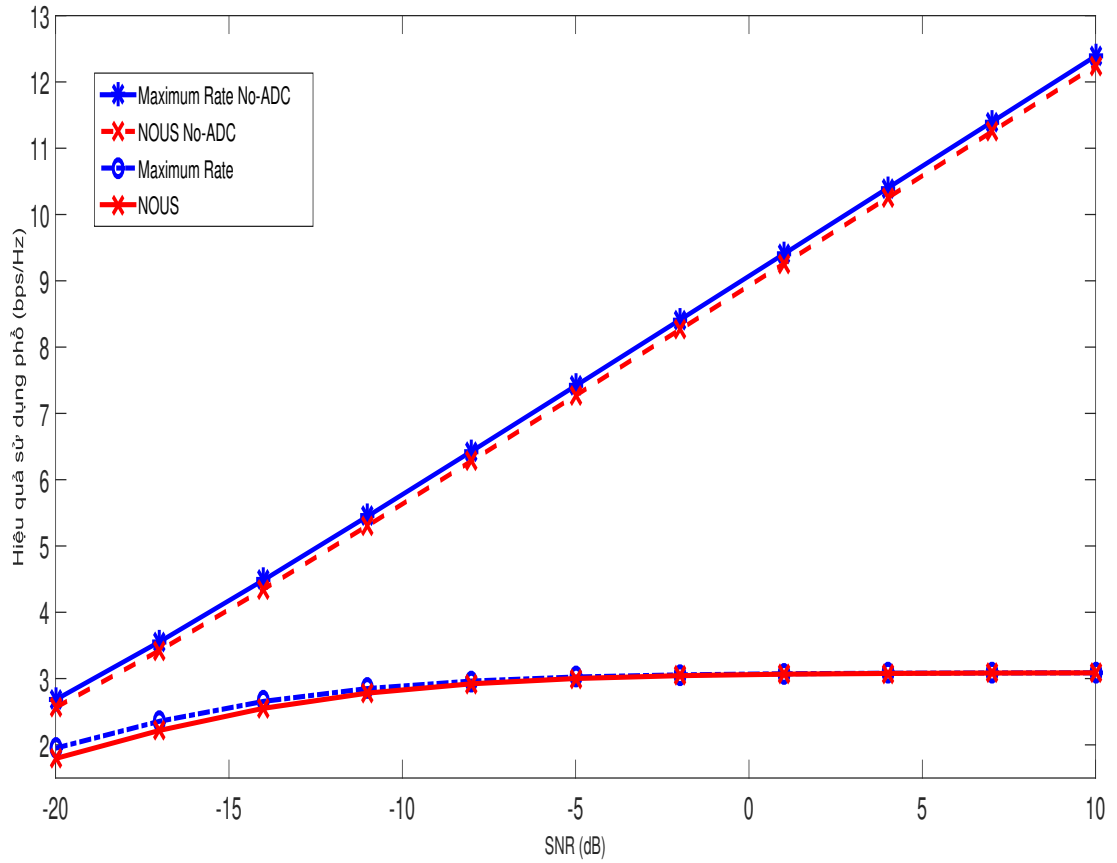
Ở hình 4.2 luận án so sánh hiệu quả sử dụng phổ trong hai trường hợp: đầu tiên là hệ thống sử dụng bộ ADC phân giải thấp với số bit $b = 2$ (bit) và trường hợp



