

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC



PHẠM MINH THẮNG

TỐI ƯU ĐỘ TRỄ TRONG MẠNG 5G

LUẬN VĂN THẠC SĨ KỸ THUẬT ĐIỆN TỬ

HÀ NỘI, 2024

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

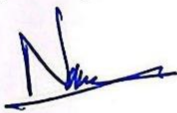
PHẠM MINH THẮNG

TỐI ƯU ĐỘ TRỄ TRONG MẠNG 5G

Chuyên ngành : Kỹ thuật điện tử
Mã số : 8520203

LUẬN VĂN THẠC SĨ KỸ THUẬT ĐIỆN TỬ

Người hướng dẫn khoa học:


PGS.TS HÀ HẢI NAM

TS. NGUYỄN ĐỨC THỦY


Nguyễn Đức Thủy

HÀ NỘI, 2024

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Số: 187/QĐ-ĐHDL

Hà Nội, ngày 25 tháng 10 năm 2024

QUYẾT ĐỊNH
Về việc thành lập Hội đồng chấm luận văn thạc sĩ khoá 10 (2022 - 2024)
ngành Kỹ thuật điện tử, mã số 8520203

HIỆU TRƯỞNG TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

Căn cứ Nghị quyết số 03/NQ-HĐT ngày 30/5/2023 của Hội đồng trường Trường Đại học Điện lực ban hành Quy chế Tổ chức và hoạt động của Trường Đại học Điện lực;

Căn cứ Quyết định số 666/QĐ-ĐHDL ngày 31/5/2022 của Trường Đại học Điện lực về việc ban hành Quy chế đào tạo trình độ thạc sĩ;

Căn cứ Quyết định số 978/QĐ-ĐHDL ngày 28/6/2023 của Trường Đại học Điện lực về việc bổ sung một số điều của Quy chế đào tạo trình độ thạc sĩ;

Căn cứ Quyết định số 729/QĐ-ĐHDL ngày 15/5/2023 của Trường Đại học Điện lực về việc phân công người hướng dẫn và tên đề tài luận văn thạc sĩ đối với các học viên cao học khóa 10 (2022-2024) ngành Kỹ thuật điện tử;

Căn cứ Hồ sơ đăng ký bảo vệ luận văn thạc sĩ của học viên và Đề nghị danh sách Hội đồng chấm luận văn thạc sĩ ngày 25/10/2024 của Khoa Điện tử Viễn thông;

Căn cứ Tờ trình số 199/TTr-QL Đào tạo ngày 25/10/2024 của Phòng Quản lý Đào tạo về việc đề nghị thành lập Hội đồng chấm luận văn thạc sĩ khóa 2022 ngành Kỹ thuật điện tử;

Theo đề nghị của Trưởng phòng Quản lý Đào tạo.

QUYẾT ĐỊNH:

Điều 1. Thành lập Hội đồng đánh giá luận văn thạc sĩ khóa 10 (2022-2024) cho học viên Cao học:

Họ và tên học viên: Phạm Minh Thắng

Giới tính: Nam

Sinh ngày: 11/11/1987

Nơi sinh: Hà Nội

Lớp: CH10.KTĐT2

Mã HV : 22CH5030001

Ngành đào tạo: Kỹ thuật điện tử

Mã số: 8520203

Tên đề tài luận văn: Tối ưu độ trễ trong mạng 5G.

Người hướng dẫn khoa học:

-NHD1: PGS.TS Hà Hải Nam - Trường Đại học Điện lực

-NHD2: TS. Nguyễn Đức Thủy - Trường Đại học Điện lực

Điều 2. Cử Ông (bà) có tên sau đây tham gia vào Hội đồng đánh giá luận văn thạc sĩ.

STT	Họ và tên (học hàm, học vị)	Đơn vị công tác	Chức danh trong HĐ
1	TS. Phạm Duy Phong	Trường Đại học Điện lực	Chủ tịch
2	TS. Đỗ Văn Tuấn	Trường Đại học Điện lực	Thư kí
3	PGS.TS. Trần Hoài Trung	Trường Đại học Giao thông Vận tải	Phản biện 1
4	TS. Hoàng Thị Phương Thảo	Trường Đại học Điện lực	Phản biện 2
5	TS. Trần Văn Nghĩa	Học viện Phòng không Không quân	Ủy viên

Điều 3. Hội đồng đánh giá luận văn thạc sĩ có trách nhiệm tổ chức đánh giá trong thời hạn 30 ngày kể từ ngày kí Quyết định. Hội đồng tự giải thể sau khi hoàn thành nhiệm vụ.

Điều 4. Trưởng các đơn vị: Phòng Quản lý Đào tạo, Phòng Kế hoạch - Tài chính, Khoa Điện tử Viễn thông, các đơn vị khác có liên quan, các thành viên trong Hội đồng chấm luận văn thạc sĩ tại Điều 2 và các cá nhân có tên tại Điều 1 chịu trách nhiệm thi hành Quyết định này. /.

Nơi nhận:

- Như Điều 4;
- HT (để báo cáo);
- Lưu: VT, QLĐào tạo, PhươngBT (05).

KT. HIỆU TRƯỞNG
PHÓ HIỆU TRƯỞNG
TRƯỜNG
ĐẠI HỌC
ĐIỆN LỰC

Nguyễn Lê Cường

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

Hà Nội, ngày tháng năm

BẢN XÁC NHẬN ĐÃ CHỈNH SỬA LUẬN VĂN

Họ và tên Học viên: Phạm Minh Thắng

Là học viên cao học ngành: Kỹ thuật điện tử

Lớp: CH10-KTĐT...

Khoá: 10 (2022-2024)

Luận văn thạc sĩ đã trình trước Hội đồng ngày: 02/11/2024

Tên đề tài: Tối ưu độ trễ trong mạng 5G

Người hướng dẫn: PGS.TS Hà Hải Nam - Trường Đại học Điện lực

TS. Nguyễn Đức Thủy - Trường Đại học Điện lực

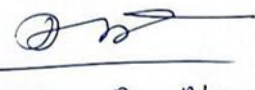
Hội đồng xác nhận nội dung đã hiệu chỉnh gồm:

Căn chỉnh đồng bộ phong chữ.

Rà soát công thức trích dẫn.

Làm rõ kết quả tính toán.

Hà Nội, ngày.....tháng năm

HỌC VIÊN	THƯ KÍ HỘI ĐỒNG	GIẢNG VIÊN HƯỚNG DẪN	CHỦ TỊCH HỘI ĐỒNG
(Ký và ghi rõ họ tên)	(Ký và ghi rõ họ tên)	(Ký và ghi rõ họ tên)	(Ký và ghi rõ họ tên)
 Phạm Minh Thắng	 Đỗ Văn Tuấn	 Nguyễn Đức Thủy	 Hà Hải Nam
			 Phạm Duy Phong

LỜI CẢM ƠN

Trước hết, xin cho phép Em được bày tỏ lòng biết ơn sâu sắc tới PGS.TS. Hà Hải Nam và TS. Nguyễn Đức Thủy. Hai thầy đã không ngừng đồng hành và hỗ trợ Em trong suốt hành trình nghiên cứu khoa học. Sự tận tâm và sáng suốt trong việc hướng dẫn, cùng những lời khuyên quý báu, đã giúp Em vững vàng hơn trong kiến thức và xác định rõ ràng hướng đi cho luận văn của mình.

Tiếp theo, lòng biết ơn của Em cũng dành tới tập thể giảng viên của Trường Đại học Điện lực, đặc biệt là quý thầy cô giáo của lớp Cao học Kỹ thuật điện tử khóa 10. Quý thầy cô không chỉ là những người truyền đạt tri thức mà còn là nguồn cảm hứng, tạo điều kiện cho tôi hoàn thiện luận văn này. Xin kính chúc quý thầy cô luôn mạnh khỏe, thành công và tiếp tục gặt hái được nhiều thành tựu hơn nữa trong sự nghiệp của mình.

Cuối cùng, không thể không nhắc đến sự ủng hộ, động viên không ngừng từ bạn bè, đồng nghiệp và gia đình đã giúp Em vượt qua mọi khó khăn và thách thức.

Xin trân trọng cảm ơn!

LỜI CAM ĐOAN

Tôi xin cam đoan:

Bản luận văn tốt nghiệp này là công trình nghiên cứu của cá nhân tôi, được thực hiện dựa trên cơ sở nghiên cứu lý thuyết, thực tế dưới sự hướng dẫn của PGS.TS Hà Hải Nam và TS. Nguyễn Đức Thủy.

Các số liệu, kết luận của luận văn là trung thực, dựa trên sự nghiên cứu những mô hình, kết quả đã đạt được của các nước trên thế giới và trải nghiệm của bản thân, chưa từng được công bố dưới bất kỳ hình thức nào trước khi trình bày bảo vệ trước “Hội đồng đánh giá luận văn thạc sỹ kỹ thuật”.

Hà Nội, ngày ... tháng ... năm 2024

Tác giả



Phạm Minh Thắng

MỤC LỤC

LỜI CẢM ƠN	1
LỜI CAM ĐOAN	Error! Bookmark not defined.
DANH MỤC TỪ VIẾT TẮT.....	3
DANH MỤC HÌNH	5
DANH MỤC BẢNG.....	6
MỞ ĐẦU	7
CHƯƠNG 1: CÁC TIẾN BỘ TRONG PHÁT TRIỂN CÔNG	
NGHỆ 5G	9
1.1 Giới thiệu	9
1.2 Sự phát triển của mạng 5G.....	9
1.2.1 LTE phiên bản 11 đến 14.....	11
1.2.2 5G NR Pha 1: Phiên bản 15	13
1.2.3 5G NR Pha 2: Phiên bản 16, 17	14
1.3 Các công nghệ chủ chốt trong mạng 5G.....	16
1.3.1 Massive MIMO.....	16
1.3.2 Truyền sóng dạng búp (Beamforming).....	17
1.3.3 Sóng cực ngắn mmWave	19
1.3.4 Phân tách mạng theo lát cắt	20
1.3.5 Công nghệ MEC	22
1.4 Kết luận chương 1	23
CHƯƠNG 2: ĐỘ TRỄ TRONG MẠNG 5G	25
2.1 Giới thiệu	25
2.2 Các tham số chất lượng dịch vụ (QoS).....	26
2.2.1 Băng thông (Bandwidth).....	27
2.2.2 Thông lượng (Throughput)	27
2.2.3 Jitter.....	27
2.2.4 Tỷ lệ lỗi bit (Bit Error Rate - BER)	27
2.2.5 Tỷ lệ chuyển giao gói tin (Packet Delivery Ratio - PDR).....	27
2.2.6 Thời gian khám phá tuyến (Route Discovery Time)	28
2.2.7 Độ trễ (Delay)	28

2.3	Yêu cầu về hiệu năng kỹ thuật cho 5G	30
2.4	Nguồn trễ trong mạng di động	30
2.5	Tầng vật lý URLLC trong 5G NR	33
2.6	Yêu cầu về các chỉ số đánh giá URLLC	34
2.6.1	Độ trễ mặt phẳng người dùng (User Plane)	34
2.6.2	Độ trễ mặt phẳng điều khiển (Control plane)	35
2.6.3	Độ tin cậy	36
2.6.4	Tính di động và thời gian gián đoạn di động	37
2.6.5	Tốc độ dữ liệu đỉnh	38
2.7	Các ràng buộc và phương pháp để đạt được độ trễ thấp	39
2.7.1	Giải pháp RAN	40
2.7.2	Giải pháp mạng lõi	44
2.7.3	Giải pháp Caching	46
2.8	Kết luận chương 2	48
CHƯƠNG 3: ĐÁNH GIÁ ĐỘ TRỄ DỊCH VỤ V2X TRONG		
	MẠNG 5G	50
3.1	Tổng quan về V2X	50
3.2	Phân tích các yếu tố ảnh hưởng đến độ trễ	51
3.3	Tính toán độ trễ tại các phân đoạn	52
3.3.1	Độ trễ phạm vi vô tuyến	52
3.3.2	Độ trễ trong mạng truyền dẫn	55
3.3.3	Độ trễ trong mạng lõi	59
3.4	Thực nghiệm tính toán độ trễ trong các phạm vi phân đoạn mạng 5G	61
3.4.1	Đánh giá ảnh hưởng của yếu tố SCS	61
3.4.2	Đánh giá ảnh hưởng của yếu tố Tỷ lệ dung lượng liên kết (α)	62
3.4.3	Đánh giá ảnh hưởng của yếu tố mật độ xe	64
3.5	Các thách thức trong triển khai 5G tại Việt Nam	66
3.6	Kết luận chương 3	68
KẾT LUẬN CỦA LUẬN VĂN		70
DANH MỤC TÀI LIỆU THAM KHẢO		72

DANH MỤC TỪ VIẾT TẮT

Ký Hiệu	Tiếng Anh	Tiếng Việt
BER	Bit Error Rate	Tỷ lệ lỗi bit
CA	Carrier aggregation	Cộng gộp sóng mang
D2D	Device-to-Device	Kết nối thiết bị tới thiết bị
eICIC	Enhanced Inter-cell interference coordination	Kỹ thuật điều khiển giảm can nhiễu tăng cường giữa các cell
eMBB	Enhanced Mobile Broadband	Băng thông rộng di động nâng cao
HARQ	Hybrid automatic repeat request	Yêu cầu lặp lại tự động hỗn hợp
IoT	Internet of Things	Internet vạn vật
MEC	Multi-access Edge Computing	Điện toán biên đa truy cập
MIMO	Multiple Input Multiple Output	Nhiều đầu vào nhiều đầu ra
MMC	Massive Machine Communications	Truyền thông với thiết bị cực lớn
mMTC	massive Machine Type Communication	Truyền thông máy số lượng lớn
NFV	Network Functions Virtualization	Ảo hóa các chức năng mạng
NR	New Radio	Mạng vô tuyến mới
OFDM	Orthogonal Frequency Division Multiplexing	Ghép kênh phân chia theo tần số trực giao
PDR	Packet Delivery Ratio	Tỷ lệ chuyển giao gói tin
RAN	Radio Access Network	Mạng truy cập vô tuyến
RNC	Radio Network Controller	Bộ điều khiển thông tin di động
RRC	Resource Radio Control	Điều khiển tài nguyên vô tuyến
RTT	Round Trip Time	Thời gian trọn vòng
SDN	Software-Defined Networking	Mạng được xác định bằng phần mềm

URLLC	Ultra-Reliable Low-Latency	Truyền thông siêu tin cậy và độ trễ thấp
V2X	Vehicle-to-Everything	Giao tiếp xe với mọi thứ
VoLTE	Voice over Long-Term Evolution	Thoại trên nền LTE

DANH MỤC HÌNH

Hình 1-1: Kiến trúc mạng LTE-Advanced	11
Hình 1-2: Các mốc thời gian của LTE.....	11
Hình 1-3: Các mốc thời gian của 5G	14
Hình 1-4: Lộ trình phát triển dự kiến của 5G	15
Hình 1-5: So sánh vùng phủ giữa Anten truyền thống với Massive MIMO	17
Hình 1-6: Azimuth Beamforming và Elevation Beamforming	18
Hình 1-7: Dải tần trải rộng của 5G mmWave	19
Hình 1-8: Chia tách một mạng vật lý thành các mạng ảo	20
Hình 1-9: Các ứng dụng của Network Slicing	21
Hình 1-10: Mô hình của MEC	22
Hình 2-1: Mô hình cung ứng Qos (QoS Provisioning)	26
Hình 2-2: Độ trễ của truyền tin trong mạng di động.....	31
Hình 2-3: Đánh giá độ trễ mặt phẳng người dùng	34
Hình 2-4: Quy trình RACH.....	36
Hình 2-5: Cloud RAN (C-RAN)	44
Hình 2-6: Các loại caching.....	46
Hình 2-7: Các giải pháp khác nhau để đạt được độ trễ thấp trong 5G	49
Hình 3-1: V2N và V2N2V	50
Hình 3-2: Các thành phần độ trễ từ phương tiện đến mạng lõi.....	51
Hình 3-3. Mặt phẳng người dùng	53
Hình 3-4. Độ trễ mặt phẳng người dùng.....	54
Hình 3-5: Kết quả mô phỏng đánh giá $tg_{NB,tx}$	62
Hình 3-6: Kết quả mô phỏng đánh giá α	63
Hình 3-7: Mô hình giả định cho thực nghiệm tính toán trễ trong phạm vi M1.....	64
Hình 3-8: Độ trễ xử lý tại phạm vi mạng M1 với sự thay đổi tốc độ tối λ của gói tin.....	66
Hình 3-9: Tỷ lệ thiết bị 5G tại 3 nhà mạng	67

DANH MỤC BẢNG

Bảng 2-1: Yêu cầu hiệu suất kỹ thuật URLLC	30
Bảng 2-2: So sánh ưu nhược điểm của các mã hóa kênh	41
Bảng 2-3: Tổng quan các kỹ thuật trong mạng lõi cho độ trễ thấp	45
Bảng 3-1: Giá trị thông số thiết lập giả định cho SCS	61
Bảng 3-2: Thông số thiết lập giả định cho việc tính toán tốc độ đường UL của nút mạng backhaul	63
Bảng 3-3: Thông số giả định thiết lập tính toán cho mạng tại phạm vi M1 (λ bằng nhau cho cả UL và DL)	65

MỞ ĐẦU

1. Lý do chọn đề tài

Trong thời đại ngày nay, sự phát triển nhanh chóng của công nghệ viễn thông đã tạo nên một cuộc cách mạng trong ngành truyền thông di động, đặt ra những thách thức đầy thú vị và khó khăn trong việc cải thiện hiệu suất của mạng di động. Trong bối cảnh này, mạng 5G đã nổi lên như một bước tiến lớn với hứa hẹn mang lại sự đổi mới và cải thiện đáng kể trong nhiều khía cạnh, đặc biệt là về độ trễ.

Luận văn thạc sĩ này tập trung nghiên cứu và đánh giá cách mạng của mạng 5G, đặc biệt là về khả năng cải thiện độ trễ trong hệ thống. Với sự ra đời của các kỹ thuật mới như numerologies biến đổi và cấu trúc khung linh hoạt, mạng 5G hứa hẹn mang lại trải nghiệm truyền thông di động với độ trễ thấp và đáng tin cậy hơn bao giờ hết.

Luận văn này sẽ đi sâu vào những khía cạnh kỹ thuật, cấu trúc mạng và các yếu tố khác nhau ảnh hưởng đến độ trễ trong mạng 5G. Bằng cách này, sẽ có cơ hội hiểu rõ hơn về những tiến bộ quan trọng và thách thức mà mạng 5G đang đối mặt khi nỗ lực cải thiện độ trễ, đặc biệt trong bối cảnh các ứng dụng như xe tự lái, y tế từ xa và Internet vạn vật (IoT) đang trở nên ngày càng quan trọng.

Khái niệm về độ trễ được chia thành nhiều phần, mọi tối ưu hóa trong từng phần này đều quan trọng. Sự tiến hóa của mạng đã làm cho việc này trở nên dễ dàng hơn thông qua MEC (Điện toán biên) và NFV (Ảo hóa chức năng mạng), cả trong hạ tầng cốt lõi và trong các số liệu số khác nhau trên băng tần mmwave mới. Trong luận văn này, em sẽ nghiên cứu về độ trễ của hệ thống 5G và tập trung vào V2N2V (Vehicle-to-Network-to-Vehicle), đặc biệt là các phân đoạn để định lượng độ trễ, góp phần hỗ trợ các dịch vụ V2X có thể đạt được độ trễ thấp.

Hy vọng rằng, thông qua nghiên cứu này, em có thể đóng góp một phần nhỏ vào sự hiểu biết và phát triển của mạng 5G, đồng thời mở ra những cơ hội mới để tối ưu hóa độ trễ và đưa ra những giải pháp sáng tạo nhằm đáp ứng nhu cầu ngày càng cao về dịch vụ 5G trong xã hội hiện đại.

2. Mục đích nghiên cứu

Nghiên cứu tìm hiểu, đánh giá hiệu quả các phân đoạn tạo ra độ trễ và đưa ra các khuyến nghị trong việc tối ưu độ trễ trong mạng 5G.

3. Nhiệm vụ nghiên cứu

Thu thập, phân tích thông tin, hiểu và nắm được các kiến thức về mạng 5G. Tổng hợp, phân tích, nắm được các giải pháp kỹ thuật để giảm độ trễ.

Thực hiện các tính toán thực nghiệm để kiểm nghiệm các phương pháp theo lý thuyết. Đưa ra được các nhận xét, kết luận cho các nội dung nghiên cứu.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: Nghiên cứu về mạng 5G, độ trễ trong mạng 5G, các cấu trúc, yếu tố ảnh hưởng đến độ trễ.

Phạm vi nghiên cứu: Đánh giá các yếu tố ảnh hưởng đến độ trễ trong mô hình End-to-End và công nghệ V2X.

Thời gian nghiên cứu: Từ năm 2023 - 2024.

5. Phương pháp nghiên cứu

Các nội dung nghiên cứu được thực hiện bằng việc kết hợp giữa nghiên cứu lý thuyết và thực nghiệm mô phỏng.

- *Nghiên cứu lý thuyết:*

Thu thập, tổng hợp, phân loại các thông tin, kết quả nghiên cứu phù hợp với đối tượng, phạm vi, nội dung nghiên cứu đề ra theo đề cương.

Phân tích, đánh giá, phương án lựa chọn, rút ra các nhận xét, kết luận đối với nội dung thông tin, kết quả nghiên cứu.

- *Thực nghiệm mô phỏng:*

Thực hiện các mô phỏng để kiểm nghiệm, đánh giá các phương pháp, kỹ thuật theo lý thuyết để từ đó rút ra các nhận xét, kết luận cho nội dung nghiên cứu được thực hiện.

CHƯƠNG 1: CÁC TIẾN BỘ TRONG PHÁT TRIỂN CÔNG NGHỆ 5G

1.1 Giới thiệu

Hành trình của mạng thế hệ thứ năm (5G) đã chứng kiến nhiều cột mốc về sự phát triển, cải tiến về công nghệ và tiêu chuẩn hóa, vì mỗi phiên bản đã đưa ra giải pháp để cung cấp dịch vụ tốt hơn và đáp ứng các yêu cầu cao hơn. Chúng ta hiện đang sống trong một thế giới mà các cuộc họp, các khóa học trực tuyến và việc hỗ trợ y tế từ xa được xem như một phần trong cuộc sống hàng ngày. Lượng thông tin mà chúng ta truy cập nhiều hơn bao giờ hết, và tất cả điều này có thể thực hiện được nhờ vào sự tiến bộ và sáng tạo trong lĩnh vực kết nối không dây.

Mạng 5G đánh dấu một bước tiến đột phá trong lĩnh vực viễn thông, mang lại nhiều cơ hội và thách thức cho sự phát triển của công nghệ truyền thông hiện đại. Con đường tiến tới mạng 5G không chỉ là một hành trình kỹ thuật, mà còn là một hành trình mang tính chiến lược, định hình lại cách chúng ta kết nối, giao tiếp và sử dụng dịch vụ trực tuyến. Với khả năng truyền tải dữ liệu với tốc độ siêu nhanh, độ trễ thấp và khả năng kết nối hàng tỷ thiết bị cùng một lúc, mạng 5G hứa hẹn mở ra một thế giới mới với ứng dụng và dịch vụ đa dạng từ truyền hình ảnh chất lượng cao, trải nghiệm thực tế ảo, đến các ứng dụng trong giao thông, y tế, giáo dục, và công nghiệp.

Tuy nhiên, việc tiến tới mạng 5G cũng đặt ra nhiều thách thức, từ khía cạnh kỹ thuật như thiết kế mạng, đa dạng hóa tần số, đến các vấn đề như an ninh thông tin, quản lý tài nguyên mạng và ảnh hưởng đến môi trường. Trong bối cảnh này, việc hiểu rõ về quá trình triển khai, ứng dụng và ảnh hưởng của mạng 5G trở nên cực kỳ quan trọng để có thể tận dụng hết tiềm năng của công nghệ này và đồng thời đối mặt với những thách thức đang chờ đón.