Hình 4.1: Hiệu quả sử dụng phổ theo SNR

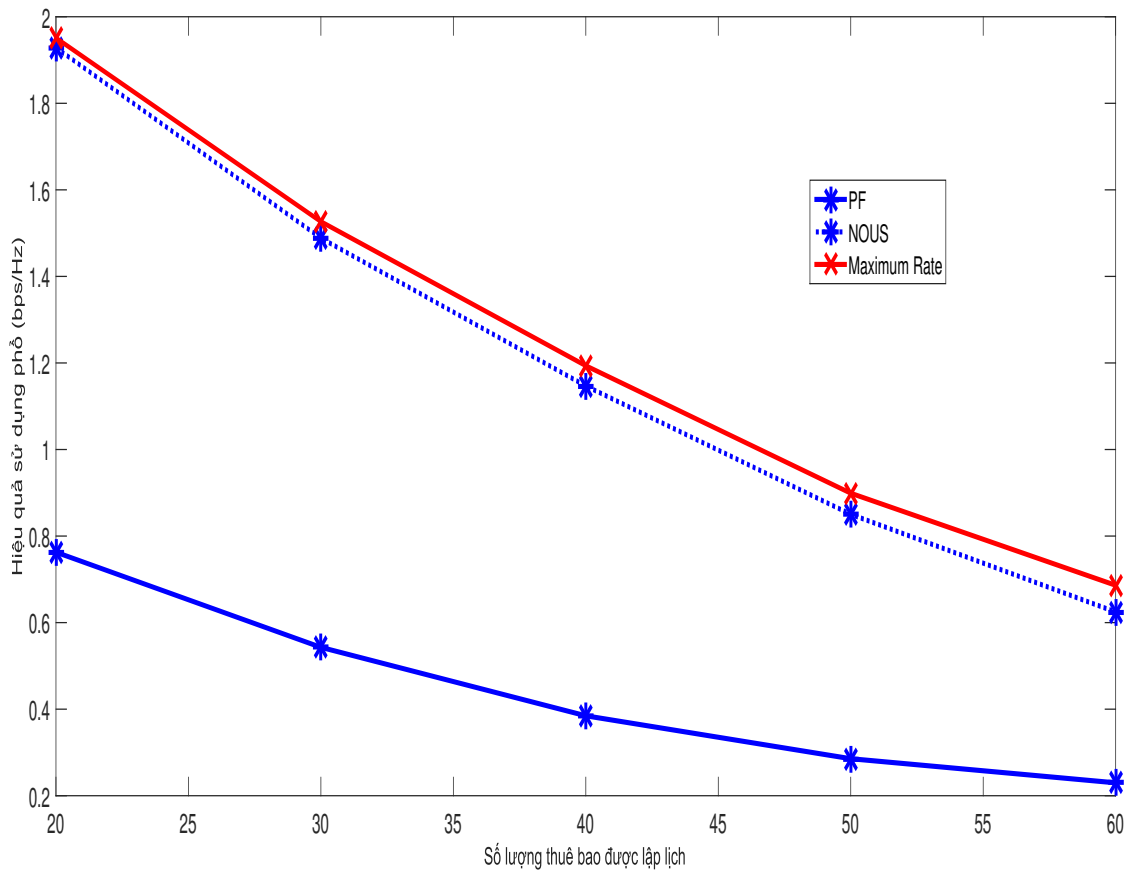
thứ hai coi như bộ ADC là hoàn hảo, không có sai số lượng tử hóa gọi là No-ADC. Có thể thấy rằng ở trường hợp No-ADC thì hiệu quả sử dụng có thể tăng tới vô cùng khi tăng SINR. Trong trường hợp bộ ADC phân giải thấp $b = 2$, sự gia tăng của SINR ở miền giá trị cao không làm tăng lên nhiều hiệu quả sử dụng phổ do sai số của quá trình lượng tử hóa. Trong cả hai trường hợp, kỹ thuật NOUS có thể xấp xỉ hiệu năng của kỹ thuật MR về phương tiện hiệu quả sử dụng phổ. Kết quả cũng cho thấy kỹ thuật NOUS có thể hoạt động tốt trong cả hai trường hợp bộ ADC có độ phân giải thấp và cao. Khoảng cách giữa kỹ thuật MR và NOUS giảm xuống khi SINR tăng lên vì NOUS hoạt động tốt khi bộ ADC có độ phân giải thấp.

Ở hình 4.3, luận án xem xét trường hợp số lượng thuê bao được phục vụ K_S



Hình 4.2: So sánh hiệu quả sử dụng phổ giữa bộ ADC phân giải thấp và No-ADC

tăng từ 20 lên 60 thuê bao, $\mathbb{K}_a = 160$ và $\rho = 3$ dB . Nó chỉ ra rằng nếu trạm gốc BS phục vụ nhiều thuê bao hơn thì hiệu quả sử dụng phổ của mỗi thuê bao sẽ giảm xuống vì công suất nhận được của mỗi thuê bao bị giảm xuống. Hơn nữa, càng khó tìm thấy nhiều thuê bao trực giao vì chúng ta đã giữ tổng số thuê bao \mathbb{K}_a không đổi trong khi lại tăng số thuê được phục vụ vì thế hiệu quả sử dụng phổ của NOUS sẽ giảm xuống một chút so với MR.



Hình 4.3: Hiệu quả sử dụng phổ theo K_s

4.5 Kết luận chương 4

Trong chương này, luận án đã đề xuất kỹ thuật lập lịch dựa trên mức độ trực giao của kênh truyền giữa các thuê bao. Từ đó có thể giảm thiểu thời gian xác định tập thuê bao tối ưu trong khi vẫn hướng đến mục tiêu tối đa hóa dung lượng hệ thống. Tính khả thi của kỹ thuật lập lịch đề xuất được minh chứng bằng hiệu quả sử dụng phổ của kỹ thuật NOUS tiệm cận với kỹ thuật MR và đánh bại được kỹ thuật PF trong cả hai trường hợp xét ảnh hưởng của bộ ADC phân giải thấp và trường hợp coi như bộ ADC là hoàn hảo (No-ADC). Và thực nghiệm cũng cho thấy nếu như số thuê bao phục vụ trong hệ thống tăng lên thì khả năng tìm thấy tập thuê bao bán trực giao sẽ giảm xuống làm giảm hiệu quả sử dụng phổ so với

kỹ thuật MR.

KẾT LUẬN

Luận án đã tiến hành nghiên cứu những kiến thức nền tảng của lớp vật lý của hệ thống thông tin di động 5G như kỹ thuật ước lượng kênh, tiền mã hóa và các kỹ thuật lập lịch hiện có. Từ đó xác định những yếu tố ảnh hưởng mới và then chốt của mạng 5G so với các kỹ thuật lập lịch hiện có. Từ đó, luận án đã đề xuất các kỹ thuật lập lịch nhằm đảm bảo chất lượng cho các thuê bao sử dụng thông tin lão hóa kênh truyền, kỹ thuật lập lịch cho phép thuê bao có thể có nhiều tốc độ tối thiểu khác nhau trong cùng một chu kỳ khung và kỹ thuật lập lịch dựa trên xác định tập thuê bao trực giao. Tóm tắt những đóng góp của luận án cũng như đề xuất một số hướng nghiên cứu tiếp theo trong tương lai được tác giả trình bày dưới đây.

Kết quả của luận án

- **Đề xuất kỹ thuật lập lịch đảm bảo chất lượng kênh truyền cho thuê bao sử dụng thông tin lão hóa kênh truyền nhằm đối phó sự biến thiên của kênh truyền trong thực tế**

Luận án đã tính toán tốc độ mà thuê bao có thể đạt được khi sử dụng thông tin lão hóa kênh truyền cũng như thuật toán kiểm tra khi thêm vào các thuê bao được phục vụ nhằm đảm bảo rằng việc phục vụ các thuê bao sử dụng thông tin lão hóa kênh truyền không làm ảnh hưởng tới chất lượng của các

thuê bao khác trong hệ thống.