1.2 Sự phát triển của mạng 5G

Sự phát triển của LTE bắt đầu trong giai đoạn của phiên bản HSPA+ (High Speed Packet Access plus) của 3GPP vào tháng 12 năm 2004[1]. Phiên bản đầu tiên của LTE được gọi là phiên bản 8 của 3GPP, được giới thiệu vào tháng 12 năm 2008. Phiên bản 9 của 3GPP hoàn thành vào tháng 12 năm 2009 là phiên

bản LTE thứ hai. Sau đó, LTE-Advanced xuất hiện trên thị trường vào tháng 6 năm 2011 bởi 3GPP dưới tên gọi phiên bản 10. Các phiên bản từ phiên bản 10 đến 14 được gọi là LTE-Advanced. 5G NR (New Radio) xuất hiện trong Release 15 của 3GPP, được công bố vào tháng 12 năm 2017. Sau đó, Release 16 của 3GPP tiếp tục phát triển và hoàn thiện chuẩn 5G NR, được công bố vào tháng 7 năm 2019. Đây là hai phiên bản quan trọng đánh dấu sự ra đời và tiến triển của công nghệ 5G NR. Hiện tại phiên bản mới nhất là 5G NR Advanced, hay còn gọi là Release 17 của 3GPP, được công bố vào tháng 6 năm 2021. Đây là một bước tiến quan trọng trong việc phát triển và cải tiến của công nghệ 5G NR. Sự tiến hóa từ LTE-Advanced cho đến 5G NR Advanced được tóm tắt như sau:

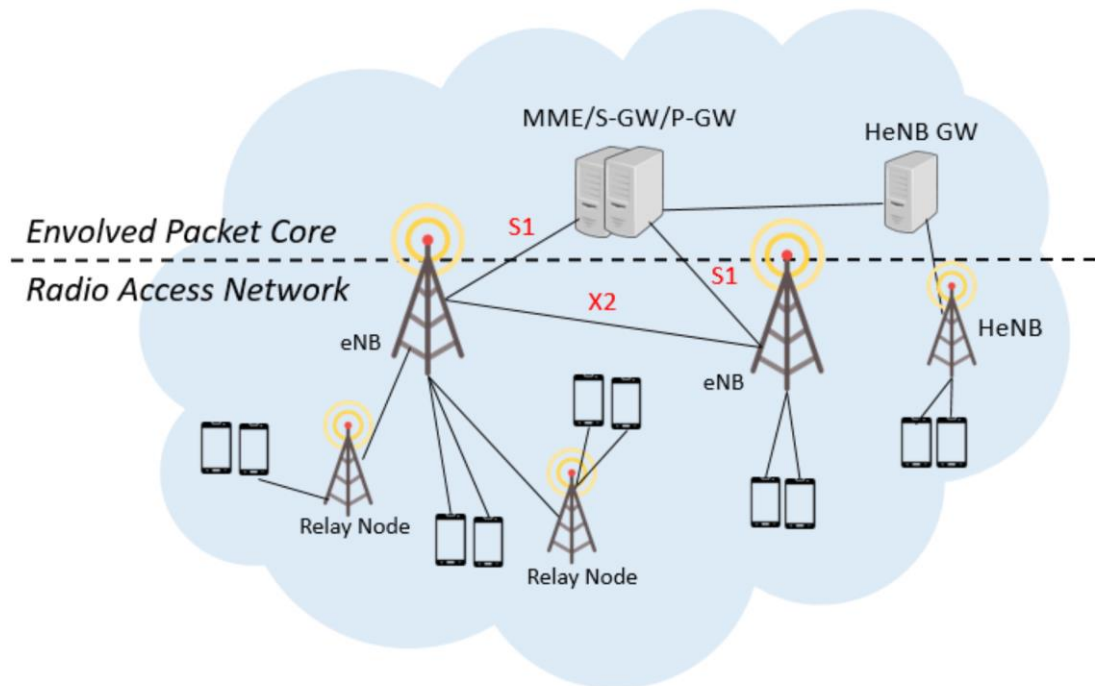
LTE phiên bản 10

LTE phiên bản 10 (LTE Advanced) đã được hoàn thiện vào cuối năm 2010, trong đó đã được thêm các cải tiến so với phiên bản 8/9 về hiệu suất và năng lực. LTE phiên 10 và các phiên bản sau được gọi là LTE-Advanced, đáp ứng tất cả các yêu cầu của IMT-Advanced. Vì vậy, LTE-Advanced là hệ thống truyền thông di động 4G bao gồm một số tính năng bổ sung so với các phiên bản trước. Các tính năng này bao gồm:

- Cộng gộp sóng mang (CA - Carrier aggregation)
- Các kỹ thuật MIMO tiên tiến
- Chuyển mạng không dây (Wireless relaying)
- Kỹ thuật điều khiển giảm can nhiễu tăng cường giữa các cell (eICIC - Enhanced Inter-cell interference coordination)
- Kỹ thuật phối hợp đa điểm (CoMP - Coordinated multipoint).

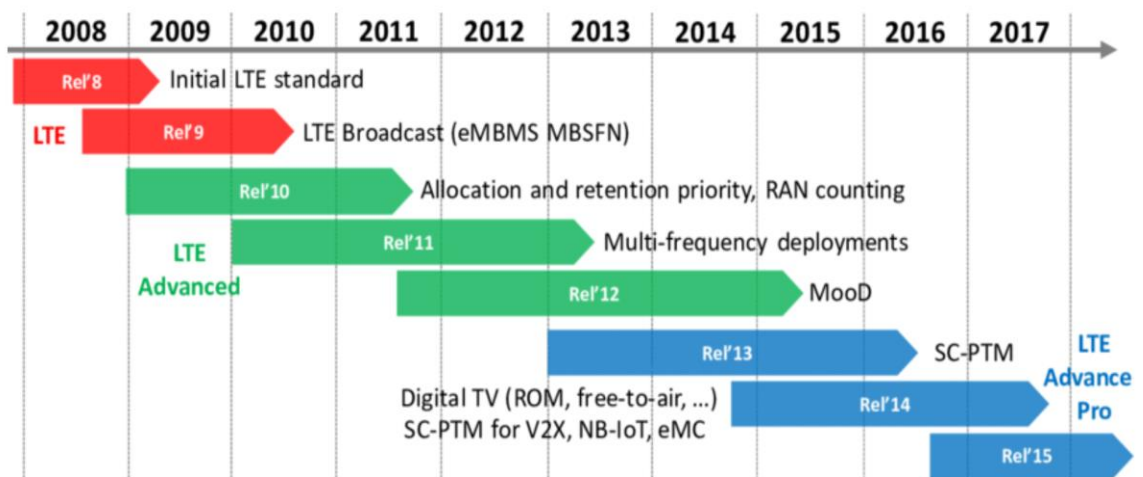
Trong tổng hợp tần số, nhiều thành phần tần số được tổng hợp để hỗ trợ băng thông truyền tải cao. LTE-Advanced có thể hỗ trợ tổng hợp tần số lên đến 100 MHz. Chuyển mạng không dây cung cấp sự mở rộng vùng phủ sóng và cải thiện hiệu suất biên của các ô.

Kiến trúc mạng của LTE-Advanced được minh họa trong hình 1-1. Tất cả các eNBs (Evolved NodeB) được kết nối với S-GW (Serving Gateway) và MME của EPC. eNB là sự kết hợp của UMTS (3G) NB và RNC (Radio Network Control)[2].



Hình 1-1: Kiến trúc mạng LTE-Advanced

1.2.1 LTE phiên bản 11 đến 14



Hình 1-2: Các mốc thời gian của LTE

Các bản cập nhật từ LTE phiên bản 11 đến phiên bản 13 đã mang lại nhiều cải tiến và tính năng mới cho chuẩn LTE, củng cố nền tảng cho sự phát triển của mạng 5G. Dưới đây là một số điểm nổi bật trong các bản release này so với LTE phiên bản 10.

LTE phiên bản 11:

- Tối ưu hóa CoMP: Cải thiện hiệu suất ở vùng biên giữa các cell, giảm tác động của nhiễu giữa các cell và tăng tốc độ truyền dữ liệu.
- Carrier Aggregation Enhancements: Tiếp tục mở rộng khả năng của Carrier Aggregation, cho phép kết hợp nhiều carrier hơn, cải thiện băng thông và tốc độ.
- Enhanced ICIC: Cải tiến quản lý nhiễu và sự phối hợp giữa các cell để tăng cường hiệu quả sử dụng tài nguyên tần số.

LTE phiên bản 12:

- D2D (Device-to-Device) Communication: Hỗ trợ giao tiếp trực tiếp giữa các thiết bị, giảm độ trễ và cải thiện hiệu quả sử dụng băng tần.
- Small Cell Enhancements: Tăng cường hiệu suất và quản lý của các small cell, bao gồm cải tiến backhaul và quản lý năng lượng.
- MIMO Enhancements: Phát triển công nghệ MIMO, bao gồm hỗ trợ cho mảng anten lớn hơn và cải tiến thuật toán, để tăng cường tốc độ và hiệu quả mạng.
- Introduction of eMTC (enhanced Machine Type Communication) và NB-IoT (Narrowband IoT): Hỗ trợ IoT với chi phí thấp, tiêu thụ năng lượng thấp và kết nối rộng rãi.

LTE phiên bản 13:

- Further enhancements to Carrier Aggregation and MIMO: Tiếp tục cải tiến các công nghệ này để tăng cường băng thông và hiệu năng thực thi.
- LAA (Licensed Assisted Access): Cho phép LTE sử dụng băng tần không giấy phép (như 5GHz) cùng với băng tần được cấp phép, giúp tăng cường băng thông và dung lượng mạng.
- Enhanced VoLTE (Voice over LTE): Cải thiện chất lượng cuộc gọi và hiệu quả mạng cho dịch vụ thoại qua LTE.
- Giới thiệu CIoT (Cellular IoT): Tiếp tục phát triển hỗ trợ cho IoT với các giải pháp như eMTC và NB-IoT, tăng cường kết nối và hiệu suất cho các thiết bị IoT.

Các cải tiến này đánh dấu bước tiến quan trọng trong việc mở rộng khả

năng và hiệu suất của mạng LTE, đặt nền móng cho sự chuyển đổi sang mạng 5G[3].

LTE phiên bản 14:

- V2X (Vehicle to Everything) Communications: Hỗ trợ giao tiếp giữa xe cộ và mọi thứ, cải thiện an toàn và hiệu quả giao thông.
- Enhanced LAA (eLAA): Cải tiến từ LAA, hỗ trợ cả uplink và downlink trong băng tần không giấy phép, cung cấp băng thông lớn hơn.
- FeMBMS (Further enhanced Multimedia Broadcast Multicast Service): Cải thiện dịch vụ phát sóng và multicast, hỗ trợ các ứng dụng như phát sóng trực tiếp và dịch vụ khẩn cấp.
- NB-IoT and eMTC Enhancements: Tiếp tục cải thiện hỗ trợ cho IoT với các cải tiến về hiệu suất, độ tin cậy và hiệu quả năng lượng[4].

1.2.2 5G NR Pha 1: Phiên bản 15

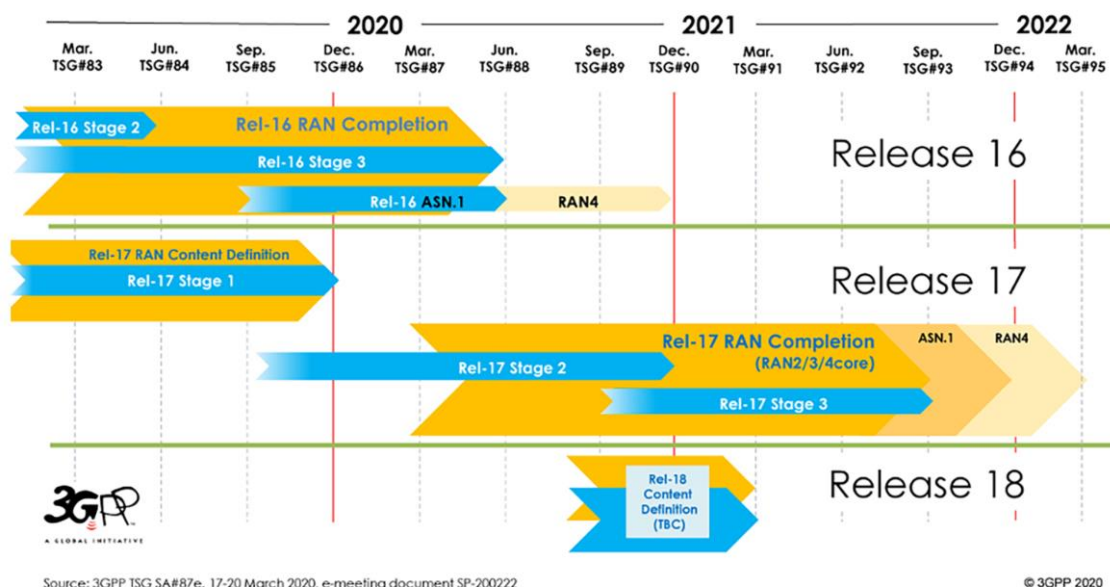
Xác định một tiêu chuẩn hoàn toàn mới cho 5G là một nhiệm vụ lớn. 3GPP đã chia tiêu chuẩn 5G thành hai phiên bản: Phiên bản 15, tương ứng với Pha 1 NR (New Radio), và Phiên bản 16, tương ứng với Pha 2 NR. Trong Pha 1 NR, có những yếu tố chung giữa LTE và NR, chẳng hạn như cả hai đều sử dụng kỹ thuật chia tần số OFDM (Orthogonal Frequency Division Multiplexing). Để thực sự triển khai phiên bản đầy đủ của NR, cần triển khai một lượng lớn phần cứng mới. Để tiếp tục sử dụng phần cứng hiện có, việc triển khai được thực hiện theo phương pháp giai đoạn. Một phiên bản không độc lập (NSA – Non-Standalone) sẽ sử dụng lõi LTE và một phiên bản độc lập (SA - Standalone) sẽ sử dụng lõi NR và hoàn toàn độc lập với mạng lõi LTE.

Để các thiết bị vẫn có thể liên lạc với nhau, các thuật ngữ mới đã được giới thiệu[5]:

- LTE eNB: Thiết bị có thể kết nối với EPC hoặc mạng lõi LTE hiện tại
- eLTE eNB: Bản tiến hóa của LTE eNB có thể kết nối với EPC và lõi NextGen
- gNB: Tương đương 5G NR của LTE eNB
- NG: Giao diện giữa lõi NextGen và gNB

- NG2: Giao diện lớp điều khiển giữa mạng lõi và RAN (S1-C trong LTE)
- NG3: Giao diện lớp người dùng giữa mạng lõi và RAN (S1-U trong LTE)

1.2.3 5G NR Pha 2: Phiên bản 16, 17



Hình 1-3: Các mốc thời gian của 5G

5G NR Phase 2, bao gồm Release 16 và Release 17 của 3GPP, mang lại nhiều cải tiến và tính năng mới, mở rộng khả năng của mạng 5G và hỗ trợ cho một loạt các ứng dụng mới. Dưới đây là tổng hợp các cải tiến chính của mỗi bản release:

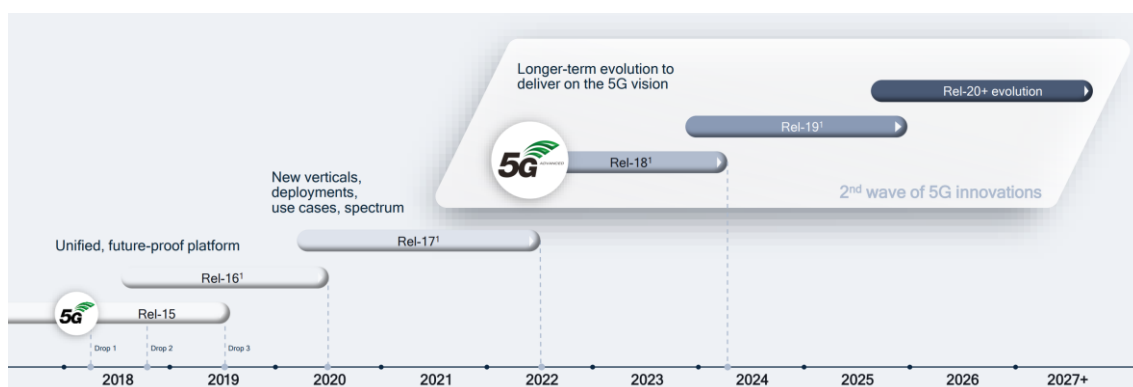
Phiên bản 16 - "5G Phase 2":

Phiên bản 16, hoàn thành vào giữa năm 2020, tập trung vào việc mở rộng 5G vào các lĩnh vực công nghiệp và các ứng dụng chuyên biệt [6].

- Ultra-Reliable Low-Latency Communication (URLLC): Cải thiện độ tin cậy và giảm độ trễ cho các ứng dụng nhạy cảm với thời gian, như tự động hóa công nghiệp và lái xe tự động.
- Vehicle-to-Everything (V2X): Hỗ trợ giao tiếp giữa xe và mọi thứ, bao gồm xe với người đi bộ, các phương tiện khác, và hạ tầng giao thông, để cải thiện an toàn và hiệu quả giao thông.
- 5G Broadcast: Cung cấp dịch vụ phát sóng trực tiếp qua mạng 5G, mở

ra khả năng phát sóng truyền hình di động mà không cần mạng Wi-Fi hoặc dữ liệu di động.

- Enhanced Mobile Broadband (eMBB): Mở rộng khả năng của eMBB, tăng cường dung lượng và băng thông cho video độ phân giải cao, VR/AR, và các dịch vụ dữ liệu nặng khác.
- Non-Public Networks (NPNs): Hỗ trợ mạng riêng 5G, cho phép các tổ chức triển khai mạng 5G riêng biệt cho các ứng dụng công nghiệp và doanh nghiệp.
- Integrated Access and Backhaul (IAB): Cải thiện khả năng triển khai mạng bằng cách sử dụng các liên kết không dây cho cả truy cập và backhaul.



Hình 1-4: Lộ trình phát triển dự kiến của 5G

Phiên bản 17:

Phiên bản 17, tiếp tục mở rộng phạm vi và khả năng của 5G NR với các cải tiến và tính năng mới, đã được cho ra mắt vào tháng 3 năm 2022.

- Cải thiện hỗ trợ cho IoT: Mở rộng khả năng kết nối cho các thiết bị IoT với các cải tiến về NR Light (RedCap), hỗ trợ các thiết bị với yêu cầu công suất thấp và băng thông hạn chế.
- Enhanced V2X: Tiếp tục cải thiện hỗ trợ cho các ứng dụng V2X, nhằm đáp ứng nhu cầu ngày càng cao về an toàn và hiệu quả giao thông.
- Multicast and Broadcast Services: Mở rộng khả năng phát sóng và multicast qua mạng 5G, hỗ trợ các ứng dụng như phát sóng trực tiếp và dịch vụ thông tin khẩn cấp.
- Cải thiện tính linh hoạt của mạng: Giới thiệu các công nghệ và cơ chế

mới để tăng cường tính linh hoạt và hiệu quả của mạng, như cải tiến trong quản lý quang phổ và kỹ thuật truy cập mạng.

Cả hai bản release này đều nhấn mạnh vào việc mở rộng 5G vào các lĩnh vực mới và cung cấp các giải pháp kỹ thuật tiên tiến để đáp ứng các yêu cầu về hiệu suất, độ tin cậy, và độ trễ cho một loạt các ứng dụng mới [7].

1.3 Các công nghệ chủ chốt trong mạng 5G

Mạng 5G sử dụng một loạt các công nghệ tiên tiến để đạt được tốc độ cao, độ trễ thấp và khả năng kết nối đáng tin cậy cho một lượng lớn thiết bị. Dưới đây là một số công nghệ chủ chốt trong mạng 5G:

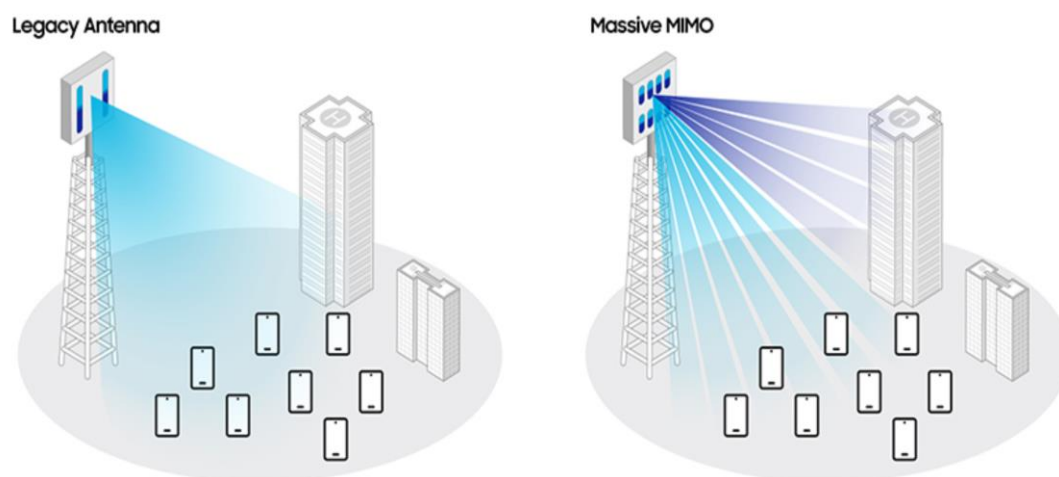
1.3.1 Massive MIMO

Massive MIMO (Multiple Input Multiple Output) sử dụng hàng trăm hoặc thậm chí hàng ngàn anten tại mỗi trạm gốc để tăng cường dung lượng và hiệu quả của tín hiệu. Công nghệ này cho phép một lượng lớn người dùng và thiết bị được kết nối đồng thời mà không làm giảm chất lượng dịch vụ.

MIMO truyền thống đã được nghiên cứu một cách rộng rãi trong các hệ thống 3G và 4G. Công nghệ này có thể giúp dung lượng tăng theo cách tuyến tính với quy mô của mảng anten nếu số lượng anten tăng lớn, nâng cao dung lượng và hiệu suất mạng di động. Các hệ thống truyền thông di động thế hệ tiếp theo sẽ được hưởng lợi từ Massive MIMO như sau [8]:

- Độ trễ: Độ trễ của truyền dữ liệu không dây bị ảnh hưởng lớn bởi hiện tượng Fading, vì việc truyền lại hoặc mã hóa kiểm soát lỗi tỷ lệ thấp sẽ được sử dụng để chống lại fading. Bằng cách sử dụng Massive MIMO, tín hiệu thu sẽ được hưởng lợi từ việc phân tập không gian và xử lý tín hiệu MIMO, chẳng hạn như truyền sóng dạng chùm (beamforming) và tiền mã hóa, có thể bù đắp cho hiện tượng fading.
- Chi phí và công suất: Mặc dù số lượng anten của Massive MIMO lên đến hàng trăm, nhưng Massive MIMO có thể đạt được lợi ích cao hơn với công suất phát thấp hơn cho mỗi anten. Trên thực tế, tổng công suất của Massive MIMO thậm chí còn ít hơn rất nhiều so với MIMO truyền thống, điều này có nghĩa là bộ khuếch đại công suất giá rẻ và tiết kiệm điện có công suất phát hàng mili-Watt sẽ thay thế bộ khuếch đại đắt tiền

hơn với công suất hàng chục Watt, nhưng vẫn còn một quãng đường dài trước khi nó được sử dụng rộng rãi.

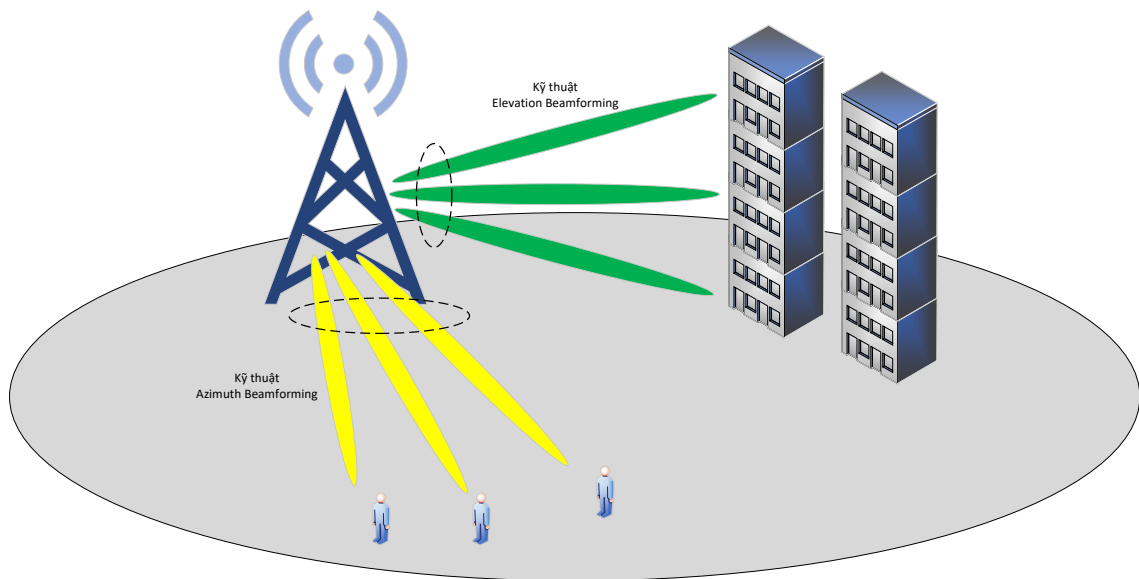


Hình 1-5: So sánh vùng phủ giữa Anten truyền thông với Massive MIMO
(nguồn *samsung.com*)

Ví dụ, trong một khu vực đô thị đông đúc với nhiều tòa nhà cao tầng, việc triển khai Massive MIMO kết hợp với beamforming sẽ cho phép tín hiệu vô tuyến được "điều hình" một cách chính xác đến các khu vực cần thiết, tối ưu hóa độ phủ và chất lượng kết nối. Điều này giúp tăng thông lượng mạng và cải thiện trải nghiệm người dùng, đặc biệt là ở những khu vực khó khăn về mặt tín hiệu như bên trong các tòa nhà hoặc ở các vùng ngoại ô.

1.3.2 Truyền sóng dạng búp (Beamforming)

Beamforming là một kỹ thuật tiên tiến trong viễn thông và mạng không dây, được sử dụng để tối ưu hóa truyền và nhận tín hiệu. Nó điều khiển hướng và độ lớn của sóng vô tuyến được phát ra từ một anten mảng để tạo ra một "chùm" (beam) tín hiệu hướng đến vị trí cụ thể của thiết bị đầu cuối, đồng thời giảm thiểu sự can nhiễu tới các thiết bị khác. Cách thức hoạt động của Beamforming là sử dụng nhiều anten để phát ra tín hiệu vô tuyến. Bằng cách điều chỉnh pha và cường độ của tín hiệu phát ra từ mỗi anten, beamforming tạo ra một mô hình sóng vô tuyến mà có thể được "định hình" và "hướng" một cách linh động để tối ưu hóa độ phủ và hiệu suất mạng tại vị trí của người nhận. Hình 1-6 minh họa hai kỹ thuật của Beamforming.



Hình 1-6: Azimuth Beamforming và Elevation Beamforming

Azimuth beamforming và Elevation beamforming là hai kỹ thuật quan trọng trong beamforming, mỗi kỹ thuật tập trung vào việc điều chỉnh và hướng tín hiệu vô tuyến theo một hướng cụ thể để tối ưu hóa độ phủ sóng và giảm nhiễu. Cả hai đều là một phần trong nỗ lực nhằm tinh chỉnh và điều khiển tín hiệu trong mạng không dây, đặc biệt trong mạng 5G và các hệ thống truyền dẫn tiên tiến, chi tiết như sau:

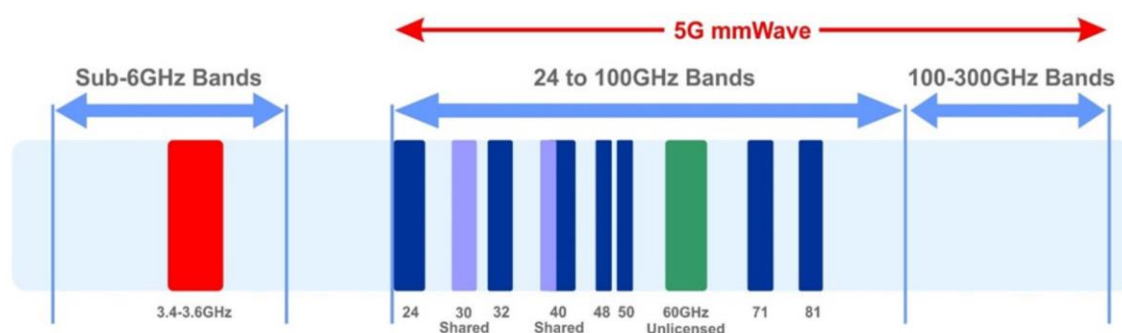
- Azimuth Beamforming: Điều khiển và hướng tín hiệu theo chiều ngang (hoặc azimuth plane). Điều này có nghĩa là việc tạo ra và hướng "chùm" tín hiệu vô tuyến dọc theo mặt đất, từ trái qua phải hoặc ngược lại, để tối ưu hóa độ phủ và tập trung tín hiệu vào khu vực cụ thể nơi người dùng hoặc thiết bị đang được phục vụ. Kỹ thuật này đặc biệt hữu ích trong các khu vực rộng lớn hoặc dân cư đông đúc, nơi cần điều chỉnh độ phủ sóng theo chiều ngang để cung cấp kết nối tốt nhất cho người dùng.
- Elevation Beamforming: Tập trung vào việc điều chỉnh và hướng tín hiệu theo chiều dọc (hoặc elevation plane). Kỹ thuật này điều khiển "chùm" tín hiệu đi lên hoặc xuống, tối ưu hóa độ phủ sóng trong các tòa nhà cao tầng hoặc các khu vực địa hình phức tạp. Kỹ thuật này rất có giá trị trong các đô thị với nhiều tòa nhà cao tầng, nơi tín hiệu cần được hướng lên các tầng cao hoặc điều chỉnh để tránh các chướng ngại vật và tối ưu hóa độ phủ sóng theo chiều dọc.

Cả azimuth và elevation beamforming đều tận dụng ưu điểm của mảng anten phức tạp để điều khiển chính xác hướng và cường độ của tín hiệu vô tuyến, từ đó cung cấp kết nối tốt nhất và giảm thiểu sự can nhiễu với các tín hiệu khác. Trong mạng 5G, việc sử dụng cả hai kỹ thuật này cho phép mạng lưới không chỉ tăng cường dung lượng và hiệu quả sử dụng phổ mà còn cải thiện đáng kể trải nghiệm người dùng bằng cách đảm bảo độ phủ sóng ổn định và đáng tin cậy, ngay cả trong điều kiện môi trường phức tạp.

1.3.3 Sóng cực ngắn mmWave

Sóng mmWave, hay sóng milimet, là một trong những đột phá quan trọng nhất trong công nghệ truyền thông không dây, đặc biệt là với sự ra đời của mạng 5G. Định nghĩa theo bước sóng từ 1 đến 10 mm, tương ứng với băng tần tần số từ khoảng 30 GHz đến 300 GHz, sóng mmWave mở ra cánh cửa cho tốc độ dữ liệu không tưởng trước đây, cùng với băng thông rộng lớn, mang lại khả năng hỗ trợ một lượng lớn dữ liệu được truyền đi nhanh chóng và hiệu quả [8].

Các tần số trên 6 GHz thường được gọi chung là băng tần mmWave, nhưng trong thực tế băng tần mmWave được các quốc gia xem xét cấp phép sử dụng cho các mạng 5G chỉ sử dụng các tần số bắt đầu từ 24 GHz, trong đó chủ yếu là ở các băng tần 24 GHz, 26 GHz, 28 GHz, 37 GHz, 39 GHz và 47 GHz.



Hình 1-7: Dải tần trải rộng của 5G mmWave

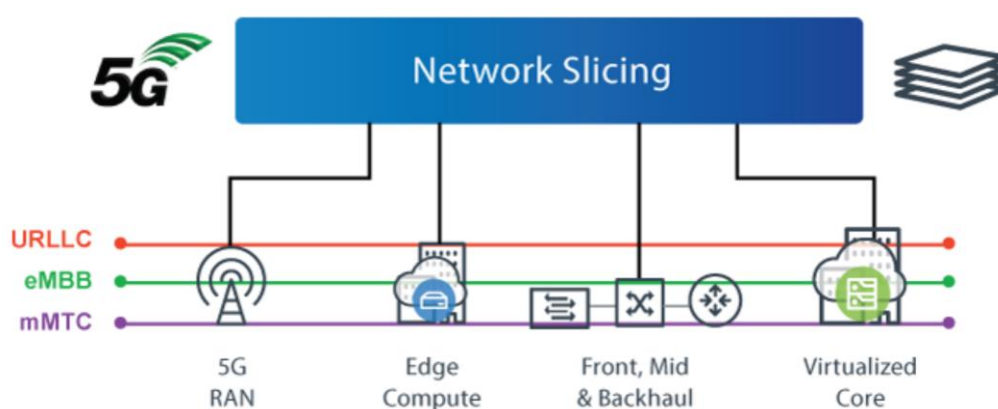
Ứng dụng của sóng mmWave trong mạng 5G là minh chứng rõ ràng nhất cho tiềm năng của công nghệ này. Trong các khu vực đô thị đông đúc, nơi mà nhu cầu về dữ liệu tốc độ cao và dung lượng mạng lớn không ngừng tăng lên, sóng mmWave có thể cung cấp giải pháp hiệu quả. Các nhà mạng có thể sử dụng băng tần này để cung cấp tốc độ truy cập internet siêu nhanh cho người dùng, cho phép tải xuống và tải lên dữ liệu với tốc độ chưa từng có. Hơn nữa, trong lĩnh vực

công nghiệp, sóng mmWave mở ra khả năng cho các ứng dụng yêu cầu độ trễ cực thấp và độ tin cậy cao, như tự động hóa công nghiệp, lái xe tự động, và các hệ thống quản lý giao thông thông minh.

Tuy nhiên, việc triển khai sóng mmWave không phải không có thách thức. Một trong những hạn chế lớn nhất là khả năng truyền đi xa và vượt qua vật cản của sóng mmWave không cao. Sóng này dễ bị suy giảm nhanh chóng qua không khí và gần như không thể đi qua các vật cản vững chắc như tường và cửa sổ. Điều này đòi hỏi một thiết kế mạng cẩn thận và việc triển khai các trạm gốc với mật độ cao hơn để đảm bảo độ phủ sóng đầy đủ và liên tục trong các khu vực dân cư và công nghiệp. Để giải quyết vấn đề này, công nghệ beamforming trở thành một phần không thể thiếu trong việc triển khai sóng mmWave. Beamforming cho phép điều khiển chính xác hướng của tín hiệu vô tuyến, tập trung năng lượng sóng vào các khu vực cụ thể mà không gây ra sự can nhiễu cho các khu vực lân cận. Kỹ thuật này không chỉ cải thiện độ phủ sóng mà còn tăng cường hiệu suất và giảm thiểu nhiễu trong môi trường truyền dẫn.

1.3.4 Phân tách mạng theo lát cắt

Với sự ra đời của mạng 5G, khả năng cung cấp một loạt các dịch vụ từ Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), đến Massive Machine Type Communications (mMTC) đòi hỏi một cách tiếp cận linh hoạt và hiệu quả hơn trong việc quản lý tài nguyên mạng. Trong bối cảnh này, "Network Slicing" đã nổi lên như một giải pháp đầy hứa hẹn.



Hình 1-8: Chia tách một mạng vật lý thành các mạng ảo (nguồn: blueplatnet.com)

Phân tách mạng theo lát cắt (NS – Network Slicing) cho phép tạo ra nhiều mạng ảo độc lập trên cùng một hạ tầng mạng vật lý. Về bản chất, đây là một hình thức ảo hóa mạng cho phép các nhà khai thác chia mạng của họ thành nhiều mạng ảo, độc lập, hay còn gọi là các “lát” (slice). Mỗi "lát" là một mạng riêng biệt với cấu hình và quản lý tài nguyên riêng, được tối ưu hóa cho một loại dịch vụ cụ thể hoặc nhóm người dùng cụ thể. Điều này cho phép các nhà cung cấp dịch vụ mạng đáp ứng nhu cầu đa dạng của người dùng và ứng dụng mà không cần phải xây dựng hạ tầng riêng biệt cho mỗi loại dịch vụ.



Hình 1-9: Các ứng dụng của Network Slicing (nguồn: samsung)

Ví dụ, một slice mạng dành cho URLLC có thể được cấu hình với độ trễ thấp và độ tin cậy cao để hỗ trợ các ứng dụng như lái xe tự động và tự động hóa công nghiệp, trong khi một slice khác dành cho eMBB tập trung vào việc cung cấp băng thông cao cho video độ phân giải cao và dịch vụ truyền thông dữ liệu. Đồng thời, mMTC yêu cầu kết nối đồng thời một lượng lớn thiết bị IoT với yêu cầu băng thông thấp và tiêu thụ năng lượng thấp, điều này cũng có thể được thực hiện thông qua một slice mạng riêng biệt.

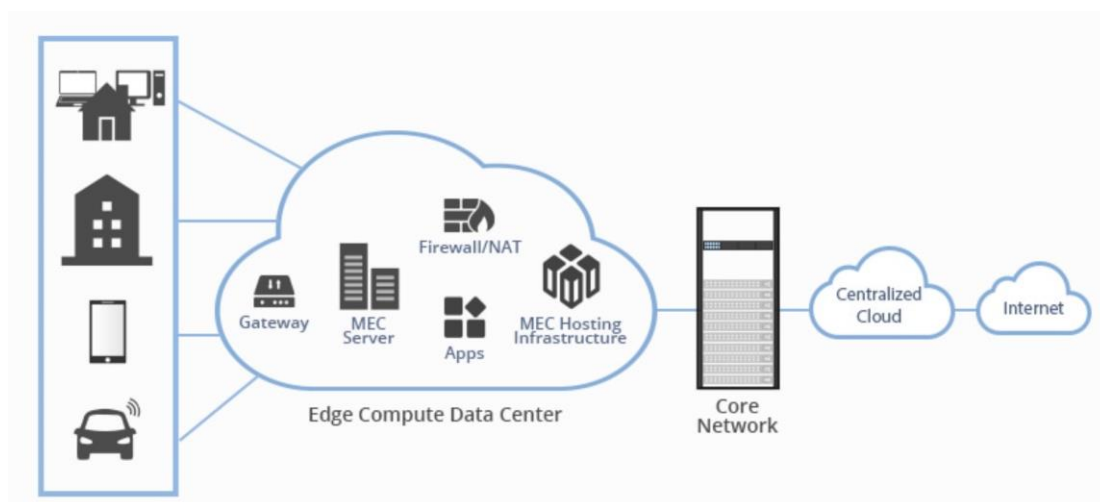
Để triển khai Network Slicing một cách hiệu quả, các nhà cung cấp dịch vụ cần phải sở hữu công nghệ quản lý và điều phối tài nguyên mạng tiên tiến, cùng với khả năng đảm bảo an ninh và bảo mật cho từng slice. Điều này đòi hỏi một hệ thống quản lý mạng (Network Management System - NMS) thông minh, có khả năng tự động hóa việc cấu hình, giám sát và tối ưu hóa các slice mạng dựa trên yêu cầu dịch vụ và điều kiện mạng thực tế. Bên cạnh đó, việc tích hợp

Network Slicing với các công nghệ tiên tiến khác như Massive MIMO, beamforming và edge computing cũng là yếu tố quan trọng giúp tăng cường hiệu suất và hiệu quả của mạng 5G. Massive MIMO và beamforming tăng cường khả năng truyền dẫn dữ liệu và mở rộng phạm vi phủ sóng, trong khi edge computing giúp giảm độ trễ và tải trọng trên mạng lõi bằng cách xử lý dữ liệu gần với người dùng cuối.

Network Slicing trong mạng 5G mở ra cơ hội mới cho các nhà cung cấp dịch vụ trong việc đáp ứng nhu cầu đa dạng của người dùng và các ứng dụng công nghệ cao, từ đó thúc đẩy sự phát triển của xã hội số và kinh tế số. Tuy nhiên, việc triển khai và quản lý Network Slicing cũng đòi hỏi sự hiểu biết sâu rộng về công nghệ, cũng như khả năng đổi mới và thích ứng liên tục từ phía các nhà cung cấp dịch vụ.

1.3.5 Công nghệ MEC

MEC (Multi-access Edge Computing) được biết đến với cái tên điện toán biên đa truy cập là kiểu kiến trúc mạng sử dụng trong lĩnh vực CNTT. Điện toán biên là mô hình dịch vụ và ứng dụng triển khai tại lớp mạng biên (Edge) gần với người dùng.



Hình 1-10: Mô hình của MEC (nguồn vienthongxanh.vn)

Mục đích chính của mô hình điện toán biên là để giảm độ trễ và tăng hiệu suất cho các ứng dụng. Đặc biệt là các ứng dụng tương tác hoặc nhạy cảm với độ trễ như chơi game trực tuyến, IoT hoặc thực tế ảo. Để đạt được này, MEC đưa ứng dụng triển khai tại các nút mạng ngay gần người dùng thay vì triển khai tại

trung tâm dữ liệu.

Về cơ bản thì ta sẽ hiểu rằng, thay vì các ứng dụng và dịch vụ triển khai cơ sở dữ liệu tại các trung tâm dữ liệu. Vậy khi người dùng truy cập và sử dụng dịch vụ, họ sẽ phải đi qua rất nhiều nút mạng mới có thể truy cập vào cơ sở dữ liệu. Rồi dữ liệu trả về cũng phải đi qua nhiều nút mạng như vậy. Càng qua nhiều nút thì độ trễ càng lớn. Tuy rằng các độ trễ chỉ được tính bằng mili giây. Nhưng trong các ứng dụng như chơi game trực tuyến, việc độ trễ tăng sẽ làm giảm đi trải nghiệm người chơi. Hay nếu gọi video call đặc biệt là gọi ra nước ngoài thì việc bị trễ là một điều rất khó chịu với người dùng. MEC hoạt động bằng cách đưa tất cả dịch vụ triển khai ở mạng biên sát với người dùng. Tức là người dùng sẽ không phải đi qua nhiều nút mạng nữa. Điều này làm giảm độ trễ rất nhiều.

Các MEC Server sẽ được đặt tại các trạm cơ sở (BTS) hoặc là một thiết bị mạng gần với người dùng và được tích hợp với mạng di động. MEC sẽ tận dụng tài nguyên cơ ở mạng tại lớp biên để.

Vì MEC tích hợp với mạng di động và đặt tại các trạm BTS. Nên MEC sẽ được hưởng rất nhiều lợi ích từ khả năng truyền dữ liệu tốc độ cao của mạng 5G. Vậy nếu xu hướng mạng di động đang dịch chuyển từ 4G sang 5G thì việc phát triển MEC cũng sẽ thuận lợi hơn rất nhiều. Có thể nói rằng MEC và 5G đang cùng nhau phát triển mạnh mẽ.

1.4 Kết luận chương 1

5G đã phát triển mạnh mẽ từ phiên bản 10 của LTE-Advanced đến phiên bản 17, với những bước đột phá đáng kể. Phiên bản 10 LTE-Advanced mang lại tốc độ cao hơn và khả năng kết nối tốt hơn. Các phiên bản tiếp theo tiếp tục cải thiện hiệu suất, từ tăng cường độ trễ thấp và độ tin cậy trong phiên bản 15, đến khả năng hỗ trợ IoT diện rộng và công nghệ truyền thông cực nhanh trong phiên bản 16. Phiên bản 17 mới nhất mở rộng hơn nữa các tính năng 5G, tối ưu hóa hiệu suất mạng, và hỗ trợ các ứng dụng mới như thực tế ảo và tự động hóa công nghiệp. Trong chương này, tôi đã cung cấp một cái nhìn tổng quan về một số công nghệ cho các mạng truyền thông di động 5G. Massive MIMO cải thiện hiệu suất mạng thông qua việc sử dụng nhiều ăng-ten để truyền và nhận tín hiệu đồng thời, gia tăng tốc độ và dung lượng. Beamforming tập trung tín hiệu vào các

hướng cụ thể, tăng cường hiệu quả truyền dẫn và giảm nhiễu. Sóng cực ngắn mmWave cung cấp băng thông rộng hơn, cho phép truyền dữ liệu tốc độ cao, nhưng yêu cầu hạ tầng dày đặc do phạm vi phủ sóng hạn chế. Network Slicing cho phép tạo ra các lát cắt mạng ảo, tối ưu hóa tài nguyên cho từng ứng dụng khác nhau. Cuối cùng, Multi-access Edge Computing (MEC) giảm độ trễ bằng cách xử lý dữ liệu gần người dùng hơn, cải thiện khả năng phản hồi và hiệu suất cho các ứng dụng thời gian thực.

CHƯƠNG 2: ĐỘ TRỄ TRONG MẠNG 5G

2.1 Giới thiệu

Trong bối cảnh công nghệ thông tin và truyền thông ngày càng phát triển, việc triển khai mạng 5G không chỉ đánh dấu một bước tiến vượt bậc trong lĩnh vực viễn thông mà còn mở ra cánh cửa cho hàng loạt các ứng dụng mới, từ Internet vạn vật (IoT) cho đến tự động hóa công nghiệp và xe tự lái. Trong số các yếu tố quyết định hiệu suất của mạng 5G, độ trễ đóng vai trò cực kỳ quan trọng, vì nó không chỉ ảnh hưởng đến chất lượng trải nghiệm người dùng mà còn là yếu tố quyết định khả năng ứng dụng của công nghệ trong các tình huống yêu cầu độ nhạy cao với thời gian. Do đó, việc nghiên cứu và giảm thiểu độ trễ trong mạng 5G trở thành một trong những chủ đề nghiên cứu quan trọng, đặc biệt là trong bối cảnh các ứng dụng mới đòi hỏi sự đáp ứng nhanh chóng và chính xác.

Độ trễ trong mạng 5G, thường được định nghĩa là thời gian mà một gói tin mất từ khi được gửi từ nguồn đến khi đến được điểm đích, là một thách thức đối với việc đảm bảo hiệu suất mạng ở mức tối ưu. Điều này đặc biệt quan trọng trong các ứng dụng như URLLC, nơi mà độ trễ cực thấp là cần thiết để hỗ trợ các dịch vụ như phẫu thuật từ xa, xe tự lái, và tự động hóa công nghiệp, nơi mà mỗi miligiây đều có giá trị. Trong môi trường sản xuất, ví dụ, sự chậm trễ trong việc truyền thông tin có thể dẫn đến sự cố trong quá trình sản xuất, ảnh hưởng đến an toàn lao động và hiệu quả kinh tế.

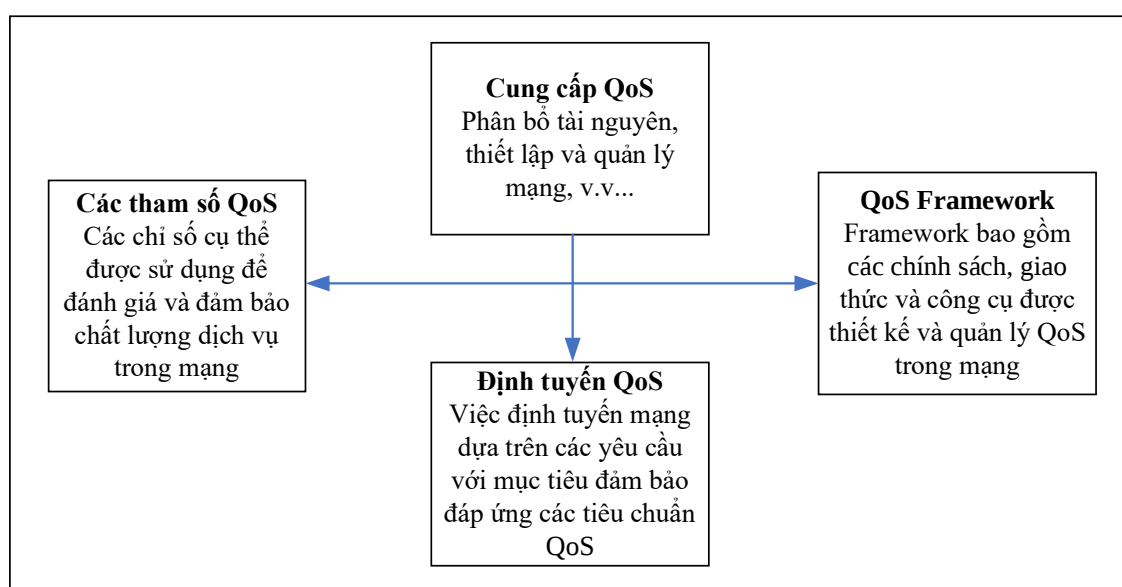
Mạng 5G, với mục tiêu giảm độ trễ xuống dưới một miligiây, đòi hỏi các cải tiến đáng kể về mặt kiến trúc mạng và công nghệ truyền dẫn. Các phương pháp như Định toán biên di động (MEC), Ảo hóa chức năng mạng (NFV), và Mạng được xác định bằng phần mềm (SDN) được coi là giải pháp chính trong việc giảm thiểu độ trễ, bởi chúng cho phép xử lý dữ liệu gần người dùng hơn, giảm bớt tải trọng trên mạng lõi và tối ưu hóa lưu lượng truy cập. Ngoài ra, việc áp dụng các công nghệ như Massive MIMO và beamforming trong mạng 5G cũng góp phần quan trọng trong việc giảm độ trễ, nhờ khả năng tối ưu hóa đường truyền và tăng cường dung lượng mạng.