- **Kỹ thuật lập lịch đảm bảo đa tốc độ tối thiểu cho thuê bao**

Luận án đã tính toán công suất tối ưu cho từng thuê bao cho phù hợp với yêu cầu về tốc độ tối thiểu của thuê bao. Kỹ thuật cũng đã sử dụng thông tin về độ ưu tiên của lưu lượng khi phục vụ thuê bao để đảm bảo rằng lưu lượng có độ ưu tiên cao hơn sẽ được phục vụ trước các lưu lượng có độ ưu tiên thấp hơn. Đây là điểm mới so với các nghiên cứu trước đây tập trung về độ trễ của lưu lượng.

- **Kỹ thuật lập lịch cho hệ thống mmWave sử dụng bộ ADC phân giải thấp**

Luận án đã nghiên cứu hệ thống mmWave sử dụng bộ ADC phân giải thấp và tính toán tốc độ của từng thuê bao và cả hệ thống có thể đạt được. Từ đó luận án đã xây dựng điều kiện bán trực giao về kênh truyền giữa các thuê bao của hệ thống mmWave sử dụng bộ ADC phân giải thấp. Điều này giúp giảm thiểu thời gian tìm ra tập thuê bao phục vụ tối ưu so với các kỹ thuật truyền thống trước đây như MR hay PF.

Hướng nghiên cứu tiếp theo

Bên cạnh các kỹ thuật lập lịch đảm bảo tốc độ cho thuê bao theo yêu cầu của chất lượng dịch vụ sử dụng cho mạng Massive MIMO thì tác giả sẽ nghiên cứu, cải tiến để áp dụng cho các hệ thống mmWave Massive MIMO với nhiều sự khác biệt trong mô hình hệ thống và đặc tính về kênh truyền. Cân nhắc các yếu tố về đặc tính kênh truyền để tối ưu hơn nữa dung lượng của hệ thống cũng là hướng nghiên cứu tiếp theo của luận án.

CÁC CÔNG TRÌNH KHOA HỌC ĐÃ CÔNG BỐ

BÀI BÁO KHOA HỌC

[J1] Hung Pham, Bac Dang Hoai, Ban Nguyen Tien, "Scheduling for Massive MIMO using channel aging under QoS constraints", Tạp chí Khoa học và Công nghệ Vietnam số 57 (5) (2019) pp 617-630.

[J2] Hung Pham, Bac Dang Hoai, Ban Nguyen Tien, "Scheduling for Massive MIMO under Power and QoS constraints", Tạp chí Khoa học Công nghệ Thông tin và Truyền thông (1) (2020) pp 71-77

[J3] Hung Pham, Bac Dang Hoai, Vu Khanh Quy, and Ban Nguyen Tien, "Semiorthogonal User Scheduling for Millimeter Wave using Low-resolution ADC", Tạp chí viễn thông JoCM quyển 16, số 1, Tháng 1, 2021

Tài liệu tham khảo

- [1] 3GPP, *Study on New Radio Access Technology, TR 38.804, V14.0.0*, 3GPP, Mar. (2017).
- [2] 3GPP, *Technical Specification Group Radio Access Network TTS 38.300, V15.2.0*, 3GPP, June (2018).
- [3] Afroz and et al, “Comparative analysis of downlink packet scheduling algorithms in 3GPP LTE networks”, *International Journal of Wireless & Mobile Networks*, 7.5 (2015).
- [4] J. Akhtar and K. Rajawat, “Quality-of-Service Constrained User and Antenna Selection in Downlink Massive-MIMO Systems”. *2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, (2019), pp. 1–6.
- [5] M. Alkhaled, E. Alsusa and W. Pramudito, “Adaptive user grouping algorithm for the downlink massive MIMO systems”, *2016 IEEE Wireless Communications and Networking Conference*, (2016), pp. 1–6.
- [6] A. Alkhateeb, O. El Ayach and et al, “Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems”, *IEEE Journal of Selected Topics in Signal Processing*, 8.5 (2014), pp. 831–846.
- [7] A. Alkhateeb, G. Leus and R. W. Heath, “Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems”, *IEEE Transactions on Wireless Communications*, 14.11 (2015), pp. 6481–6494.
- [8] Ameigeiras and et al, “3GPP QoS-based scheduling framework for LTE”, *EURASIP Journal on Wireless Communications and Networking*, 2016.1 (2016), p. 78.
- [9] J. G. Andrews, S. Buzzi and et al, “What Will 5G Be?”, *IEEE Journal on Selected Areas in Communications*, 32.6 (2014), pp. 1065–1082.
- [10] M. Andrews and et al, “Providing quality of service over a shared wireless link”, *IEEE Communications Magazine*, 39.2 (2001), pp. 150–154.
- [11] O. E. Ayach, S. Rajagopal and et al, “Spatially Sparse Precoding in Millimeter Wave MIMO Systems”. *IEEE Transactions on Wireless Communications*, 13.3 (2014), pp. 1499–1513.
- [12] B. Hassibi and B. M. Hochwald, “How much training is needed in multiple-antenna wireless links?”, *IEEE Transactions on Information Theory*, 49.4 (2003), pp. 951–963.
- [13] M. Benmimoune, E. Driouch and et al, “Joint transmit antenna selection and user scheduling for Massive MIMO systems”, *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, (2015), pp. 381–386.
- [14] Bjornson, “The Massive MIMO Paradigm Fundamentals and State of the Art”, *GLOBECOM* ((2016)), p. 82.
- [15] E. Björnson, E. G. Larsson and M. Debbah, “Massive MIMO for Maximal Spectral Efficiency: How Many Users and Pilots Should Be Allocated?”, *IEEE Transactions on Wireless Communications*, 15.2 (2016), pp. 1293–1308.

- [16] F. Boccardi, R. W. Heath and et al, “Five disruptive technology directions for 5G”, *IEEE Communications Magazine*, 52.2 (2014), pp. 74–80.
- [17] M. Bohge et al, “Dynamic resource allocation in OFDM systems: an overview of cross-layer optimization principles and techniques”, *IEEE Network*, 21.1 (2007), pp. 53–59.
- [18] J. Brady, N. Behdad and A. M. Sayeed, “Beamspace MIMO for Millimeter-Wave Communications: System Architecture, Modeling, Analysis, and Measurements”, *IEEE Transactions on Antennas and Propagation*, 61.7 (2013), pp. 3814–3827.
- [19] S. A. Busari and et al, “Millimeter-Wave Massive MIMO Communication for Future Wireless Systems: A Survey”, *IEEE Communications Surveys Tutorials*, 20.2 (2018), pp. 836–869.
- [20] Z. Chen, E. Bjornson and E. G. Larsson, “Dynamic Scheduling and Power Control in Uplink Massive MIMO with Random Data Arrivals”, *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, (2019), pp. 1–6.
- [21] J. Choi, B. L. Evans and A. Gatherer, “ADC bit allocation under a power constraint for mmWave massive MIMO communication receivers”, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2017), pp. 3494–3498.
- [22] J. Choi, B. L. Evans and A. Gatherer, “Resolution-Adaptive Hybrid MIMO Architectures for Millimeter Wave Communications”, *IEEE Transactions on Signal Processing*, 65.23 (2017), pp. 6201–6216.
- [23] J. Choi and B. L. Evans, “User Scheduling for Millimeter Wave MIMO Communications with Low-Resolution ADCs”, *2018 IEEE International Conference on Communications (ICC)*, (2018), pp. 1–6.
- [24] 2019 Cisco Mobile VNI, *The Cisco® Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update*, Cisco, (2019).
- [25] M. Costa, “Writing on dirty paper (Corresp.)”, *IEEE Transactions on Information Theory*, 29.3 (1983), pp. 439–441.
- [26] L. Fan, S. Jin and et al, “Uplink Achievable Rate for Massive MIMO Systems With Low-Resolution ADC”, *IEEE Communications Letters*, 19.12 (2015), pp. 2186–2189.
- [27] Jerry R. Hampton, *Introduction to MIMO Communications*, Cambridge University Press, (2015).
- [28] R. W. Heath, N. González-Prelcic and et al, “An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems”, *IEEE Journal of Selected Topics in Signal Processing*, 10.3 (2016), pp. 436–453.
- [29] N. Q. Hien, “Massive MIMO: Fundamentals and System Designs”, PhD thesis. Linköping, (2015).
- [30] J. Hoydis, S. Brink and M. Debbah, “Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?”, *IEEE Journal on Selected Areas in Communications*, 31.2 (2013), pp. 160–171.
- [31] ITU-R, *Recommendation ITU-R M.2083-0*, ITU, (2015).
- [32] W. C. Jakes, *Microwave mobile communications*, New York: Wiley, (1974).
- [33] C. Kong, C. Zhong and et al, “Effect of channel aging on the sum rate of uplink massive MIMO systems”, *2015 IEEE International Symposium on Information Theory (ISIT)*, (2015), pp. 1222–1226.
- [34] C. Kong, C. Zhong and Z. Zhang, “Performance of ZF precoder in downlink massive MIMO with non-uniform user distribution”, *Journal of Communications and Networks*, 18.5 (2016), pp. 688–698.