Tuy nhiên, việc giảm độ trễ không chỉ đòi hỏi sự cải tiến về mặt công nghệ

mà còn cần một sự hiểu biết sâu sắc về yêu cầu dịch vụ cụ thể của từng ứng dụng. Điều này đòi hỏi một hệ thống mạng có khả năng tự động điều chỉnh và tối ưu hóa dựa trên điều kiện mạng và nhu cầu của người dùng, đồng thời duy trì sự cân bằng giữa hiệu suất và chi phí. Trong bối cảnh đó, Network Slicing, một tính năng nổi bật của mạng 5G, cho phép tạo ra các "lát" mạng ảo với các yêu cầu và chính sách quản lý tài nguyên riêng biệt, trở thành một công cụ quan trọng để đáp ứng các yêu cầu đa dạng về độ trễ và hiệu suất trong các ứng dụng khác nhau.

Nội dung của chương sẽ nghiên cứu về độ trễ trong mạng 5G, không chỉ được xem xét từ góc độ thực thi kỹ thuật mà còn từ góc độ ứng dụng thực tiễn. Việc giảm thiểu độ trễ không chỉ cải thiện chất lượng trải nghiệm người dùng mà còn mở ra khả năng triển khai các ứng dụng mới, đóng góp vào sự phát triển của xã hội số và nền kinh tế thông minh. Do đó, nghiên cứu và phát triển các giải pháp nhằm giảm độ trễ trong mạng 5G đang trở thành một trong những ưu tiên hàng đầu trong lĩnh vực viễn thông và công nghệ thông tin.

2.2 Các tham số chất lượng dịch vụ (QoS)



Hình 2-1: Mô hình cung ứng QoS (QoS Provisioning)

Chất lượng của các tham số dịch vụ khác nhau đóng một vai trò quan trọng trong thiết kế của các mạng thế hệ tiếp theo. Các ứng dụng của mạng 5G đòi hỏi lượng dữ liệu rất lớn. Để đáp ứng yêu cầu của các ứng dụng trong mạng 5G, việc bảo đảm các tham số chất lượng dịch vụ (QoS) là một mục tiêu quan trọng (Hình 2-1).

Dưới đây là các tham số QoS thường được sử dụng trong các mạng không dây [9].

2.2.1 Băng thông (Bandwidth)

Băng thông là một trong những tham số chất lượng dịch vụ quan trọng nhất cần thiết để đáp ứng nhu cầu của các ứng dụng trong mạng. Băng thông tỉ lệ thuận trực tiếp với lượng dữ liệu được truyền tải hoặc dung lượng của kênh (tức dung lượng của RAN). Đơn vị đo là bit/s.

2.2.2 Thông lượng (Throughput)

Mức độ hiệu quả của mạng phụ thuộc vào giá trị thông lượng. Thông lượng cao đại diện cho số lượng tối đa các bit được truyền tải một cách thành công so với tổng số bit có sẵn trong toàn bộ bản tin trên mỗi đơn vị thời gian. Các thiết bị phần cứng, băng thông, tỷ lệ tín hiệu trên nhiễu (SNR), và những yếu tố khác ảnh hưởng đến giá trị hiệu suất truyền. Thông lượng đóng vai trò quan trọng trong việc đo lường tốc độ truyền thực tế.

2.2.3 Jitter

Mọi người đều mong muốn truyền dữ liệu liên tục mà không có bất kỳ độ trễ nào, nhưng jitter là một tham số làm trì hoãn toàn bộ một giao tiếp trao đổi dữ liệu của một ứng dụng cụ thể nào đó. Trong truyền dữ liệu gói, Jitter còn được gọi là biến thiên trễ gói (PDV), được hiểu là độ trễ giữa các gói tin. Việc truyền gói tin mà không có độ trễ tối thiểu giữa các gói tin phụ thuộc vào yếu tố này. Yếu tố này phải nằm trong khoảng cho phép, đặc biệt là trong trường hợp của các ứng dụng truyền thông như video và âm thanh trực tuyến. Jitter là một trong những tham số quan trọng để đánh giá chất lượng dịch vụ.

2.2.4 Tỷ lệ lỗi bit (Bit Error Rate - BER)

Thông số BER là tỷ số giữa số bit bị lỗi trên tổng số bit phát đi. Thông số này đặc trưng cho chất lượng truyền dẫn của tuyến. Thông số này phải được giảm thiểu để đáp ứng yêu cầu của các ứng dụng.

2.2.5 Tỷ lệ chuyển giao gói tin (Packet Delivery Ratio - PDR)

Như tên gọi, thông số PDR được tính bằng tỷ lệ giữa số lượng gói tin được nhận thành công so với tổng số gói tin đã được gửi qua mạng. PDR là một

thông số có ý nghĩa nhất để đánh giá độ tin cậy và quyết định hiệu năng thực thi của mạng.

2.2.6 Thời gian khám phá tuyến (Route Discovery Time)

Thời gian cần thiết để xác định một tuyến đường từ nguồn đến đích được gọi là thời gian khám phá tuyến. Trong trường hợp của mạng di động, mạng thường xuyên thay đổi do sự dịch chuyển của các nút mạng. Thuật toán định tuyến phát hiện tuyến đường thích hợp trước khi truyền. Thời gian khám phá tuyến phải là tối thiểu để kết nối các nút mạng khác nhau. Để đáp ứng các ứng dụng thế hệ mới, cần thiết phải triển khai các thuật toán định tuyến thông minh.

2.2.7 Độ trễ (Delay)

Độ trễ là thời gian chờ đợi xảy ra trong một mạng truyền thông. Độ trễ mạng là sự chậm trễ trong giao tiếp mạng. Độ trễ mạng cho biết thời gian cần để truyền dữ liệu qua mạng. Mạng có thời gian trễ hoặc độ ỳ (latency) cao sẽ có độ trễ cao. Các doanh nghiệp muốn hoạt động giao tiếp mạng diễn ra nhanh hơn và có độ trễ thấp để cải thiện năng suất và nâng cao tính hiệu quả trong hoạt động kinh doanh. Rất nhiều ứng dụng yêu cầu độ trễ mạng thấp như các livestream, điều khiển từ xa, các ứng dụng theo thời gian thực. Độ trễ mạng cao làm giảm hiệu năng ứng dụng và khiến ứng dụng có nguy cơ cao xảy ra lỗi. Độ trễ giữa hai đầu cuối từ nguồn đến đích được gọi là Độ Trễ Đầu Cuối (End-to-End Delay - EED).

Độ trễ phụ thuộc vào 4 thành phần, công thức toán học như sau:

$$\text{Độ trễ tổng} = \text{Độ trễ lan truyền} + \text{Độ trễ truyền} + \text{Độ trễ xử lý} + \text{Độ trễ xếp hàng}$$

(2-1)

Dưới đây là giải thích ngắn gọn về từng thành phần:

- *Độ trễ lan truyền (Propagation Delay)*: Thời gian cần thiết để tín hiệu di chuyển từ nguồn đến đích. Nó bị ảnh hưởng bởi khoảng cách giữa các thiết bị và tốc độ của môi trường mà qua đó tín hiệu được truyền đi.
- *Độ trễ truyền (Transmission Delay)*: Thời gian cần thiết để đẩy các bit lên mạng. Nó bị ảnh hưởng bởi tốc độ của kết nối mạng.

- *Độ trễ xử lý (Processing Delay)*: Thời gian cần thiết cho các bộ định tuyến, switch và các thiết bị mạng khác để xử lý và chuyển tiếp dữ liệu.
- *Độ trễ xếp hàng (Queueing Delay)*: Thời gian một gói tin phải chờ đợi trong hàng đợi trước khi có thể được truyền đi. Điều này xảy ra khi có sự tắc nghẽn trong mạng.

Để đạt được một mạng chất lượng tốt, độ trễ phải nằm trong khoảng tối thiểu. Nghiên cứu sẽ tập trung vào tìm hiểu về độ trễ trong mạng 5G. Ví dụ như để kích hoạt các dịch vụ liên quan đến Internet xúc giác (Tactile Internet) thì độ tin cậy cao và độ trễ thấp là điều kiện tiên quyết. Độ trễ, được tính từ thời điểm thông tin hoặc gói tin bắt đầu được gửi từ nguồn cho đến khi gói tin đó đến được đích. Đây là chỉ số quan trọng đánh giá thời gian truyền dữ liệu trong mạng.

Trong nỗ lực đáp ứng yêu cầu khắt khe về độ trễ, một loạt các biện pháp đã được triển khai để giảm thiểu thời gian truyền không dây. Một trong số đó là đề xuất giảm độ trễ truyền từ mức 1 ms, tiêu chuẩn trong LTE, xuống còn 0.1-0.25 ms hoặc thậm chí thấp hơn. Điều này đòi hỏi các phương pháp điều chế và truyền sóng phải được thiết kế để hỗ trợ gói tin ngắn và đảm bảo độ trễ thấp. Các hệ thống truyền thông rộng băng thông hiện tại chủ yếu dựa trên OFDM do khả năng chống chịu sự suy giảm do truyền đa đường. Tuy nhiên, để đạt được mục tiêu truyền dữ liệu với độ trễ thấp trong khoảng một mili giây, cần phải có gói tin cực kỳ nhỏ, yêu cầu một cơ chế truyền vật lý với chu kỳ truyền chỉ kéo dài vài micro giây. Điều này, không chỉ giải quyết các vấn đề về thiết kế ký hiệu tham chiếu, ước lượng kênh, mã hóa kênh mà còn phải xem xét kỹ lưỡng thiết kế của lớp vật lý (PHY) cho ứng dụng Internet xúc giác trong hệ thống 5G. Một giải pháp để vượt qua những thách thức này và đạt được mục tiêu độ trễ 1 ms là hệ thống dựa trên OFDM với số liệu học (OFDM numerology) thực hiện các kỹ thuật điều chỉnh trong OFDM, bao gồm thời gian ký hiệu, khoảng cách giữa các tần số con, cũng như độ dài của khung phụ và ước lượng kênh với thời gian ngắn nhất cùng với việc giải mã kênh hiệu quả (ví dụ, sử dụng mã chập – Convolutional code). Điều này đã được chứng minh là một hệ thống dựa trên OFDM với số liệu học được điều chỉnh có khả năng đạt được độ trễ 1 ms là khả thi [10].

Để hỗ trợ sự tồn tại đồng thời của các ứng dụng cùng với việc đạt được độ trễ 1 ms, cần phải có một cấu trúc khung linh hoạt. Phương pháp được đề xuất có

thể là điều chỉnh động Thời gian truyền (TTI) phù hợp với các yêu cầu dịch vụ cụ thể, cung cấp một giải pháp mềm dẻo cho việc quản lý độ trễ trong mạng 5G, từ đó tối ưu hóa trải nghiệm người dùng và hiệu quả của hệ thống truyền thông.

2.3 Yêu cầu về hiệu năng kỹ thuật cho 5G

Dựa trên những khả năng chính và IMT-2020, yêu cầu về hiệu suất kỹ thuật được định nghĩa trong khuyến nghị ITU-R M.2410 [11]. Các yêu cầu về hiệu suất kỹ thuật được tóm tắt trong Bảng 2-1. Chi tiết về yêu cầu về hiệu suất kỹ thuật để đạt được mục tiêu thực thi mạng 5G hiện đang được tổ chức ITU-R, 3GPP tiếp tục nghiên cứu theo các kịch bản triển khai cụ thể. Chẳng hạn như các yêu cầu liên quan đến ba kịch bản sử dụng trong tài liệu 3GPP TR 38.913 cho thấy mong muốn của 3GPP trong việc thể hiện năng lực cung cấp dịch vụ cao hơn so với yêu cầu của ITU. Mô tả chi tiết về yêu cầu của 3GPP vượt quá phạm vi của nghiên cứu này do vậy sẽ không được đề cập chi tiết ở đây.

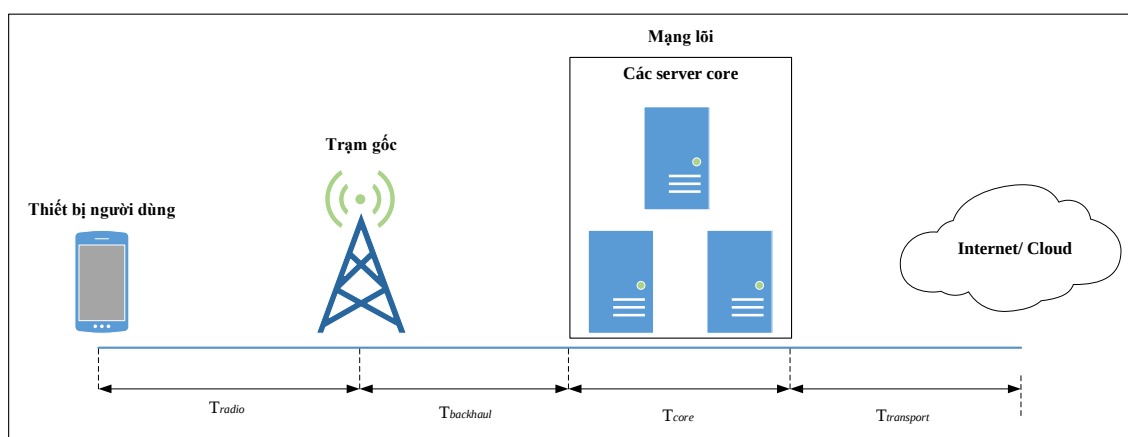
Bảng 2-1: Yêu cầu hiệu suất kỹ thuật URLLC

Yêu cầu về hiệu suất kỹ thuật	Tải xuống (Download)	Tải lên (Upload)	So sánh với yêu cầu của IMT-Advanced
Độ trễ mặt phẳng người dùng (User plane)	1 ms	1 ms	Giảm hơn 10 lần so với IMT-Advanced
Độ trễ mặt phẳng điều khiển (Control plane)	20 ms	20 ms	Giảm hơn 5 lần so với IMT-Advanced
Thời gian gián đoạn di động	0	0	
Độ tin cậy	99.9999% trong 1 ms	99.9999% trong 1 ms	

2.4 Nguồn trễ trong mạng di động

Trong hệ thống mạng di động, độ trễ có thể được chia thành hai phần chính:

- Độ trễ của mặt phẳng người dùng (U-plane)
- Độ trễ của mặt phẳng điều khiển (C-plane)



Hình 2-2: Độ trễ của truyền tin trong mạng di động

Việc đo lường độ trễ của mặt phẳng người dùng bằng thời gian truyền một chiều của một gói tin để trở nên khả dụng ở tầng IP giữa biên của Mạng truy nhập vô tuyến mặt đất cải tiến (E-UTRAN)/UE và nút mạng E-UTRAN. Độ trễ của mặt phẳng điều khiển được định nghĩa là thời gian chuyển đổi của một UE từ trạng thái chờ đợi sang trạng thái hoạt động. Ở trạng thái Chờ (Idle), một UE không kết nối với kiểm soát tài nguyên vô tuyến (RRC). Sau khi kết nối RRC được thiết lập, UE chuyển từ trạng thái Chờ sang trạng thái Đã kết nối và sau đó chuyển sang trạng thái Active. Mặt phẳng người dùng (U-plane) là trọng tâm cho giao tiếp có độ trễ thấp, vì hiệu suất ứng dụng chủ yếu phụ thuộc vào độ trễ của mặt người dùng [12].

Độ trễ của việc truyền một gói tin trong mạng di động có thể được đạt được bằng cách kết hợp RAN, backhaul, mạng lõi và trung tâm dữ liệu/Internet. Như được tham chiếu trong Hình 2-2.

Sau khi phân tích hình vẽ, có thể thấy tổng thời gian truyền một chiều cho gói tin trong hệ thống di động được tính toán theo công thức sau [12]:

$$T = T_{radio} + T_{backhaul} + T_{core} + T_{transport} \quad (2-2)$$

Trong đó:

T_{radio} là thời gian truyền một gói tin giữa eNB/gNB và UEs và chủ yếu ở

lớp truyền thông tầng vật lý. Nó được đóng góp bởi eNBs/gNBs, UEs và môi trường. Nó bao gồm thời gian truyền, thời gian xử lý tại eNB/UE, các lần truyền lại, và độ trễ truyền đi. Độ trễ xử lý tại eNB bao gồm mã hóa kênh, khớp tốc độ, trộn lẫn, kiểm tra đa dạng chu kỳ (CRC), ghép trước, ánh xạ Modulation, ánh xạ Layer, ánh xạ phân tử nguồn, và tạo tín hiệu OFDM.

$T_{backhaul}$ là thời gian để xây dựng kết nối giữa eNB/gNB và mạng lõi (tức là EPC/CN). Thông thường, mạng lõi và eNB/gNB được kết nối bằng cáp đồng hoặc sợi quang.

T_{core} là thời gian xử lý của mạng lõi. Nó được đóng góp bởi các thực thể mạng lõi khác nhau như thực thể quản lý di động (MME), nút hỗ trợ GPRS phục vụ (SGSN) và SDN/NFV. Các bước xử lý của mạng lõi bao gồm bảo mật NAS, kiểm soát EPS bearer, gắn kết di động, phân phối địa chỉ IP của UE, xử lý di động ở trạng thái chờ đợi, và lọc gói tin.

$T_{transport}$ là độ trễ giao tiếp dữ liệu giữa mạng lõi và Internet/đám mây. Nó phụ thuộc vào khoảng cách giữa mạng lõi và máy chủ, băng thông và giao thức giao tiếp để ảnh hưởng đến độ trễ này.

Công thức ở trên thể hiện thời gian truyền một chiều, như vậy với giao tiếp giữa 2 thiết bị đầu cuối (E2E) thì thời gian sẽ là khoảng $TE2E = T \times 2$.

T_{radio} là tổng của thời gian truyền, thời gian lan truyền, thời gian xử lý ước lượng kênh, thời gian mã hóa và giải mã (cho lần đầu tiên) và thời gian truyền lại (do mất gói tin). Cụ thể, thời gian T_{radio} cho một người dùng có thể được biểu diễn như sau:

$$T_{radio} = T_q + T_f + T_{trans} + T_{proc} + T_{mpt} \quad (2-3)$$

Trong đó:

T_q là thời gian hàng đợi phụ thuộc vào số người dùng được nhiều kênh.

T_f là thời gian cho việc điều chỉnh khung, phụ thuộc vào cấu trúc và chế độ (chế độ FDD, chế độ TDD).

T_{trans} là thời gian xử lý truyền và truyền dữ liệu, sử dụng ít nhất một TTI (Time Transmission Interval) - phụ thuộc vào điều kiện kênh vô tuyến, kích thước tải, tài nguyên có sẵn, lỗi truyền và truyền lại.

T_{proc} là thời gian xử lý tại trạm cơ sở.

T_{mpt} là thời gian xử lý của thiết bị người dùng. Cả thời gian tại trạm cơ sở và thiết bị người dùng đều phụ thuộc vào năng lực của trạm gốc và thiết bị.

2.5 Tầng vật lý URLLC trong 5G NR

Trái ngược với các hệ thống 4G LTE, yêu cầu về độ trễ, độ tin cậy và thông lượng cần được xem xét đồng bộ trong 5G NR, điều này đòi hỏi một sự thay đổi cơ bản trong kiến trúc tầng vật lý (gói, slot và khung). Cụ thể, cần phải đề xuất được một giải pháp với cấu trúc gói có độ trễ linh hoạt cho quá trình giải mã nhanh trong một cấu trúc khung linh hoạt.

Một trong những giải pháp như vậy là cơ chế đặt trước tài nguyên dựa trên sơ đồ lập lịch (URLLC). Cơ chế này cho phép thiết lập hai chế độ đặt trước tài nguyên là chế độ bán tĩnh và chế độ động. Trong chế độ bán tĩnh, trạm gốc thi thoảng mới phát đi thông tin về cấu hình như cấu trúc khung, khung tần số, chu kỳ dịch vụ v.v.. Trong khi đó chế độ đặt chỗ động thông tin về tài nguyên URLLC được cập nhật thường xuyên bằng kênh điều khiển của người dùng đã lên lịch. Ví dụ, nếu một gói MBB bao gồm 14 ký hiệu thì 10 ký hiệu được sử dụng để truyền eMBB và phần còn lại được dành riêng cho URLLC. Nhược điểm của sơ đồ này là khi không có truyền URLLC trong chu kỳ đã lên lịch, các tài nguyên được đặt riêng cho URLLC sẽ bị lãng phí. Khi so sánh với sơ đồ đặt chỗ bán tĩnh hoặc bán tĩnh, việc đặt chỗ động yêu cầu thêm chi phí điều khiển để chỉ ra thông tin đặt tài nguyên trước. Ngoài ra, để đảm bảo độ tin cậy của tín hiệu điều khiển thì bản thân chi phí điều khiển cũng phải được tính đến.

Ngoài ra, khi dịch vụ URLLC được khởi động, gói URLLC sẽ được truyền ngay lập tức mà không có bất kỳ sự chậm trễ nào. Để đạt được điều này, cần phải có một kế hoạch cụ thể giúp giảm thiểu độ trễ truyền của gói URLLC, bao gồm cả cơ chế để giảm độ trễ khi truyền lại. Ở đây cần nhấn mạnh tầm quan trọng của việc thiết kế và tối ưu hóa lớp vật lý cho dịch vụ URLLC trong mạng 5G NR, một yếu tố quan trọng để đảm bảo trải nghiệm người dùng cuối và hiệu năng thực thi dịch vụ không bị gián đoạn trong môi trường truyền thông hiện đại. Điều này không chỉ yêu cầu sự đổi mới về mặt kỹ thuật mà còn cần có sự hiểu biết sâu sắc về yêu cầu của các ứng dụng thời gian thực và khả năng ứng phó linh hoạt với

các thách thức trong môi trường mạng đa dạng [13].

2.6 Yêu cầu về các chỉ số đánh giá URLLC

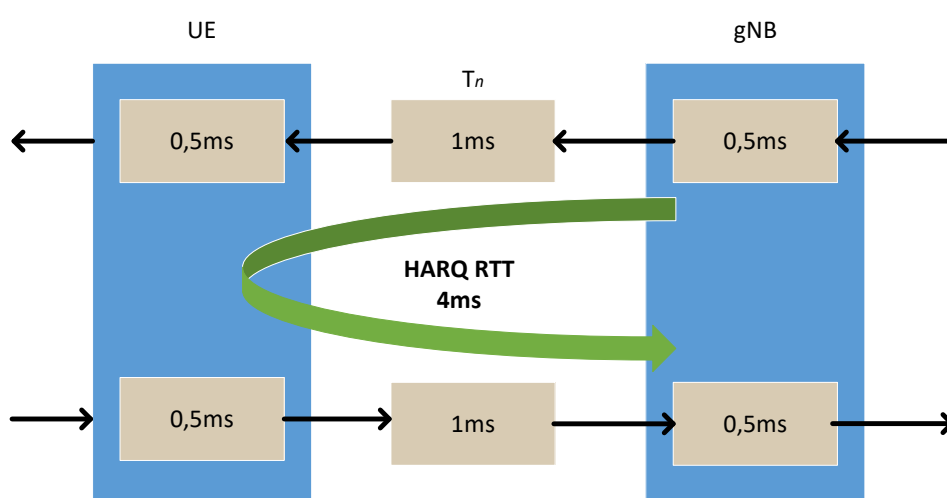
2.6.1 Độ trễ mặt phẳng người dùng (User Plane)

Theo 3GPP TR 38.913, độ trễ mặt phẳng người dùng được định nghĩa là thời gian cần thiết để chuyển một gói tin của tầng ứng dụng từ điểm vào (entry point) của lớp giao thức vô tuyến cho đến điểm thoát (exit point) lớp giao thức vô tuyến cả trong hướng UL và DL theo hướng giao diện vô tuyến, trong đó cả thiết bị và trạm gốc không bị hạn chế bởi chế độ DRX (Discontinuous Reception, một chế độ trong đó UE ngủ trong một khoảng thời gian).

- Theo ITU, yêu cầu độ trễ mặt phẳng người dùng cho eMBB là 4ms và cho URLLC là 1ms trong NR.
- Theo 3GPP, yêu cầu độ trễ mặt phẳng người dùng cho eMBB là 4ms và cho URLLC là 0.5ms trong NR.

Yêu cầu về độ trễ URLLC là rất nghiêm ngặt và cần một kiến trúc/mô hình mạng mạnh mẽ/linh hoạt để đạt được yêu cầu này. URLLC là một trong những yếu tố kích hoạt cho cuộc cách mạng công nghiệp thứ tư nhưng đòi hỏi khoản đầu tư lớn từ các công ty viễn thông để cung cấp dịch vụ độ trễ thấp [12].

Mô hình dưới đây có thể sử dụng để tính toán độ trễ mặt phẳng người dùng, mô hình này có thể áp dụng chung cho cả LTE và 5G.



Hình 2-3: Đánh giá độ trễ mặt phẳng người dùng

HARQ là yêu cầu lặp lại tự động hỗn hợp. RTT là thời gian vòng tròn hoàn chỉnh mà một gói tin cần để đi từ nguồn đến đích và quay trở lại nguồn sau khi nhận được xác nhận (ACK - acknowledgement) hoặc yêu cầu truyền lại (NACK - negative acknowledgement) từ đích. Đây là một chỉ số quan trọng đối với việc đo lường hiệu suất và độ trễ của một mạng di động.

Công thức toán học cụ thể để tính HARQ RTT trong mặt phẳng người dùng có thể phụ thuộc vào nhiều yếu tố, bao gồm cấu hình mạng, độ trễ truyền dẫn, độ trễ xử lý, và cơ chế HARQ được sử dụng. Tuy nhiên, không có một công thức "chuẩn" duy nhất cho tất cả các môi trường mạng, vì HARQ RTT có thể bị ảnh hưởng bởi các yếu tố động như tình trạng mạng, chất lượng kênh truyền, và cấu hình cụ thể của giao thức HARQ. Cách tính HARQ RTT có thể ví dụ như công thức sau:

$$HARQ\ RTT = \text{Thời gian truyền} + \text{Thời gian xử lý} + \text{Thời gian truyền} + \text{ACK/NACK} + \text{Thời gian chờ HARQ} \quad (2-4)$$

Tuy nhiên, trong thực tế, việc tính toán HARQ sẽ phức tạp hơn và cần thêm thông tin chi tiết về mạng và các công cụ đo lường.

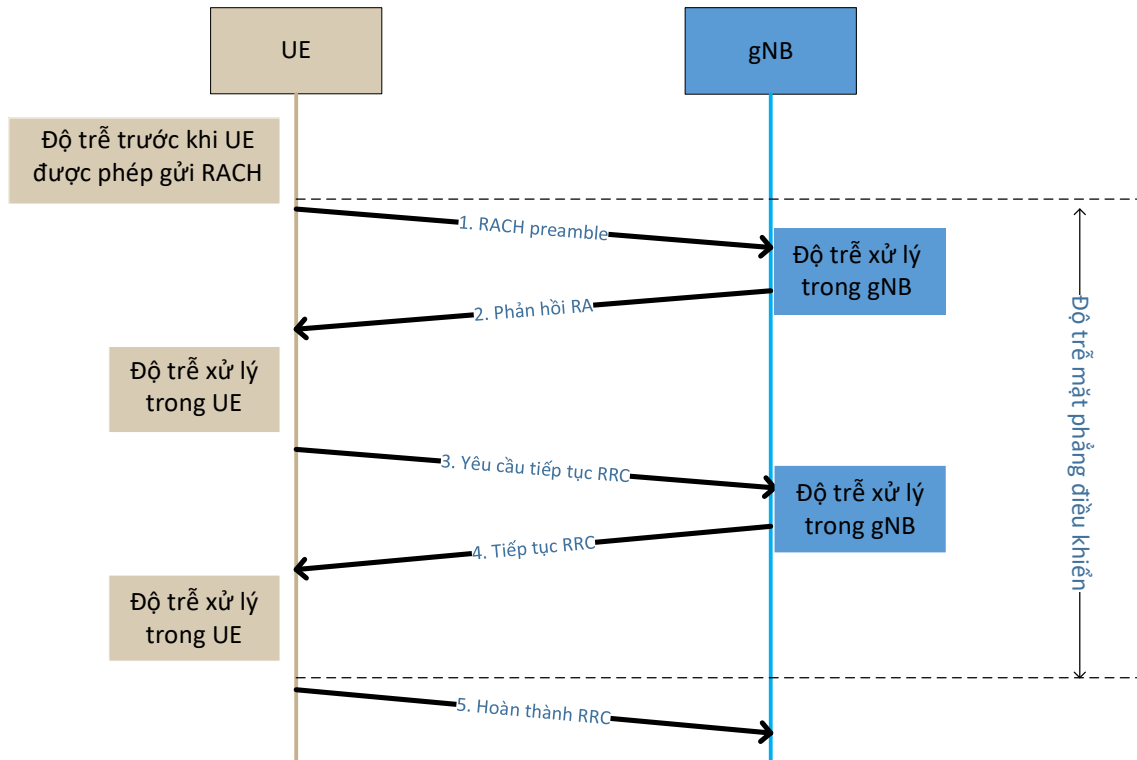
2.6.2 Độ trễ mặt phẳng điều khiển (Control plane)

Theo 3GPP 38 913, độ trễ mặt phẳng điều khiển là thời gian để chuyển từ trạng thái tiết kiệm pin (ví dụ IDLE) sang bắt đầu trạng thái truyền dữ liệu liên tục (ví dụ ACTIVE). Mục tiêu cho độ trễ này là 10ms. Đối với 3GPP 5G NR, độ trễ mặt control plane có thể được đánh giá từ trạng thái RRC_INACTIVE đến trạng thái RRC_CONNECTED. Hình 2-6 cung cấp quy trình RACH, quy trình này được sử dụng để UE chuyển trạng thái.

- UE gửi một tín hiệu truy cập ngẫu nhiên đến mạng thông qua kênh RACH. Tín hiệu này được gọi là "preamble". RACH preamble là một tín hiệu cố định được sử dụng bởi UE để thông báo cho trạm gốc về ý định truy cập mạng của nó. Preamble giúp trạm gốc nhận biết có UE nào đang cố gắng thiết lập kết nối
- Sau khi nhận được tín hiệu preamble, trạm gốc (gNodeB) phản hồi với một thông điệp chấp nhận truy cập, chỉ định tài nguyên cụ thể cho UE để gửi thông tin bổ sung cần thiết cho việc thiết lập kết nối thông qua quy trình

thiết lập RRC (Radio Resource Control).

- UE sử dụng tài nguyên được chỉ định để gửi thông tin chi tiết hơn về yêu cầu truy cập của mình, bao gồm loại dịch vụ và khả năng của thiết bị. Hoàn thành kết nối.



Hình 2-4: Quy trình RACH

Quy trình RACH quan trọng vì nó là bước đầu tiên trong việc thiết lập kết nối mạng, cho phép UE từ trạng thái chờ (Idle) chuyển sang trạng thái hoạt động (Active). Độ trễ trong quy trình RACH có thể ảnh hưởng đến tổng độ trễ mặt phẳng điều khiển, do đó việc tối ưu hóa và giảm thiểu thời gian trong quy trình RACH là quan trọng để cải thiện hiệu suất và trải nghiệm người dùng trong mạng di động [14].

2.6.3 Độ tin cậy

Độ tin cậy liên quan đến khả năng truyền dữ liệu của một luồng dữ liệu trong một khoảng thời gian đã xác định trước với xác suất thành công cao.

Độ tin cậy là xác suất truyền thành công của một gói tin tầng 2/tầng 3 trong một khoảng thời gian yêu cầu cụ thể, đó là thời gian cần thiết để chuyển một gói dữ liệu nhỏ từ điểm vào SDU của lớp giao thức vô tuyến tầng 2/tầng 3 đến điểm

ra SDU của lớp giao thức vô tuyến tầng 2/tầng 3 của giao diện vô tuyến tại một chất lượng kênh nhất định [15]. Yêu cầu này được định nghĩa với mục đích đánh giá kịch bản sử dụng URLLC.

Yêu cầu về độ tin cậy thường được xác định là khả năng truyền thành công một gói tin trong khoảng thời gian cụ thể với xác suất rất cao. 3GPP, là tổ chức đứng sau việc phát triển các tiêu chuẩn cho 5G đã xác định yêu cầu độ tin cậy cho URLLC là 99.999% ($1-10^{-5}$ xác suất lỗi) trong vòng 1ms cho môi trường thử nghiệm đô thị Macro-URLLC. Điều này có nghĩa là, trong hầu hết các trường hợp, gần như tất cả các gói tin phải được truyền một cách đáng tin cậy trong khoảng thời gian dưới 1ms, ngay cả ở biên của vùng phủ sóng, nơi chất lượng tín hiệu có thể không lý tưởng.

2.6.4 Tính di động và thời gian gián đoạn di động

Tính di động (mobility) khả năng của mạng và thiết bị người dùng (UE) để duy trì kết nối mạng liên tục và chất lượng dịch vụ trong khi di chuyển qua các khu vực phủ sóng khác nhau hoặc giữa các cell khác nhau. Tiêu chuẩn này đặc biệt quan trọng trong việc xác định các yêu cầu và đặc tính kỹ thuật cho mạng 5G New Radio (NR) để hỗ trợ di động.

Các lớp di động được định nghĩa như sau:

- Tĩnh: không động (tốc độ 0 km/giờ)
- Người đi bộ: từ 0 km/giờ đến 10 km/giờ
- Xe chuyển động: từ 10 km/giờ đến 120 km/giờ
- Xe chuyển động nhanh: từ 120 km/giờ đến 500 km/giờ

Di động nhanh đến 500 km/giờ chủ yếu dành cho tàu hỏa chuyển động nhanh (high speed train) [12].

Thời gian gián đoạn di động (Mobility Interruption Time) là khoảng thời gian mà kết nối dữ liệu của người dùng bị gián đoạn trong quá trình chuyển giao (handover) từ một trạm gốc này sang trạm gốc khác hoặc từ một loại mạng này sang loại mạng khác (ví dụ, từ LTE sang 5G). Đây là khoảng thời gian từ khi kết nối dữ liệu cuối cùng bị mất ở trạm gốc ban đầu cho đến khi kết nối dữ liệu được thiết lập thành công ở trạm gốc mới và dữ liệu có thể tiếp tục được truyền.

Mạng di động hiện đại như 4G LTE và 5G NR, việc giảm thiểu thời gian gián đoạn để đảm bảo trải nghiệm người dùng liền mạch, đặc biệt là cho các ứng dụng đòi hỏi độ trễ thấp và độ tin cậy cao như video trực tuyến, trò chơi trực tuyến, và các ứng dụng của Internet vạn vật (IoT). Thời gian gián đoạn thấp giúp đảm bảo rằng người dùng không nhận thấy sự gián đoạn hoặc mất mát chất lượng dịch vụ khi di chuyển qua các khu vực phủ sóng.

Trong mạng 5G, với các cải tiến về kiến trúc mạng và công nghệ, nhằm mục tiêu đạt được thời gian gián đoạn cực kỳ thấp, thậm chí dưới 1 mili giây cho các kịch bản URLLC và eMBB.

2.6.5 Tốc độ dữ liệu đỉnh

Tốc độ dữ liệu đỉnh (Peak data rate), là tốc độ truyền dữ liệu tối đa mà một mạng viễn thông có thể cung cấp dưới điều kiện lý tưởng. Đây là một chỉ số quan trọng đánh giá khả năng của mạng trong việc xử lý dữ liệu với tốc độ cao nhất có thể, và thường được sử dụng để so sánh hiệu suất giữa các công nghệ mạng khác nhau.

Trong mạng 4G LTE, 5G NR và các công nghệ viễn thông khác, tốc độ dữ liệu đỉnh (peak data rate) phản ánh khả năng của mạng trong các tình huống lý tưởng nhất, nơi không có sự cản trở nào từ nhiễu, tắc nghẽn mạng, hoặc các yếu tố giảm chất lượng kênh. Nó thường được đo khi một thiết bị người dùng duy nhất (UE) kết nối với mạng và không có bất kỳ thiết bị nào khác cạnh tranh tài nguyên vô tuyến. Ví dụ trong mạng LTE, tốc độ đỉnh có thể đạt tới 1 Gbps (Gigabit per second) hoặc cao hơn với sự hỗ trợ của các kỹ thuật như carrier aggregation, MIMO (Multiple Input Multiple Output) hoặc các phương pháp điều chế tiên tiến. Đối với mạng 5G NR, tốc độ đỉnh được kỳ vọng sẽ vượt qua mức 1Gbps, hướng tới mục tiêu thiết kế lên tới 20 Gbps hoặc cao hơn nhờ vào việc sử dụng các băng tần vô tuyến mới, kỹ thuật truyền dẫn tiên tiến và việc sử dụng băng thông rộng hơn.

Tốc độ dữ liệu đỉnh không chỉ phản ánh khả năng truyền dữ liệu nhanh chóng của mạng mà còn là một chỉ số quan trọng cho các ứng dụng đòi hỏi băng thông cao như truyền phát video độ phân giải cao, trò chơi trực tuyến và các ứng dụng thực tế ảo/augmented reality (VR/AR).

2.7 Các ràng buộc và phương pháp để đạt được độ trễ thấp

Ở phần này chúng ta sẽ tìm hiểu về các phương pháp và ràng buộc của việc đạt được độ trễ thấp. Có sự đánh đổi cơ bản giữa các khả năng như mức độ phủ sóng, độ trễ, độ tin cậy và hiệu suất sử dụng tần số trong mạng không dây. Điều này có nghĩa là nếu một yếu tố được tối ưu hóa để cải thiện hiệu năng của nó có thể dẫn đến sự suy giảm hiệu năng của một yếu tố khác. Vì độ trễ liên quan đến việc control overhead (tiền tố lặp CP, chế độ truyền và pilot symbol) chiếm một phần lớn thời gian truyền của một gói tin (khoảng 0.4 ms mỗi lần truyền gói) nên việc xem xét một gói tin với thời gian truyền vô tuyến ít hơn 1 ms là không cần thiết. Nếu chúng ta thiết kế một gói tin với thời gian truyền là 0.5 ms, hơn 60% tài nguyên sẽ được sử dụng bởi control overhead. Hơn nữa, việc truyền lại mỗi lần truyền gói mất khoảng 8 ms và việc loại bỏ việc truyền lại sẽ ảnh hưởng đáng kể đến lỗi gói tin. Do đó, chúng ta cần có những sửa đổi và cải tiến toàn diện trong cấu trúc và chiến lược truyền gói. Có thể liệt kê như sau [12]:

- Đầu tiên, cần thiết kế một khung truyền dẫn vô tuyến mới bằng việc giảm control overhead và có thời gian truyền nhỏ hơn. Điều này có thể đạt được bằng cách giảm số lượng thông điệp điều khiển cần thiết cho việc phân bổ tài nguyên, hoặc thậm chí kết hợp hoặc loại bỏ hoàn toàn các quy trình này. Việc giảm thiểu control overhead và tối ưu hóa lập lịch người dùng giúp giảm độ trễ tổng thể.
- Thứ hai, giảm xác suất lỗi gói tin trong lần truyền đầu tiên bằng các kỹ thuật truyền mới giảm thiểu độ trễ do truyền lại và sử dụng các dạng sóng mới. Điều này giúp tăng cường hiệu quả của lần truyền đầu tiên, giảm nhu cầu về truyền lại, từ đó giảm độ trễ.
- Thứ ba, cần phải xác định các kỹ thuật để ưu tiên dữ liệu quan trọng hơn so với dữ liệu bình thường, đảm bảo rằng dữ liệu quan trọng được gửi đi ngay lập tức. Điều này đặc biệt quan trọng trong các ứng dụng đòi hỏi độ trễ thấp và độ tin cậy cao.
- Thứ tư, đồng bộ hóa và tính chất trực giao là những yếu tố không thể thiếu của OFDM, nhưng đây cũng là rào cản lớn để đạt được độ trễ thấp. Mặc dù chế độ giao tiếp không đồng bộ có thể có lợi hơn về mặt độ trễ so với hoạt động đồng bộ, nó đòi hỏi nguồn lực về phổ tần và năng lượng

bổ sung.

- Thứ năm, độ trễ trong truyền dẫn dữ liệu cũng phụ thuộc vào độ trễ giữa mạng lõi và trạm gốc. Mạng đệm (Caching network) có thể được sử dụng để lưu trữ dữ liệu phổ biến tại biên mạng. Điều này cho phép dữ liệu được truy cập nhanh chóng hơn khi cần thiết, giảm độ trễ trong việc truy cập dữ liệu.

Căn cứ vào các yêu cầu trên, hiện tại có thể đề xuất 3 giải pháp chính sau:

- 1) Giải pháp RAN
- 2) Giải pháp mạng lõi
- 3) Giải pháp caching.

2.7.1 Giải pháp RAN

Các giải pháp RAN (Radio Access Network) nhằm mục đích tối ưu hóa mạng truy nhập vô tuyến để đáp ứng các yêu cầu về hiệu suất, độ tin cậy và độ trễ thấp của mạng 5G. Việc thiết kế cấu trúc khung hoặc gói tin mới giúp tăng hiệu quả truyền dẫn dữ liệu, trong khi các thiết kế dạng sóng và kỹ thuật truy cập đa dạng hỗ trợ việc truyền dẫn đồng thời cho nhiều người dùng. Kỹ thuật điều chế và mã hóa mmWave mở ra cơ hội sử dụng băng tần cao hơn, nơi có nhiều băng thông sẵn có nhưng đòi hỏi công nghệ tiên tiến để đối phó với thách thức về phạm vi truyền và sự thay đổi của môi trường. QoS và QoE được củng cố để đảm bảo rằng người dùng cuối nhận được dịch vụ đáng tin cậy và chất lượng trải nghiệm cao, trong khi các kỹ thuật giao tiếp nhận biết vị trí và RAN đám mây nhằm vào việc tối ưu hóa quản lý tài nguyên và dịch vụ dựa trên vị trí. Một số giải pháp nhằm tối ưu RAN được giới thiệu ở dưới đây [12].

2.7.1.1 Kiến trúc khung

Trong các giải pháp RAN, lựa chọn thông thường là điều chỉnh giao diện vật lý, hầu hết các giải pháp được đề xuất đều tập trung vào các lớp kiểm soát truy cập trung bình (MAC). Trong mạng di động LTE, thời lượng của khung radio là 10 ms. Một khung được chia thành 10 subframe với kích thước 1 ms, tiếp theo được chia thành các đơn vị 0.5 ms được gọi là một khối tài nguyên (RB). Mỗi RB kéo dài 0.5 ms (6 hoặc 7 ký hiệu OFDM) trong miền thời gian và 180 KHz (12 tần số con liên tiếp, mỗi tần số con có độ rộng 15 KHz) trong miền tần số. Dựa

trên điều này, khoảng cách subcarrier (Δf) là 15 KHz, thời gian ký hiệu OFDM Tofdm là $1/\Delta f = 66.67 \mu s$, kích thước FFT là 2048, tốc độ lấy mẫu f_s là:

$$\Delta f \times N_{IFFT} = 33.72 \text{ MHz} \quad (2-5)$$

với khoảng thời gian lấy mẫu T_s là $1/f_s$.

Vì vậy, để giảm TTI và đạt được thời gian đáp ứng thấp, khoảng cách subcarrier Δf có thể được thay đổi thành 30 KHz. Điều này dẫn đến thời gian ký hiệu OFDM tương ứng T_{OFDM} là $33.33 \mu s$ và kích thước FFT N_{fft} sẽ là 1024, trong khi tốc độ lấy mẫu f_s được giữ 30.72 MHz tương tự như hệ thống LTE. Thời lượng của khung $T_s = 10 \text{ ms}$ có thể được chia thành 40 subframes trong đó mỗi subframe có thời lượng T_{sf} là 0.25 ms và chứa 6 hoặc 7 ký hiệu. Hai loại tiền tố lặp CP có thể được sử dụng trong cấu hình này theo thời gian.

$$T_{cp1} = \frac{5}{64} \times N_{IFFT} \times T_s \approx 2.604 \mu s \quad (2-6)$$

$$T_{cp2} = \frac{4}{64} \times N_{IFFT} \times T_s \approx 2.083 \mu s \quad (2-7)$$

2.7.1.2 Điều chế và mã hóa kênh

Mặc dù việc sử dụng các gói tin nhỏ là một phương pháp tiềm năng để đạt được độ trễ thấp, nhưng cần có sự điều chế và mã hóa phù hợp cho việc truyền gói tin nhỏ để đảm bảo độ tin cậy chấp nhận được. Các nghiên cứu cho rằng chủ yếu có 3 loại mã hóa được đề xuất cho 5G. Theo đó mã LDPC (Low-Density Parity-Check) hoặc kiểm tra tính chẵn mật độ thấp và mã Polar có hiệu suất tốt hơn so với mã Turbo trong trường hợp các gói tin nhỏ, nhưng đối với các gói tin lớn hoặc trung bình thì điều ngược lại là đúng. Đối với các ứng dụng độ trễ thấp thì việc sử dụng gói tin nhỏ là một yêu cầu, các khía cạnh khác như hiệu năng thực thi trong thử nghiệm thực tế, độ phức tạp, tính linh hoạt cần được nghiên cứu thêm. So sánh ưu, nhược điểm giữa các mã đưa ra như trong Bảng 2-2 dưới đây [16].

Bảng 2-2: So sánh ưu nhược điểm của các mã hóa kênh

	Mã LDPC	Mã Polar	Mã Turbo
Ưu điểm	LDPC được biết	Polar Codes được	Turbo Codes cung cấp

	đến với hiệu suất gần với giới hạn Shannon, đặc biệt ở tỷ lệ bit/sai sót thấp (BER). Mã này cũng hiệu quả trong việc giảm thiểu độ trễ và tối ưu hóa cho truyền dẫn dữ liệu tốc độ cao	chọn làm kỹ thuật mã hóa kênh tiêu chuẩn cho kênh kiểm soát trong 5G NR. Chúng cung cấp hiệu suất mã hóa tốt ở mức độ trễ thấp và được chứng minh là hiệu quả cho việc truyền dữ liệu nhỏ và yêu cầu độ tin cậy cao.	hiệu suất mã hóa rất tốt, đặc biệt ở tỷ lệ BER cao. Chúng đã được sử dụng rộng rãi trong các hệ thống 3G và 4G và được biết đến với khả năng phục hồi cao trong điều kiện kênh nhiễu.
Nhược điểm	LDPC có thể yêu cầu phần cứng phức tạp và tốn kém hơn để thực hiện, đặc biệt là trong các hệ thống yêu cầu mã hóa và giải mã nhanh chóng.	Trong một số trường hợp, Polar Codes có thể không hiệu quả bằng LDPC hoặc Turbo Codes ở tỷ lệ bit/sai sót cao (BER).	Turbo Codes thường yêu cầu nhiều tài nguyên tính toán hơn và có độ trễ giải mã cao hơn so với LDPC và Polar Codes.

- LDPC và Polar Codes thường được ưu tiên cho 5G do hiệu năng mã hóa xuất sắc ở mức độ trễ thấp và hiệu quả tốt với các gói tin nhỏ, đáp ứng nhu cầu của các ứng dụng URLLC.
- Turbo Codes có ý nghĩa đối với các ứng dụng yêu cầu độ tin cậy cao (tỷ lệ BER thấp), nhưng chúng có thể không phải là lựa chọn tốt nhất cho các mạng 5G với yêu cầu độ trễ thấp và hiệu quả truyền dẫn dữ liệu tốc độ cao.
- Sự lựa chọn giữa các kỹ thuật này phụ thuộc vào yêu cầu cụ thể của hệ thống và sự cân nhắc giữa hiệu năng thực thi, độ phức tạp và chi phí.