- [35] Chuili Kong and et al, “Sum-Rate and Power Scaling of Massive MIMO Systems with Channel Aging”, *IEEE Transactions on Communications*, 63 (Oct. 2015).
- [36] E. G. Larsson, O. Edfors and et al, “Massive MIMO for next generation wireless systems”, *IEEE Communications Magazine*, 52.2 (2014), pp. 186–195.
- [37] B. Lee, L. Ngo and B. Shim, “Antenna group selection based user scheduling for massive MIMO systems”, *2014 IEEE Global Communications Conference*, (2014), pp. 3302–3307.
- [38] G. Lee, Y. Sung and J.Seo, “Randomly-Directional Beamforming in Millimeter-Wave Multiuser MISO Downlink”, *IEEE Transactions on Wireless Communications*, 15.2 (2016), pp. 1086–1100.
- [39] G. Lee, Y. Sung and M. Kountouris, “On the Performance of Random Beamforming in Sparse Millimeter Wave Channels”, *IEEE Journal of Selected Topics in Signal Processing* 10.3 (2016), pp. 560–575.
- [40] H. Lee, S. Park and S. Bahk, “Enhancing Spectral Efficiency Using Aged CSI in Massive MIMO Systems”, *2016 IEEE Global Communications Conference (GLOBECOM)*, (2016), pp. 1–6.
- [41] H. Lee, S. Park and S. Bahk, “Enhancing Spectral Efficiency Using Aged CSI in Massive MIMO Systems”, *2016 IEEE Global Communications Conference (GLOBECOM)*, (2016), pp. 1–6.
- [42] T. L. Marzetta, “How Much Training is Required for Multiuser MIMO?”, *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, (2006), pp. 359–363.
- [43] T. L. Marzetta, “Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas”, *IEEE Transactions on Wireless Communications*, 9.11 (2010), pp. 3590–3600.
- [44] J. Mo, P. Schniter and et al, “Channel estimation in millimeter wave MIMO systems with one-bit quantization”, *2014 48th Asilomar Conference on Signals, Systems and Computers*, (2014), pp. 957–961.
- [45] J. Mo, P. Schniter and R. W. Heath, “Channel Estimation in Broadband Millimeter Wave MIMO Systems With Few-Bit ADCs”, *IEEE Transactions on Signal Processing*, 66.5 (2018), pp. 1141–1154.
- [46] H. Q. Ngo, E. G. Larsson and T. L. Marzetta, “Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots”, *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, (2013), pp. 293–298.
- [47] H. Q. Ngo, E. G. Larsson and T. L. Marzetta, “Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems”, *IEEE Transactions on Communications*, 61.4 (2013), pp. 1436–1449.
- [48] O. Orhan, E. Erkip and S. Rangan, “Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance”, *2015 Information Theory and Applications Workshop (ITA)*, (2015), pp. 191–198.
- [49] A. K. Papazafeiropoulos and T. Ratnarajah, “Deterministic Equivalent Performance Analysis of Time-Varying Massive MIMO Systems”, *IEEE Transactions on Wireless Communications*, 14.10 (2015), pp. 5795–5809.
- [50] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems”, *IEEE Communications Magazine*, 49.6 (2011), pp. 101–107.
- [51] T. S. Rappaport and et al, “Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!”, *IEEE Access*, 1 (2013), pp. 335–349.
- [52] T. S. Rappaport, F. Gutierrez and et al, “Broadband Millimeter-Wave Propagation Measurements and Models Using Adaptive-Beam Antennas for Outdoor Urban Cellular Communications”, *IEEE Transactions on Antennas and Propagation*, 61.4 (2013), pp. 1850–1859.