2.7.1.3 Điều chỉnh bộ phát

Kiến trúc hệ thống dựa trên công nghệ sóng milimét cho việc chuyển mạch đã được giới thiệu trong những năm gần đây, trong đó tín hiệu điều khiển sử dụng beamforming số có độ phân giải thấp (nhằm mục đích ghép kênh các gói tin điều khiển nhỏ) cùng với beamforming tương tự trong mặt phẳng dữ liệu (để kích hoạt điều chế bậc cao hơn). Điều này giảm đáng kể dữ liệu mào đầu chứa dữ liệu điều khiển, dẫn đến việc giải phóng thêm tài nguyên cho việc truyền dữ liệu. Các tiến bộ gần đây trong giao tiếp song công toàn phần (FD – Full Duplex) mang lại lợi ích bằng cách nhân đôi dung lượng, cũng như cơ chế giảm độ trễ và cung cấp phản hồi, đồng thời duy trì mức độ an toàn cho lớp vật lý. Nhiều kỹ thuật mạng 5G được đề xuất, chẳng hạn như beamforming và công nghệ Massive MIMO giúp giảm nhiễu giao thoa, có thể hỗ trợ thực hiện hệ thống FD đáng tin cậy. Ngoài ra, việc lập lịch thông minh cho các gói tin quan trọng về độ trễ và thông lượng cùng với việc điều chỉnh tỷ lệ và phân bổ công suất phù hợp có thể giúp đạt được kết quả tốt trong việc tăng dung lượng và giảm độ trễ. Tuy nhiên các giải pháp giao tiếp song công cải thiện thông lượng, giảm độ trễ và duy trì an ninh lớp vật lý vẫn gặp phải một số vấn đề như nhiễu chéo giữa bộ phát (Tx) và bộ thu (Rx), nhiễu nội bộ, suy giảm và mất đường truyền [17].

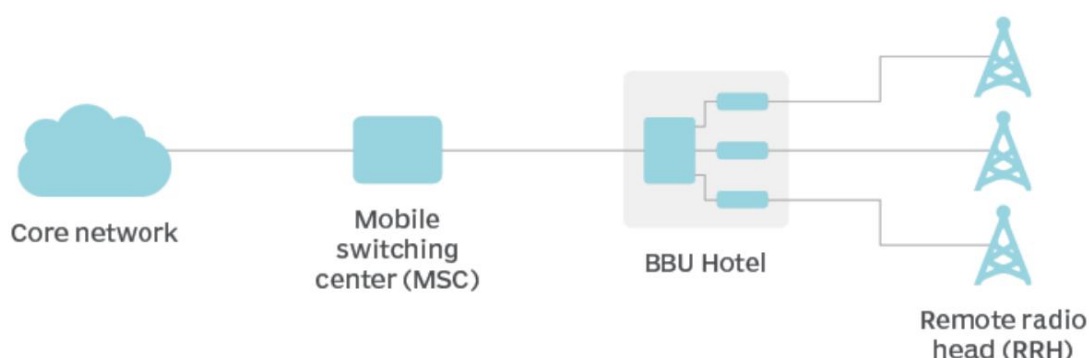
2.7.1.4 Mạng truy nhập vô tuyến đám mây

Mạng Truy cập vô tuyến đám mây (C-RAN) được giới thiệu cho 5G như minh họa trong Hình 2-5. C-RAN kết hợp các đơn vị xử lý băng tần cơ sở của một nhóm trạm gốc vào một máy chủ trung tâm, đồng thời giữ lại phần đầu cuối radio tại các trạm gốc. Tuy nhiên, điều này yêu cầu các liên kết kết nối với độ trễ 250 μ s để hỗ trợ các dịch vụ độ trễ thấp của 5G. Để đáp ứng yêu cầu độ trễ nghiêm ngặt trong C-RAN, hai tối ưu hóa được đề xuất, và kết quả thực nghiệm cho thấy sự hiệu quả của các phương pháp này trong việc tối ưu hóa độ trễ. Việc tách lớp PHY và MAC trong C-RAN với Ethernet front haul được đề xuất. Kết quả đầy hứa hẹn để khẳng định rằng C-RAN có thể hỗ trợ các dịch vụ quan trọng về độ trễ trong 5G. Mặc dù sóng milimét sẽ là nguồn đóng góp chính trong việc đạt được mục tiêu của 5G, phổ tần dưới 6 GHz luôn là lựa chọn chính do ít suy giảm hơn, hỗ trợ khoảng cách xa, và tương thích với ăng-ten. Hơn nữa, các mạng di động truyền thống thường được triển khai trong các băng tần có giấy phép đắt đỏ và sử dụng mạng lõi đáng tin cậy được tối ưu hóa để cung cấp các dịch vụ nhạy

cảm với độ trễ thấp như dịch vụ truyền tiếng nói [12].

Một lược đồ phân bổ phổ tần động và thông minh dựa trên QoE cũng đã được giới thiệu để có thể hỗ trợ phân bổ phổ tần ở cả cấp độ trạm gốc và thiết bị. Kỹ thuật này không chỉ cải thiện việc sử dụng phổ tần, mà còn có thể duy trì QoE mong muốn bao gồm cả khía cạnh độ trễ.

Cloud RAN (C-RAN)



Hình 2-5: Cloud RAN (C-RAN)

2.7.2 Giải pháp mạng lõi

Để đáp ứng tầm nhìn của 5G bao gồm độ trễ siêu thấp cùng với những cải tiến trong mạng truy cập radio, các thay đổi mạnh mẽ cũng được đề xuất trong mạng lõi. Mạng lõi mới bao gồm một số thực thể mới như SDN, MEC và NFV cũng như các kỹ thuật backhaul mới. Những cải tiến này nhằm vào việc giảm thời gian xử lý, bỏ qua một số lớp giao thức để đảm bảo hoạt động liền mạch. Các giải pháp mạng lõi cho độ trễ thấp sẽ được xem xét chi tiết hơn trong phần còn lại của mục này [12].

2.7.2.1 Các thực thể của mạng lõi

SDN (Software-Defined Networking) và NFV (Network Functions Virtualization) được coi là các ứng viên chính trong thiết kế mạng lõi 5G để hỗ trợ giảm độ trễ trong phạm vi mạng này. Các nghiên cứu về thực thể mạng lõi có thể hỗ trợ đạt được độ trễ thấp được tóm tắt trong Bảng 2-3. Cần tách biệt mặt phẳng điều khiển và mặt phẳng dữ liệu vì hai mặt phẳng này cần đáp ứng các tiêu chí QoS khác nhau. Cụ thể, mặt phẳng điều khiển cần độ trễ thấp để xử lý các bản

tin báo hiệu, trong khi mặt phẳng dữ liệu yêu cầu thông lượng cao để xử lý dữ liệu. Do đó, để thiết kế các mặt phẳng này một cách hiệu quả thì cần tách biệt cả hai. SDN và NFV có thể được ứng dụng trong kiến trúc EPC để tách biệt mặt phẳng dữ liệu và mặt phẳng điều khiển và duy trì hoạt động liên mạch của các chức năng mạng lõi.

Bảng 2-3: Tổng quan các kỹ thuật trong mạng lõi cho độ trễ thấp

Cấu trúc mạng lõi	Dựa trên SDN	Kiến trúc của 5G được đề xuất dựa trên SDN với tầm nhìn 5G để đáp ứng thông lượng lớn, kết nối rộng và độ trễ thấp.
	Dựa trên NFV	NFV loại bỏ sự phụ thuộc vào nền tảng phần cứng và làm cho việc triển khai các chức năng PEC trở nên dễ dàng cũng như chia sẻ tài nguyên trong RAN, điều này có thể giảm độ trễ E2E với hiệu suất thông lượng được cải thiện.
	Dựa trên MEC/FOG	MEC/fog cung cấp tính toán và lưu trữ gần người dùng và cũng tách biệt mặt phẳng dữ liệu khỏi mặt phẳng điều khiển, những điều này giúp giảm độ trễ.

Bằng cách sử dụng SDN trong EPC, mặt phẳng điều khiển và mặt phẳng người dùng có thể được tách biệt. Bộ điều khiển SDN có thể hoạt động như một giao diện giữa các mặt phẳng đã tách biệt. Ngoài những lợi ích của việc tách mặt phẳng người dùng và điều khiển dựa trên SDN/NFV như khả năng mở rộng độc lập, linh hoạt trong phân phối luồng và quản lý di động người dùng tốt hơn; việc tách biệt này cũng có thể ảnh hưởng đáng kể đến việc giảm độ trễ. Tách biệt các mặt phẳng nói trên tạo điều kiện cho ứng dụng công nghệ điện toán biên di động (MEC) để giảm độ trễ. Tuy nhiên, việc thêm một bộ điều khiển SDN vào mạng có thể bổ sung độ trễ trong toàn hệ thống.

Quản lý di động trong mạng lõi dựa trên SDN có thể tiềm ẩn một số yếu tố gây ra độ trễ, chẳng hạn như độ trễ xử lý trong hệ thống quản lý di động dựa trên SDN. Thực hiện các giải pháp chủ động và đáp ứng tương thích trong quản lý di động sử dụng Mininet và Open Flow cho tổng độ trễ xử lý ở khoảng giá trị trung

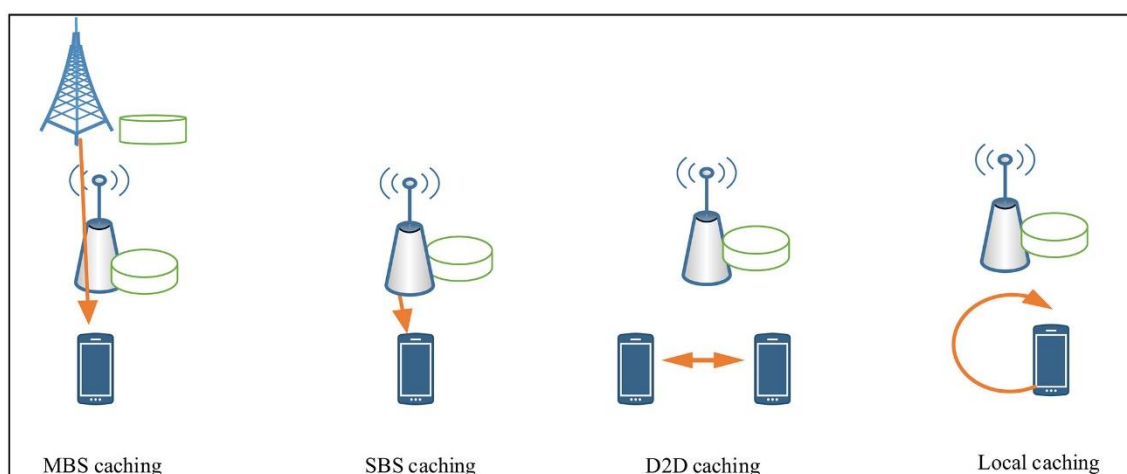
bình với độ tin cậy 95%.

2.7.2.2 Backhaul

Backhaul giữa các trạm cơ sở và mạng lõi chịu trách nhiệm truyền tải thông báo và dữ liệu từ lõi và Internet. Số lượng lớn các trạm gốc Small Cell và các Macro Cell có thể tạo ra dung lượng tới mạng lõi gấp hàng nghìn lần dung lượng của mỗi trạm, cùng với đó là số lượng kết nối các dịch vụ đòi hỏi độ trễ trong phạm vi mạng 5G tạo ra áp lực đối với năng lực truyền tải backhaul để đạt được độ trễ thấp. Trong các kịch bản triển khai, các liên kết vô tuyến, cáp đồng và sợi quang sử dụng cho kết nối backhaul cần được thiết kế, tính toán để đáp ứng các yêu cầu truyền tải. Backhaul của 5G đòi hỏi đáp ứng các yêu cầu dung lượng cao hơn, độ trễ thấp hơn, giải quyết các vấn đề đồng bộ, bảo mật và dự phòng.

2.7.3 Giải pháp Caching

Ngoài vấn đề thiếu hụt phổ tần vô tuyến, khả năng không đủ của các liên kết backhaul cũng có thể được coi là một trở ngại lớn cho việc truyền thông độ trễ thấp. Độ trễ lớn có thể xuất phát từ việc quá nhiều người dùng yêu cầu truy cập trong giờ cao điểm. Do đó, việc giảm độ trễ là yếu tố quan trọng cho chất lượng dịch vụ (QoS) và trải nghiệm người dùng (QoE) trong mạng 5G. Caching có thể được coi là một trong những công nghệ tiềm năng để thiết kế một sự thay đổi cách mạng cho việc giảm độ trễ trong hệ thống truyền thông thế hệ tiếp theo.



Hình 2-6: Các loại caching

Mục tiêu chính của giải pháp này là giảm độ trễ trong việc truy cập dữ liệu,

tăng tốc độ truy cập nội dung và giảm tải trên mạng lõi bằng cách giảm số lượng truy vấn dữ liệu cần phải đi qua mạng. Dưới đây là chi tiết về các loại caching gồm MBS caching, SBS caching, D2D caching và Local caching [12].

1. MBS caching

- Định nghĩa: Lưu trữ dữ liệu tại các trạm gốc lớn (MBS – Macro Base Stations), những trạm này cung cấp vùng phủ rộng và thường xử lý lượng truy cập lớn.
- Lợi ích: Cải thiện hiệu quả truy cập dữ liệu trong phạm vi rộng và giảm độ trễ cho người dùng ở các khu vực phủ sóng của MBS.
- Ứng dụng: Hiệu quả cho việc phân phối nội dung đại chúng hoặc dữ liệu thường xuyên được yêu cầu trên một MSB.

2. SBS Caching

- Định Nghĩa: Lưu trữ dữ liệu tại các trạm gốc nhỏ (SBS - Small Base Stations), những trạm này thường được triển khai trong các khu vực có mật độ người dùng cao hoặc khu vực yêu cầu băng thông lớn.
- Lợi ích: Giảm độ trễ bằng cách lưu trữ nội dung gần người dùng hơn và giảm tải cho mạng lõi và liên kết backhaul.
- Ứng dụng: Phù hợp với các khu vực đô thị, trung tâm thương mại, hoặc sự kiện lớn, nơi có nhu cầu truy cập dữ liệu cao.

3. D2D Caching

- Định Nghĩa: Cho phép trao đổi dữ liệu trực tiếp giữa các thiết bị người dùng mà không cần thông qua mạng lõi, sử dụng công nghệ D2D (Device-to-Device).
- Lợi ích: Tăng cường hiệu suất mạng bằng cách giảm độ trễ và tải trên mạng lõi, đồng thời tận dụng tài nguyên vô tuyến và băng thông hiệu quả.
- Ứng dụng: Hữu ích trong các kịch bản như phát video trực tiếp và chia sẻ nội dung giữa người dùng trong cùng một khu vực địa lý.

4. Local Caching (Caching cục bộ)

- Định nghĩa: Lưu trữ dữ liệu trực tiếp tại thiết bị người dùng hoặc tại các

nút mạng gần người dùng nhất, ví dụ như trạm gốc hoặc router.

- Lợi ích: Giảm độ trễ trong việc truy cập dữ liệu do không cần truy vấn dữ liệu từ xa, và giảm tải truy cập lên mạng lõi.
- Ứng dụng: Phù hợp với việc lưu trữ nội dung địa phương phổ biến hoặc dữ liệu thường xuyên được sử dụng bởi người dùng cụ thể.

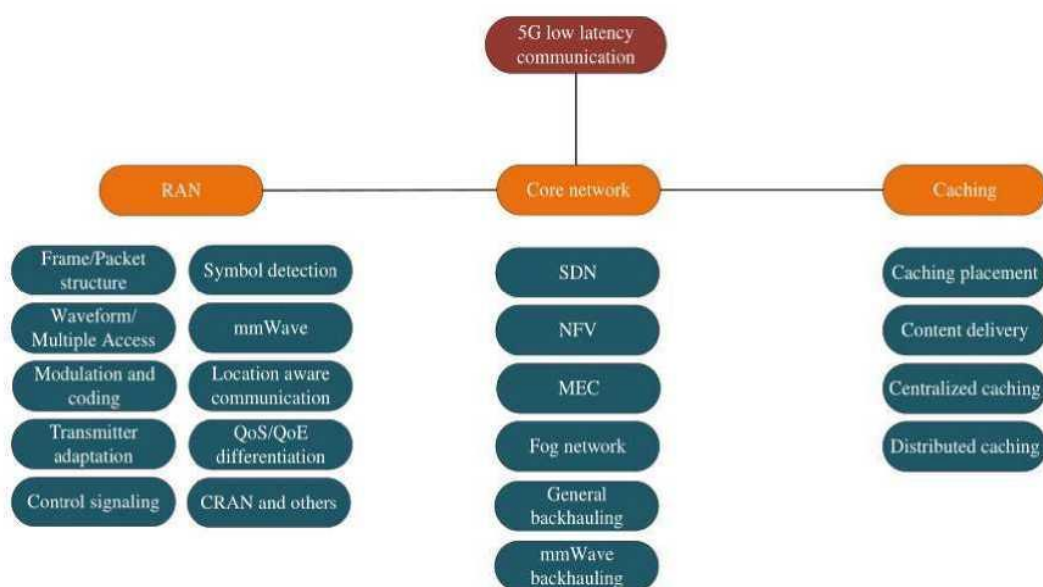
Mỗi loại caching này có những đặc điểm và lợi ích riêng, tùy thuộc vào cấu trúc và yêu cầu của mạng để cải thiện trải nghiệm người dùng và hiệu năng thực thi của mạng bằng cách giảm độ trễ và tối ưu hóa sử dụng tài nguyên.

2.8 Kết luận chương 2

Mạng 5G ngoài việc cung cấp băng thông cực rộng, các yêu cầu về độ trễ cực nhỏ của một số dịch vụ sẽ tác động lớn đến các khía cạnh thiết kế, đáp ứng thực thi của hệ thống trong một phạm vi phổ sóng rộng. Độ trễ thấp sẽ cho phép triển khai trên nó các dịch vụ dựa trên những giải pháp công nghệ mới nổi như VR/AR, chăm sóc y tế và phẫu thuật từ xa; trong một số tình huống nhất định, độ trễ không quá 1 ms là yêu cầu cần được đáp ứng. Nhiều cải tiến trong các bối cảnh mạng khác nhau cần được giải quyết để duy trì độ trễ thấp này. Nội dung nghiên cứu của chương này giới thiệu, mô tả, phân tích các tham số, chỉ tiêu, yêu cầu cũng như các thành phần tác động đến trễ trong mạng 5G. Để có thể đạt được độ trễ thấp đối với mạng 5G, nội dung nghiên cứu trình bày các giải pháp khác nhau phù hợp với các chiến lược triển khai trong các phạm vi như RAN, mạng lõi và bộ nhớ đệm để đạt được độ trễ thấp. Các giải pháp điển hình được trình bày là kiến trúc phân tách mặt phẳng trễ, hệ thống điều chế và mã hóa, phương pháp tiếp cận thực thi kênh điều khiển, mã hóa kênh, RAN đám mây và các khía cạnh thực thi kỹ thuật khác để tăng cường hiệu năng thực thi của mạng cho mục tiêu đạt được độ trễ truyền tải thấp. Hình 2-7 tổng kết các kỹ thuật, giải pháp hướng tới mục tiêu giảm trễ truyền tải, tương ứng với các phạm vi thực thi trong mạng 5G.

Lớp vật lý trong mạng 5G có tầm quan trọng đặc biệt nhưng các công nghệ thực thi ở lớp này hiện đang gặp khó khăn trong việc kết hợp URLLC vào SG NR để đáp ứng các tiêu chí nghiêm ngặt về duy trì độ trễ thấp. Theo đó một số khía cạnh thực thi ở lớp vật lý cần được nghiên cứu cải tiến, chẳng hạn như kỹ

thuật điều khiển tạo chùm tia tương thích với thành phần dữ liệu hoặc kỹ thuật tái cấu hình từng phần dựa trên giao thức URLLC. Ngoài ra, cần nghiên cứu kiến trúc máy thu phát áp dụng các kỹ thuật tiên tiến để hỗ trợ thích ứng các loại hình dữ liệu cũng như kỹ thuật giải mã song song.



Hình 2-7: Các giải pháp khác nhau để đạt được độ trễ thấp trong 5G

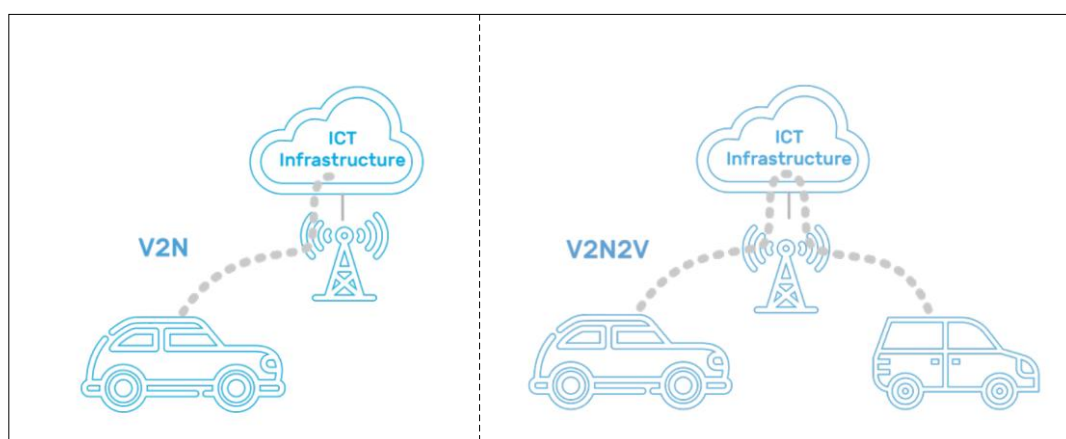
Các mô hình mạng SDN, NFV kết hợp với đường truyền backhaul tốc độ cao cũng đã được giới thiệu trong mạng lõi với mục đích đáp ứng các tiêu chí độ trễ thấp của 5G. Các kỹ thuật tiên tiến áp dụng cho mạng lõi thế hệ mới có thể mang lại nhiều lợi ích, chẳng hạn như thiết kế mạng phân tán, thiết bị phần mềm độc lập với khuôn khổ phần cứng, giải pháp cách ly mặt phẳng dữ liệu khỏi mặt phẳng thiết bị đều đều giúp giảm độ trễ. Tuy nhiên, những vấn đề về công nghệ, kỹ thuật, giải pháp được nêu hầu hết có tính mở, cần phải có những nghiên cứu bổ sung, thử nghiệm với những kịch bản triển khai cụ thể để hoàn thiện chúng.

CHƯƠNG 3: ĐÁNH GIÁ ĐỘ TRỄ DỊCH VỤ V2X TRONG MẠNG 5G

3.1 Tổng quan về V2X

Mối tương quan giữa công nghệ 5G và dịch vụ V2V (Vehicle-to-Vehicle), cũng như ứng dụng rộng lớn hơn là V2X (Vehicle-to-Everything), tạo điều kiện cho một kỷ nguyên mới trong ngành giao thông và vận tải, nơi độ trễ thấp đóng vai trò quan trọng trong việc đảm bảo an toàn và hiệu quả. Với sự ra đời của V2X Application Server (V2X AS), khả năng của 5G trong việc hỗ trợ các dịch vụ liên lạc giữa xe cộ và mọi thứ xung quanh nó trở nên rõ ràng và thiết thực hơn bao giờ hết.

Công nghệ V2X cho phép các phương tiện tham gia giao thông trao đổi thông tin thời gian thực với xe gần đó (V2V), thiết bị ven đường (V2I) và người đi bộ (V2P) thông qua giao tiếp không dây. V2X có tiềm năng to lớn trong việc nâng cao an toàn đường bộ và đạt được mức độ tự động hóa lái xe cao hơn. Trong môi trường V2V và V2X, độ trễ thấp không chỉ cải thiện khả năng giao tiếp và tương tác giữa các phương tiện mà còn là yếu tố quyết định trong việc ngăn ngừa tai nạn và cải thiện luồng giao thông. 5G, với đặc điểm độ trễ thấp, băng thông rộng, và khả năng kết nối đáng tin cậy, mở ra tiềm năng to lớn triển khai các dịch vụ V2V và V2X, cho phép giao tiếp gần như tức thời giữa các phương tiện và với cơ sở hạ tầng xung quanh.

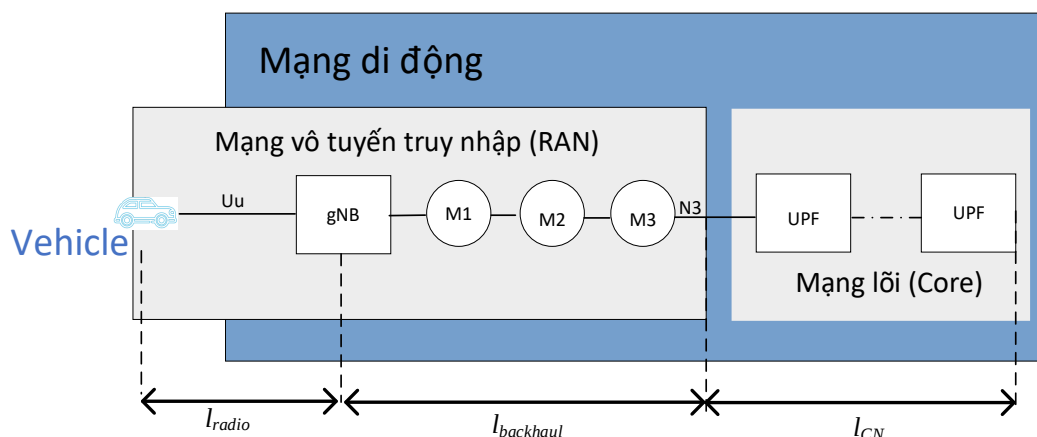


Hình 3-1: V2N và V2N2V

Tuy nhiên, để đạt được mục tiêu độ trễ thấp trong môi trường thực, việc triển khai và cấu hình mạng 5G cần phải được tối ưu hóa. Điều này có thể thực hiện bằng tối ưu hóa các yếu tố ảnh hưởng đến độ trễ như tăng cường khả năng

cân bằng tải, tính toán các kịch bản có thể xảy ra và tài nguyên cần để đáp ứng yêu cầu về độ trễ, độ tin cậy cũng như độ khả dụng của hệ thống. Hình 3-1 thể hiện mô hình triển khai dịch vụ này trong mạng 5G.

3.2 Phân tích các yếu tố ảnh hưởng đến độ trễ



Hình 3-2: Các thành phần độ trễ từ phương tiện đến mạng lõi.

Về mặt kỹ thuật như mô tả trong Hình 3-2, một chiếc xe (vehicle) cũng tương tự như các thiết bị UE khác sử dụng công nghệ vô tuyến. Độ trễ của một vehicle đến mạng lõi được tính như sau:

$$Latency = l_{radio} + l_{backhaul} + l_{corenetwork}$$

Với mỗi thành phần sẽ bị ảnh hưởng bởi các yếu tố khác nhau:

Độ trễ vô tuyến (l_{radio}) phụ thuộc vào các yếu tố như chất lượng tín hiệu, khoảng cách giữa vehicle và trạm phát sóng, cũng như sự nhiễu sóng trong môi trường. Đối với công nghệ 5G, các trạm phát sóng được thiết kế để cung cấp độ trễ thấp hơn so với các công nghệ trước đây, nhờ vào tần số cao và khả năng kết nối đa điểm. Tuy nhiên, trong thực tế, các yếu tố môi trường như địa hình và vật cản có thể gây ra sự biến đổi đáng kể trong độ trễ vô tuyến.

Độ trễ trong mạng truyền dẫn (backhaul – bh) (l_{bh}) là thời gian cần thiết để truyền tải dữ liệu từ trạm phát sóng đến mạng lõi. Độ trễ này bị ảnh hưởng bởi độ dài và chất lượng của tuyến truyền dẫn backhaul. Trong các khu vực đô thị, nơi có cơ sở hạ tầng mạng phát triển, độ trễ backhaul thường thấp hơn so với các khu vực nông thôn. Tối ưu hóa các tuyến truyền dẫn và sử dụng các công nghệ như mạng cáp quang hoặc kết nối microwave có thể giúp giảm thiểu độ trễ backhaul.

Độ trễ trong mạng lõi (Core Network – CN) (l_{CN}) liên quan đến thời gian xử lý dữ liệu trong mạng lõi của nhà mạng. Yếu tố này phụ thuộc vào cấu trúc và khả năng xử lý của mạng lõi, bao gồm các máy chủ, bộ định tuyến và thiết bị chuyển mạch. Mạng lõi 5G được thiết kế với kiến trúc phân tán và sử dụng các công nghệ như phân tách mạng theo lát cắt (network slicing) để tối ưu hóa quá trình xử lý dữ liệu và giảm độ trễ.

3.3 Tính toán độ trễ tại các phân đoạn

Trong phần này sẽ xem xét công thức tính toán theo từng phân đoạn. Số mũ X được sử dụng trong phần này để chỉ Đường lên (UL) hoặc Đường xuống (DL). Trong thực tế độ trễ UL và DL là khác nhau do nhiều yếu tố như băng thông không đối xứng, quá trình xử lý, cách thức lập lịch (scheduling), v.v... Giả định trong nghiên cứu này là thời gian cho UL và DL bằng nhau.

3.3.1 Độ trễ phạm vi vô tuyến

Độ trễ vô tuyến được xem xét ở đây, như đã đề cập ở chương 2, bao gồm hai loại chính: Độ trễ mặt phẳng người dùng (User plane latency) và Độ trễ mặt phẳng điều khiển (Control plane latency).

Độ trễ mặt phẳng người dùng là thời gian cần thiết để truyền tải dữ liệu từ thiết bị đầu cuối đến mạng và ngược lại. Đây là một yếu tố then chốt đối với các ứng dụng yêu cầu độ trễ thấp như thực tế ảo (VR), thực tế tăng cường (AR), và đặc biệt là các ứng dụng V2X (Vehicle-to-Everything).

Độ trễ mặt phẳng điều khiển liên quan đến thời gian cần thiết để thiết lập, duy trì và kết thúc các kết nối mạng, cũng như truyền tải các tín hiệu điều khiển. Điều này bao gồm các thủ tục như thiết lập cuộc gọi, xác thực người dùng và quản lý di động.

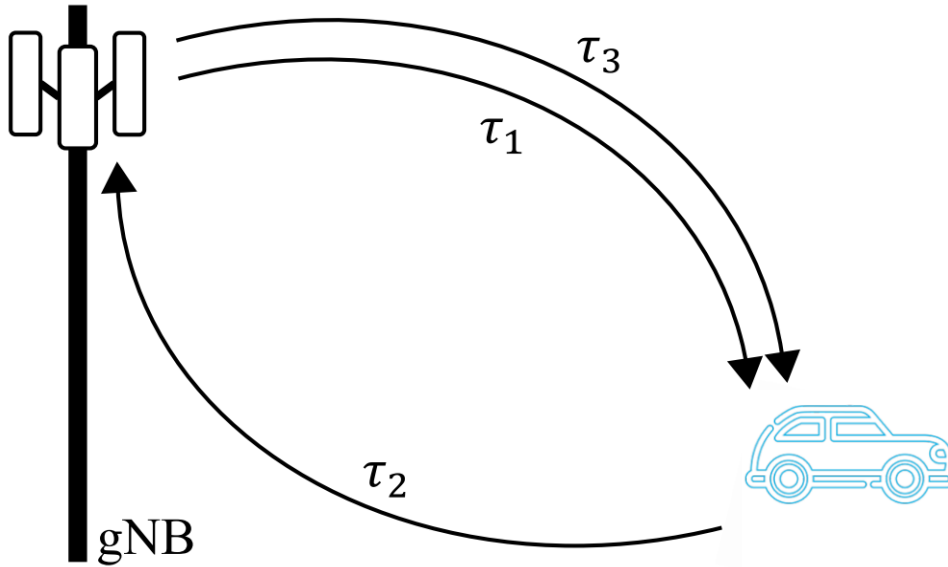
3.3.1.1 Độ trễ mặt phẳng người dùng

Độ trễ mặt phẳng người dùng (UPL) bao gồm quá trình truyền (τ_1), yêu cầu HARQ (τ_2) và truyền lại (τ_3) giữa hai thực thể, như được hiển thị trong Hình 2. Các yêu cầu cho chỉ số hiệu suất này là 4 ms đối với dịch vụ eMBB và 1 ms đối với URLLC.

Thời gian truyền có thể được tính theo công thức sau: [18]

$$l_{UPL} = \tau_1 + p(\tau_2 + \tau_3) \quad (3-1)$$

Với p là tần suất truyền lại.



Hình 3-3. Mặt phẳng người dùng

Với lần truyền đầu tiên độ trễ là:

$$\tau_1 = t_{gNB,tx} + t_{FA1} + t_{TTI} + t_{UE,rx} \quad (3-2)$$

trong đó $t_{gNB,tx}$ là thời gian xử lý trong gNB, t_{FA1} là thời gian cần thiết để căn chỉnh khung, t_{TTI} là thời gian truyền dữ liệu và $t_{UE,rx}$ là thời gian xử lý trong Vehicle.

Thời gian xử lý trong gNB $t_{gNB,tx}$ được tính toán như sau:

$$t_{gNB,tx} = \frac{\max\left[N_2(2048+144).k.2^{-\mu}.\frac{T_c}{2048.\Delta_f}, 0\right]}{2} \quad (3-3)$$

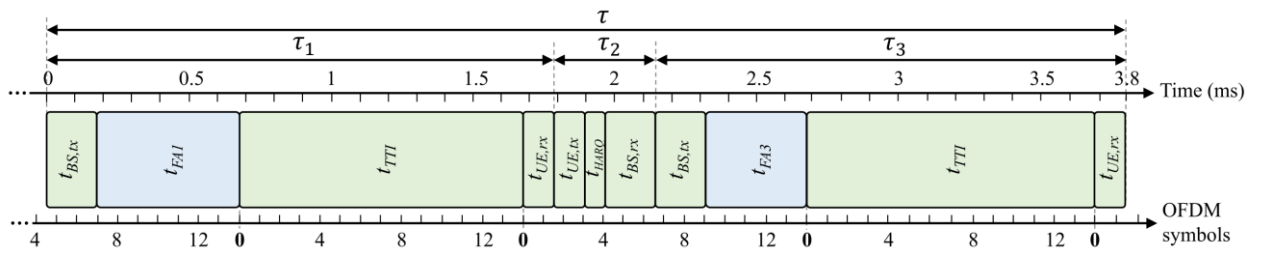
trong đó:

- N_2 là số lượng chu kỳ thời gian xử lý (processing cycles) cần thiết trong gNodeB để thực hiện các tác vụ cụ thể, tham số này phụ thuộc vào SCS [19].

Các giá trị của N_2 mẫu dựa trên SCS với $N=13$ cho $\mu=0$, $N=13$ cho $\mu=1$, $N=20$ cho $\mu=2$, and $N=24$ cho $\mu=3$

- T_c là số đơn vị thời gian có giá trị $T_c = \frac{1}{\Delta f_{max} \cdot N_f}$.
- Hằng số $k = \frac{T_s}{T_c} = 64$ với $T_s = 1/(\Delta f_{ref} \cdot N_{f,ref})$, $\Delta f_{ref} = 15 \cdot 10^3 \text{ Hz}$ và $N_{(f,ref)} = 2048$. [20]
- Δf_{max} bằng $480 \cdot 10^3$ và $N_f = 4096$ đối với mạng 5G [20].
- Giá trị của μ và Δf được thể hiện trong [20].

Lưu ý rằng độ trễ trong gNB là như nhau khi truyền và nhận. Điều này cũng áp dụng cho Vehicle.



Hình 3-4. Độ trễ mặt phẳng người dùng

Thời gian xử lý được tính theo công thức:

$$t_{UE,rx} = \frac{\max\left[N_1(2048+144) \cdot k \cdot 2^{-\mu} \cdot \frac{T_c}{2048 \cdot \Delta f}, 0\right]}{2} \quad (3-4)$$

Trong đó N_1 là tham số phụ thuộc vào SCS. [19]

Nếu không cần truyền lại thì: $t_{UPL} = \tau_1$.

Nếu thiết bị cần phải truyền lại qua HARQ request, vậy phải cần thêm thời gian:

$$\tau_2 = t_{UE,tx} + t_{FA2} + t_{HARQ} + t_{gNB,rx} \quad (3-5)$$

trong đó:

- t_{FA2} là thời gian điều chỉnh khung
- t_{HARQ} là thời gian của 1 yêu cầu (1 OFDM symbol).

Sau khi nhận và xử lý HARQ request, gNB sẽ truyền lại với thời gian:

$$\tau_3 = t_{gNB,tx} + t_{FA3} + t_{TTI} + t_{UE,rx} \quad (3-6)$$

3.3.1.2 Độ trễ mặt phẳng điều khiển

Độ trễ mặt phẳng điều khiển (CPL) có thể được biểu diễn theo công thức sau: [18] [19]

$$l_{CPL} = \sum_{i=1}^{10} T_i + \sum_{j=1}^3 T_{FAj} \quad (3-7)$$

Quy trình RACH như đã trình bày trong Hình 2-4 của chương 2, quá trình tính toán độ trễ CP được chia thành 10 bước liên quan đến lập lịch, xử lý và các khía cạnh truyền dẫn, 3 bước bổ sung liên quan đến việc căn chỉnh khung. Quá trình truyền chỉ có thể bắt đầu trong các ký hiệu OFDM nơi mà tiền tố PRACH được sử dụng.

3.3.2 Độ trễ trong mạng truyền dẫn

Độ trễ của mạng truyền dẫn (backhaul - BH) l_{BH} là tổng của l_{BH}^{UL} và l_{BH}^{DL} . l_{BH}^{UL} là thời gian các gói tin được xử lý tại gNB và thời gian truyền các gói tin đó qua backhaul về phía Core. l_{BH}^{DL} là thời gian truyền các gói tin qua backhaul tới gNB và quá trình xử lý chúng tại gNB. l_{BH}^X có thể được biểu diễn dưới dạng: [21]

$$l_{BH}^X = t_{BH,prop}^X + t_{BH,transit}^X \quad (3-8)$$

Trong đó:

- $t_{BH,prop}^X$ là độ trễ lan truyền (propaganda) hoặc thời gian để gói tin di chuyển qua các liên kết kết nối các node của BH.
- $t_{BH,transit}^X$ là độ trễ chuyển tiếp và biểu thị thời gian mà các gói tin dành cho các nút trong BH. Các gói tin trong trường hợp này được coi là không yêu cầu truyền lại vì đây là trường hợp của hầu hết các liên kết mạng trong thực tế [22] [23].
- $t_{BH,transit}^X$ tính đến độ trễ liên quan đến quá trình xử lý truyền và nhận của các nút BH, bao gồm:
 - $t_{BH,p}$ (processing) là khoảng thời gian từ lúc nhận được gói tin đến lúc đưa vào hàng đợi để truyền đi.
 - $t_{BH,q}$ (queuing) là thời gian gói tin phải chờ trong hàng đợi để được truyền đi.
 - $t_{BH,tt}$ (transmission time) là thời gian cần thiết để truyền gói tin từ nguồn đến đích.

$t_{BH,prop}^X$ phụ thuộc vào tổng khoảng cách d_{BH} mà các gói tin truyền qua BH và tốc độ lan truyền v trong môi trường BH, được tính bằng công thức:

$$t_{BH,prop}^X = \frac{d_{BH}}{v} \quad (3-9)$$

Để ước tính $t_{BH,transit}^X$, nghiên cứu ở đây sử dụng mô hình hóa backhaul như một hệ thống xếp hàng, sử dụng mô hình hàng đợi M/M/1 [24]. Mỗi node trong mạng Backhaul được biểu diễn dưới dạng một máy chủ duy nhất xử lý các gói tin đến theo quy trình Poisson với tốc độ tới trung bình λ . Nút phục vụ hoặc gửi các gói tin nhận được ở tốc độ dịch vụ trung bình μ tuân theo phân phối chuẩn. Việc sử dụng một nút được tính bằng $\rho = \frac{\lambda}{\mu}$. Hàng đợi M/M/1 xem xét với dung lượng bộ đệm vô hạn. Các nút trong BH sẽ không ổn định nếu $\rho > 1$ (tức là $\lambda > \mu$). Điều này xảy ra vì lượng tồn đọng bộ đệm sẽ tăng lên vô tận và sẽ có các gói tin không bao giờ rời khỏi nút trong một khoảng thời gian hữu hạn. Sử dụng λ và μ , độ trễ chuyển tiếp trung bình khi các gói tin đi qua tập hợp n nút trong BH có thể được biểu thị theo định lý Jackson dưới dạng [23]:

$$t_{BH,transit_n}^X = n(t_{BH,p} + t_{BH,q} + t_{BH,tt}) = n \cdot t_{BH,p} + \sum_i \frac{1}{\mu_i^X - \lambda_i^X} \quad (3-10)$$

Trong đó λ_i^X và μ_i^X là tốc độ đến của lưu lượng truy cập và tốc độ dịch vụ của nút i , $i=1\dots,n$. Theo [10], cần tính λ_i^X và μ_i^X cho mỗi nút i để ước tính $t_{BH,transit_n}^X$. Để đạt được điều đó, ta có thể sử dụng các thuộc tính của quá trình Poisson. Nếu chúng ta xem xét kiến trúc phân cấp cho BH, có thể xác định rằng node i tổng hợp lưu lượng truy cập từ một tập hợp g nút BH trong chiều UL. Trong trường hợp này, λ_i^{UL} có thể được tính là $\sum_{k=1}^g \lambda_k^{UL}$. Ở chiều DL, nút i ở trên sẽ phân chia lưu lượng truy cập giữa các nút g . Tốc độ đến của lưu lượng truy cập mà nút i phải gửi qua mỗi liên kết g được xác định bởi quy trình Poisson với tốc độ trung bình $\lambda_i^{DL} = p_k \cdot \lambda_{i,total}^{DL}$. Với p_k là phần tốc độ đến tổng thể của lưu lượng truy cập $\lambda_{i,total}^{DL}$ được gửi đến liên kết đầu ra k , $k=1\dots,g$. Tất cả các gói tin đến nút BH $_i$ phải được gửi đi, vì vậy điều kiện $\sum_{k=1}^g p_k = 1$ phải được thỏa mãn. Tốc độ dịch vụ của nút BH i trong UL được tính như sau:

$$\mu_i^{UL} = \frac{\alpha_{i-(i+1)} \cdot C_{i-(i+1)}}{B} \quad (3-11)$$

Trong đó:

- B là kích thước (tính bằng bit) của các gói tin được truyền qua các liên kết trong BH.
- $C_{i-(i+1)}$ (tính bằng bit/s) biểu thị dung lượng của liên kết kết nối nút i và nút tiếp theo $i+1$ trong đường dẫn UL.
- $\alpha_{i-(i+1)}$ là phần dung lượng liên kết được phân bổ trong chiều UL để hỗ trợ lưu lượng truy cập của một dịch vụ cụ thể. (α : shape parameter - tham số hình dạng hay tham số độ dốc)

Tốc độ dịch vụ μ_{i+1}^{DL} của nút $i+1$ trong chiều DL, được tính như trong [6] nhưng sử dụng $\alpha_{(i+1)-i}$ để hiển thị hướng ngược lại của liên kết. Các giá trị khác nhau của α có thể được phân bổ để hỗ trợ lưu lượng truy cập UL và DL trên cùng một liên kết giữa các nút i và $i+1$, với điều kiện tổng của cả hai giá trị đều nhỏ hơn 1.

3.3.2.1 Xem xét trễ qua gNB

Tại gNB, d_{BH} có thể coi xấp xỉ bằng 0 m, và do đó $t_{BH,prop_{gNB}}^X = 0$ s.

Tại hướng UL, các gói tin nhận được tại gNB được xử lý và truyền tới Core.

Tại hướng DL, các gói tin đến từ Core được xử lý tại gNB và sau đó được truyền qua giao diện vô tuyến, do vậy chỉ tính thời gian xử lý. Khi đó, theo (3-11), độ trễ chuyển tiếp trong UL và DL có thể được tính như sau: [21]

$$t_{BH,transit_{gNB}}^{UL} = t_{BH,p} + \frac{1}{\mu_{gNB}^{UL} - \lambda_{gNB}^{UL}} \quad (3-12)$$

$$t_{BH,transit_{gNB}}^{DL} = t_{BH,p} \quad (3-13)$$

Tốc độ đến của lưu lượng truy cập đến gNB trong UL khi đó là:

$$\lambda_{gNB}^{UL} = N_{UE} \cdot T_p \cdot (1 - P_{rel}) \quad (3-14)$$

Trong đó N_{UE} là số lượng UE hoặc phương tiện nằm trong phạm vi phủ sóng của gNB. Cho rằng tất cả các phương tiện trong phạm vi phủ sóng của gNB đều tạo ra các gói tin bit B với tốc độ trung bình là T_p gói tin/giây. P_{rel} là độ tin cậy trong giao diện vô tuyến UL. P_{rel} có thể gần đúng với BLER khi phương tiện

chỉ thực hiện một lần truyền trên mỗi gói tin hoặc tới $1 - \text{BLER}^{n+1}$ khi phương tiện thực hiện n lần truyền bổ sung với sự lặp lại/truyền lại gói tin gốc. Theo (3-11), tốc độ dịch vụ của gNB được biểu thị bằng $\mu_{gNB}^{UL} = \frac{\alpha_{gNB-UPF} \cdot C_{gNB-UPF}}{B}$.

3.3.2.2 Xem xét độ trễ qua M1

Độ trễ lan truyền khi ra khỏi M1 có thể được tính theo (3-10) dưới dạng: [21]

$$t_{BH,prop_{M1}}^X = \frac{d_{gNB-M1}}{v} \quad (3-15)$$

Trong đó d_{gNB-M1} là khoảng cách giữa gNB và nút M1. Độ trễ chuyển tiếp được tính toán khi xem xét rằng các gói tin được xử lý và truyền trong đường lên UL từ nút gNB tới M1. Sử dụng (3-10), độ trễ UL được tính như sau:

$$t_{BH,transit_{M1}}^{UL} = 2 \cdot t_p + \frac{1}{\mu_{gNB}^{UL} - \lambda_{gNB}^{UL}} + \frac{1}{\mu_{M1}^{UL} - \lambda_{M1}^{UL}} \quad (3-16)$$

Trong đó:

$$\mu_{gNB}^{UL} = \frac{\alpha_{gNB-M1} \cdot C_{gNB-M1}}{B}$$

λ_{gNB}^{UL} được định nghĩa như tính cho công thức (3-14).

Để tính λ_{M1}^{UL} , coi mỗi nút M1 được kết nối với một tập hợp g các gNB. Khi đó λ_{M1}^{UL} bằng $\sum_{i=1}^g \lambda_{gNB,i}^{UL}$. Theo (3-11) thì $\mu_{M1}^{UL} = \frac{\alpha_{M1-M2} \cdot C_{M1-M2}}{B}$.

Tại đường xuống DL, nút M1 xử lý và truyền các gói tin tới gNB. gNB xử lý các gói tin và truyền chúng qua giao diện vô tuyến. Khi đó, độ trễ chuyển tiếp trong DL được biểu thị như sau:

$$t_{BH,transit_{M1}}^{DL} = 2 \cdot t_p + \frac{1}{\mu_{M1}^{DL} - \lambda_{M1}^{DL}} \quad (3-17)$$

Trong công thức (3-17), $\mu_{M1}^{DL} = \frac{\alpha_{M1-gNB} \cdot C_{gNB-M1}}{B}$ và λ_{M1}^{DL} được tính toán khi xem xét rằng M1 phân chia lưu lượng truy cập trong DL được đánh địa chỉ tới g các gNB khác nhau. Do vậy, nó có thể được biểu thị dưới dạng $\lambda_{M1}^{DL} = p_k \cdot \lambda_{M1,total}^{DL}$, trong đó tốc độ lưu lượng truy cập tổng thể $\lambda_{M1,total}^{DL}$ đến M1 bằng với tốc độ lưu lượng khởi hành từ UPF mà nó được kết nối (tức là λ_{UPF}^{DL}).