- [53] F. Rusek, D. Persson and et al, “Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays”, *IEEE Signal Processing Magazine*, 30.1 (2013), pp. 40–60.
- [54] C. Rusu, N. González-Prelcic and R. W. Heath, “Low resolution adaptive compressed sensing for mmWave MIMO receivers”, *2015 49th Asilomar Conference on Signals, Systems and Computers*, (2015), pp. 1138–1143.
- [55] L. Dai, S. Mumtaz, J. Rodriguez, *mmWave Massive MIMO A Paradigm for 5G*, Elsevie, (2017).
- [56] A. M. Sayeed, “Deconstructing multiantenna fading channels”, *IEEE Transactions on Signal Processing* 50.10 (2002), pp. 2563–2579.
- [57] M. Shariat, A. U. Quddus and et al, “Scheduling as an important cross-layer operation for emerging broadband wireless systems”, *IEEE Communications Surveys Tutorials*, 11.2 (2009), pp. 74–86.
- [58] J. Sohn, S. W. Yoon and J. Moon, “On Reusing Pilots Among Interfering Cells in Massive MIMO”, *IEEE Transactions on Wireless Communications*, 16.12 (2017), pp. 8092–8104.
- [59] S. Stefania, *LTE The UMTS Long Term Evolution: From Theory to Practice*, 2nd Edition. Wiley, (2009), p. 290.
- [60] K. T. Truong and R. W. Heath, “Effects of channel aging in massive MIMO systems”, *Journal of Communications and Networks*, 15.4 (2013), pp. 338–351.
- [61] H. Viswanathan, S. Venkatesan and H. Huang, “Downlink capacity evaluation of cellular networks with known-interference cancellation”, *IEEE Journal on Selected Areas in Communications*, 21.5 (2003), pp. 802–811.
- [62] P. Xia, S. Yong and et al, “A practical SDMA protocol for 60 GHz millimeter wave communications”, *2008 42nd Asilomar Conference on Signals, Systems and Computers*, (2008), pp. 2019–2023.
- [63] Y. Xu, G. Yue and et al, “User grouping and scheduling for large scale MIMO systems with two-stage precoding”, *2014 IEEE International Conference on Communications (ICC)*, (2014), pp. 5197–5202.
- [64] H. Yang and T. L. Marzetta, “Performance of Conjugate and Zero-Forcing Beamforming in Large-Scale Antenna Systems”, *IEEE Journal on Selected Areas in Communications*, 31.2 (2013), pp. 172–179.
- [65] T. Yoo and A. Goldsmith, “On the optimality of multi antenna broadcast scheduling using zero-forcing beamforming”, *IEEE Journal on Selected Areas in Communications*, 24.3 (2006), pp. 528–541.
- [66] L. You, X. Gao and et al, “Pilot Reuse for Massive MIMO Transmission over Spatially Correlated Rayleigh Fading Channels”, *IEEE Transactions on Wireless Communications*, 14.6 (2015), pp. 3352–3366.