3.3.2.3 Xem xét độ trễ qua M3

Độ trễ lan truyền khi qua M3 được tính như sau: [21]

$$t_{TN,prop_{M3}}^X = \frac{d_{gNB-M3}}{v} \quad (3-18)$$

Trong đó d_{gNB-M3} là khoảng cách mà các gói tin di chuyển từ gNB đến nút M3. Trong trường hợp này, các gói tin được xử lý và truyền theo đường lên UL từ nút gNB, đi qua M1, M2 tới M3. Khi đó:

$$\overline{t_{TN,transit_{M3}}^{UL}} = 4 \cdot t_p + \sum_{\forall q \in \{gNB, M1, M2, M3\}} \frac{1}{\mu_q^{UL} - \lambda_q^{UL}} \quad (3-19)$$

Việc tổng hợp lưu lượng diễn ra trong UL tại các nút M1, M2 và M3. Coi rằng các nút này lần lượt tổng hợp lưu lượng truy cập từ một tập hợp g các node gNB, nút M1 và nút M2. λ_{gNB}^{UL} và λ_{M1}^{UL} được tính như trong mục 3.3.2.1 và 3.3.2.2.

Tốc độ đến λ của lưu lượng truy cập đến các nút M2 và M3 được tính lần lượt là. $\lambda_{M2}^{UL} = \sum_{i=1}^{m1} \lambda_{M1,i}^{UL}$ và $\lambda_{M3}^{UL} = \sum_{i=1}^{m2} \lambda_{M2,i}^{UL}$.

Tốc độ dịch vụ trung bình tại M2 và M3 lần lượt là $\mu_{M2}^{UL} = \frac{\alpha_{M2-M3} \cdot C_{M2-M3}}{B}$ và $\mu_{M3}^{UL} = \frac{\alpha_{M3-UPF} \cdot C_{M3-UPF}}{B}$.

Trên đường xuống DL, các nút M3, M2 và M1 xử lý và truyền các gói tin tới gNB. Sau đó gNB xử lý chúng trước khi các gói tin được truyền qua giao diện vô tuyến. Khi đó độ trễ chuyển tiếp trong DL có thể được tính như sau:

$$\overline{t_{BH,transit_{M3}}^{DL}} = 4 \cdot t_p + \sum_{\forall q \in \{M1, M2, M3\}} \frac{1}{\mu_q^{DL} - \lambda_q^{DL}} \quad (3-20)$$

Trong trường hợp này, tốc độ đến của lưu lượng truy cập đến các nút M3, M2 và M1 trong đường xuống DL được tính là $\lambda_{M3}^{DL} = p_k \cdot \lambda_{UPF}^{DL}$ với $k=1, \dots$, $\lambda_{M2}^{DL} = p_j \cdot \lambda_{M3}^{DL}$ với $j=1, \dots, m1$, và $\lambda_{M1}^{DL} = p_i \cdot \lambda_{M2}^{DL}$ với $i, i=1, \dots, g$. Ngoài ra, tốc độ dịch vụ trung bình của các nút này lần lượt là $\mu_{M3}^{DL} = \frac{\alpha_{(M3-M2)} \cdot C_{(M2-M3)}}{B}$, $\mu_{M2}^{DL} = \frac{\alpha_{(M2-M1)} \cdot C_{(M1-M2)}}{B}$ và $\mu_{M1}^{DL} = \frac{\alpha_{(M1-gNB)} \cdot C_{(gNB-M1)}}{B}$.

3.3.3 Độ trễ trong mạng lõi

Độ trễ trong mạng lõi l_{CN} được tính bằng độ trễ mà các gói tin gặp phải tại UL (l_{CN}^{UL}) và DL (l_{CN}^{DL}). Tương tự như độ trễ trong mạng truyền tải, l_{CN}^X có thể được biểu thị bằng tổng của thời gian lan truyền ($t_{CN,prop}^X$) và thời gian chuyển tiếp

$(t_{CN,transit}^X)$.

Thời gian $t_{CN,prop}^X$ được tính toán tương tự như trong (3-9) với d_{CN} biểu thị khoảng cách mà các gói tin di chuyển qua các liên kết kết nối các nút của CN và v tốc độ lan truyền qua CN.

Thời gian $t_{CN,transit}^X$ cũng được ước lượng bằng cách sử dụng hệ thống hàng đợi M/D/1 cho các nút của CN [25]. Hàng đợi M/D/1 cũng được đặc trưng bởi λ và μ , nhưng trong trường hợp này λ tuân theo quy trình Poisson và μ được cố định. Mô hình này giải thích tính linh hoạt được tạo ra trong 5G với tính năng ảo hóa chức năng mạng (NFV). Nó coi các nút NFV được đặt làm cổng nối giữa CN và BH và giữa CN và Internet. Các nút NFV này hoạt động như UPF và chúng được kết nối với nhau thông qua các bộ chuyển mạch mạng. Khi các gói tin chỉ đi qua UPF (cục bộ) để đi tiếp, độ trễ chuyển tiếp trung bình có thể được ước tính là [25]:

$$t_{CN,transit}^X = \frac{2}{\mu_{UPF}^X} + \frac{\lambda_{UPF}^X}{2(\mu_{UPF}^X)^2(1-\rho_{UPF}^X)} \quad (3-21)$$

Trong đó số hạng đầu tiên bao gồm độ trễ xử lý (t_p) và độ trễ thời gian truyền (t_{tt}) và số hạng thứ hai là độ trễ xếp hàng (t_q). Khi các gói tin cần đi qua CN để được xử lý trên đám mây, độ trễ chuyển tiếp trung bình được ước tính là:

$$t_{CN,transit}^X = 2 \cdot (t_{CN,transit}^X) + \frac{2}{\mu_{UPF}^X} \cdot S \quad (3-22)$$

Trong (3-18), S là số nút trung gian giữa các UPF đóng vai trò là cổng nối với BH và Internet (tức là PSA) và nó được tính như sau:

$$S = \left\lceil \frac{d_{CN}}{d_{CN,max}} \right\rceil - 1 \quad (3-23)$$

Trong đó $d_{CN,max}$ là khoảng cách tối đa giữa hai nút của CN. Lưu ý rằng số hạng đầu tiên của (3-21) coi các gói tin chỉ được xếp hàng đợi ở hai UPF đóng vai trò là cổng nối. UPF trung gian chỉ thêm độ trễ xử lý và độ trễ thời gian truyền vào độ trễ CN (số hạng thứ hai của (3-21)).

Các biểu thức lý thuyết xếp hàng nổi tiếng có thể được sử dụng để rút ra hàm phân phối độ trễ CN (tức là $\mathcal{L}_{CN}^{(t)}$). Tuy nhiên, việc tính toán các biểu thức của $\mathcal{L}_{CN}^{(t)}$ cho hệ thống xếp hàng M/D/1 là không thể điều khiển được.

Các biểu thức chung về độ trễ CN được hiển thị ở trên và áp dụng cho tất cả các hoạt động triển khai mạng 5G được xác định. Tuy nhiên, việc triển khai khác nhau về giá trị của các biến $t_{CN,prop}^X$ và $t_{CN,transit}^X$ được tính toán tiếp theo cho mỗi lần triển khai.

3.4 Thực nghiệm tính toán độ trễ trong các phạm vi phân đoạn mạng 5G

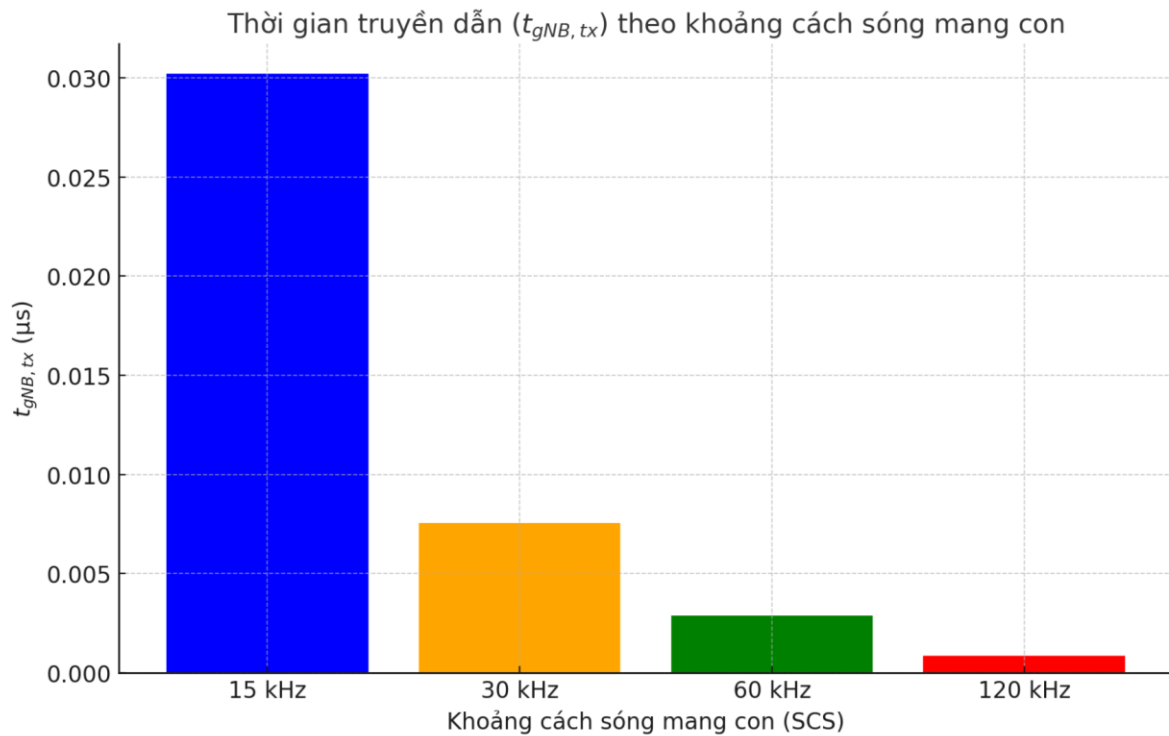
Mục tiêu của nội dung nghiên cứu sẽ thực hiện các thực nghiệm tính toán để khảo sát nhằm đánh giá yếu tố ảnh hưởng đến độ trễ trong các phân đoạn mạng như các mô hình, phân tích đưa ra trong mục 3.3 của chương này. Cụ thể, các thực nghiệm tính toán để khảo sát đánh giá một số yếu tố ảnh hưởng tới trễ trong phân đoạn mạng truyền dẫn với các phần tử gây trễ như thể hiện trong Hình 3-2. Các tính toán trong Chương này sử dụng phần mềm Mathematica Online (đường dẫn <https://www.wolfram.com/mathematica/online/>).

3.4.1 Đánh giá ảnh hưởng của yếu tố SCS

Trễ tại nút gNB được xem tương đương với thời gian xử lý dữ liệu truyền qua nó, nghĩa là tương đương với tính giá trị $gNB_{t(gNB,tx)}$ như thể hiện công thức (3-3). Với các thông số giả định về thiết lập giá trị khoảng cách sóng mang con (SCS) như thể hiện trong Bảng 3-1.

μ	$\Delta f = 2^\mu \cdot 15 [kHz]$	N_2
0	15	13
1	30	13
2	60	20
3	120	24

Bảng 3-1: Giá trị thông số thiết lập giả định cho SCS



Hình 3-5: Kết quả tính toán đánh giá $t_{gNB,tx}$

Đồ thị thể hiện kết quả tính toán trên Hình 3-5 cho thấy sự thay đổi độ trễ thời gian xử lý truyền tải qua gNB tương ứng với sự thay đổi giá trị của Δf với khoảng cách sóng mang con được thiết lập ở 4 giá trị khác nhau.

Nhận xét: Trong việc đánh giá yếu tố ảnh hưởng đến $t_{gNB,tx}$ giá trị của SCS đóng vai trò quan trọng, ảnh hưởng trực tiếp đến thời gian $t_{gNB,tx}$. Trong các ứng dụng V2N (Vehicle-to-Network), khoảng cách sóng mang con (SCS) đóng vai trò quan trọng trong việc đảm bảo thời gian xử lý nhanh tại $gNodeB$. Khi SCS tăng, độ dài ký hiệu OFDM giảm, cho phép xử lý dữ liệu nhanh hơn và giảm độ trễ, điều này rất quan trọng đối với các ứng dụng yêu cầu phản hồi thời gian thực, như cảnh báo va chạm hoặc điều hướng tự động. Tuy nhiên, với SCS nhỏ, độ dài ký hiệu tăng, phù hợp hơn trong môi trường kết nối ổn định ở khoảng cách xa, như giữa các phương tiện và hạ tầng mạng trong khu vực ngoại ô. Để cân đối SCS trong V2N, cần xem xét các yêu cầu về độ trễ, tốc độ di chuyển của phương tiện, và điều kiện môi trường mạng để tối ưu hóa hiệu suất và độ tin cậy của kết nối.

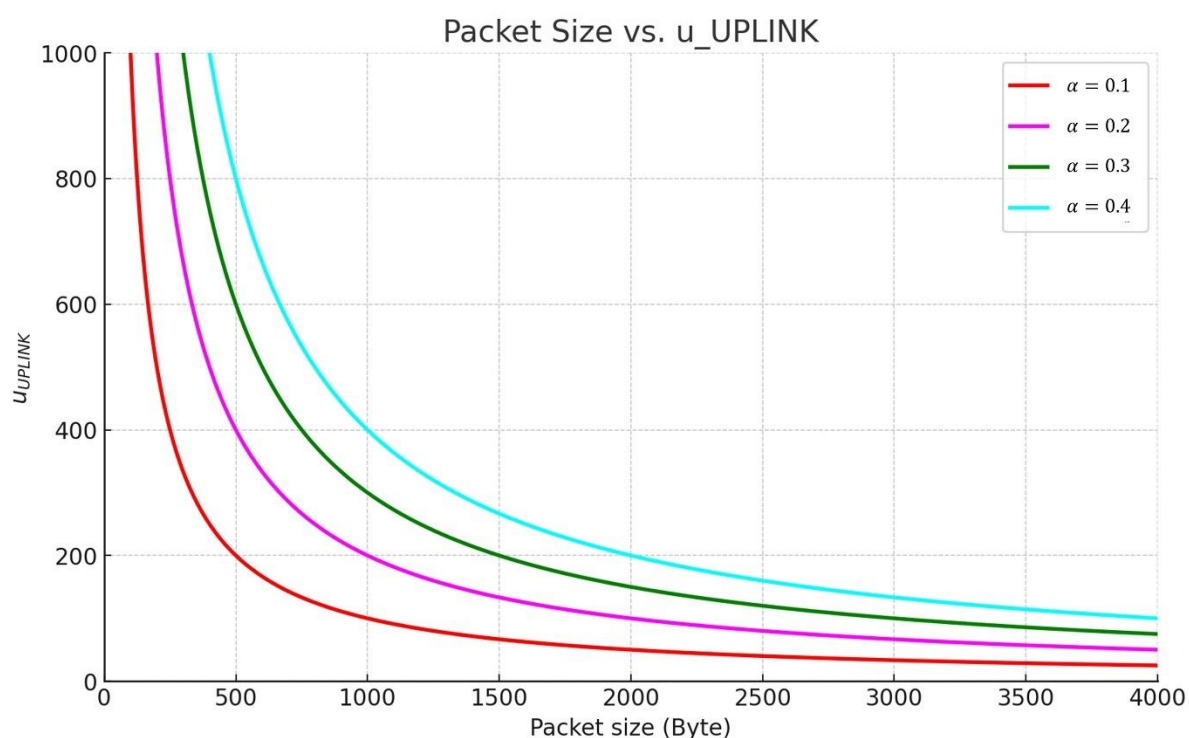
3.4.2 Đánh giá ảnh hưởng của yếu tố Tỷ lệ dung lượng liên kết (α)

Nút mạng backhaul được mô hình hóa như một máy chủ phục vụ theo mô

hình hàng đợi. Tính toán thực nghiệm sẽ thực hiện khảo sát tốc độ phục vụ của nút mạng này với các yếu tố ảnh hưởng tới nó. Cụ thể ở đây sẽ xem xét mức độ sử dụng dung lượng liên kết sẽ ảnh hưởng như thế nào tới tốc độ phục vụ trên đường UL của nút mạng. Mô hình tính toán được giả định với các thông số thiết lập như trong Bảng 3-2. Kết quả thực nghiệm tính toán thể hiện trong hình 3-6.

Bảng 3-2: Thông số thiết lập giả định cho việc tính toán tốc độ đường UL của nút mạng backhaul

Kích thước gói tin (B)	10 – 500 bytes (80-4000 bits)
Dung lượng liên kết (C)	1Gbps
Dung lượng liên kết sử dụng (α)	10% (0.1) 20% (0.2) 30% (0.3) 40% (0.4)



Hình 3-6: Kết quả tính toán đánh giá α

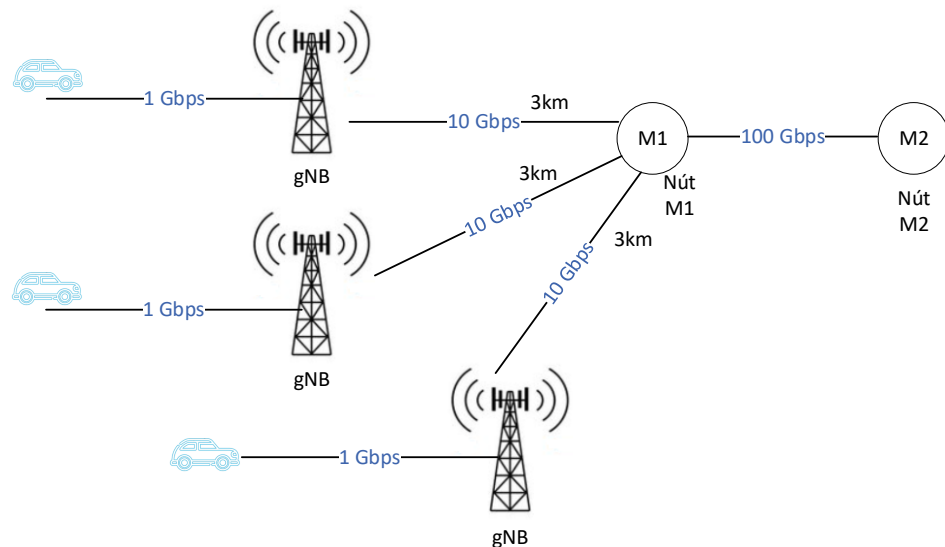
Nhận xét: Tỷ lệ phục vụ (μ) và phần dung lượng liên kết sử dụng (α) có

mối quan hệ tuyến tính, được biểu thị qua công thức (3-11) $\mu = \frac{\alpha \cdot C}{B}$. μ đại diện cho số lượng gói tin có thể được xử lý trong một giây, trong khi α là phần dung lượng liên kết được phân bổ trong chiều UL (uplink) để hỗ trợ lưu lượng truy cập của một dịch vụ cụ thể. Các giá trị α từ 0,1 đến 0,4 thể hiện mức độ ưu tiên dành cho dịch vụ. Khi α tăng, tỉ lệ dung lượng dành cho dịch vụ cũng tăng, dẫn đến μ tăng.

Dung lượng kết nối ($C = 1\text{Gbps}$) là tốc độ truyền dữ liệu tối đa của liên kết. Trong thực tế, giá trị của C là cố định, là một hằng số, do vậy tỉ lệ phục vụ chủ yếu phụ thuộc vào α và kích thước gói tin B . Kích thước gói tin B có giá trị từ 80 đến 4000 bits, phản ánh độ lớn của dữ liệu truyền đi. Khi kích thước gói tin B tăng, số lượng gói tin được xử lý trong một đơn vị thời gian giảm, làm giảm μ . Do đó, để duy trì tỉ lệ phục vụ cao, cần tối ưu hóa giá trị B tương ứng với α . Tóm lại việc điều chỉnh α và B sẽ trực tiếp ảnh hưởng đến tỉ lệ phục vụ, đặc biệt khi C là một hằng số.

Các dịch vụ V2X yêu cầu độ trễ thấp và độ tin cậy cao, do đó thường được ưu tiên tài nguyên mạng. Trong một số nghiên cứu và thử nghiệm thực tế, việc phân bổ tài nguyên (tỷ lệ α) cho V2X có thể thay đổi tùy theo kịch bản sử dụng và cấu hình mạng. Ví dụ, khi tải lưu lượng V2X tăng, có thể cần điều chỉnh α để đảm bảo các yêu cầu về độ trễ và độ tin cậy.

3.4.3 Đánh giá ảnh hưởng của yếu tố mật độ xe



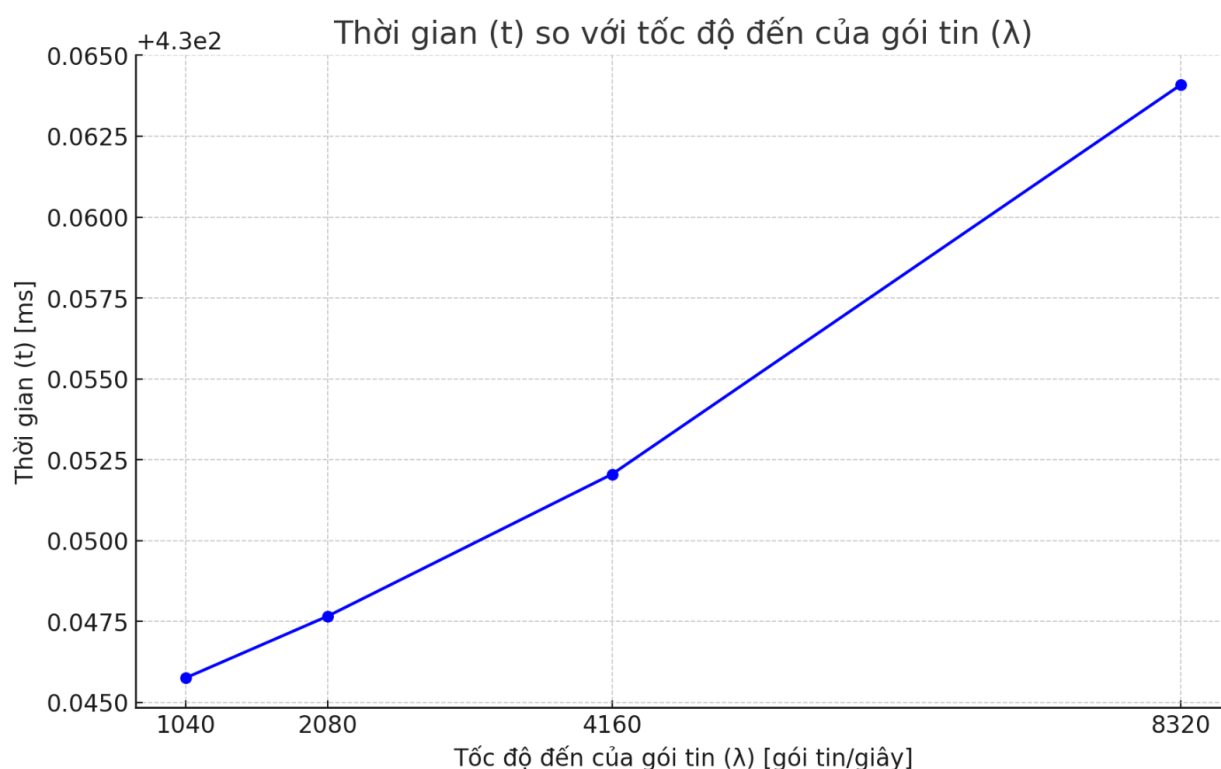
Hình 3-7: Mô hình giả định cho thực nghiệm tính toán trễ trong phạm vi

M1

Hình 3-7 thể hiện mô hình giả định để tính toán độ trễ trong phạm vi mạng vô tuyến, cụ thể là ở phạm vi M1 theo như mô hình trễ thể hiện ở hình 3-2 với các thông số giả định được thiết lập như trong Bảng 3-3. Tính toán cho các công thức (3-15) – (3-17) ta có kết quả quan hệ giữa tốc độ tới của gói tin và thời gian xử lý (độ trễ tại phạm vi mạng này như thể hiện trong Hình 3-8.

Bảng 3-3: Thông số giả định thiết lập tính toán cho mạng tại phạm vi M1 (λ bằng nhau cho cả UL và DL)

Số lượng gNB	6
Khoảng cách gNB – M1 (d_{gNB-M1})	3 km
Tốc độ (v)	Xấp xỉ 200000km/s
Tốc độ dịch vụ trung bình λ tại 1 gNB (λ_{gNB})	10, 20, 40, 80 vehicle/km tương đương 1040, 2080, 4160, 8320 packets/s
Dung lượng liên kết (C)	gNB – M1 : 10 Gbps M1 – M2: 100 Gbps
Kích thước gói tin (B)	500 bytes
Dung lượng liên kết sử dụng (α)	10% (0.1)
Thời gian xử lý tại các nút ($t_{BH,p}$)	0.2 ms



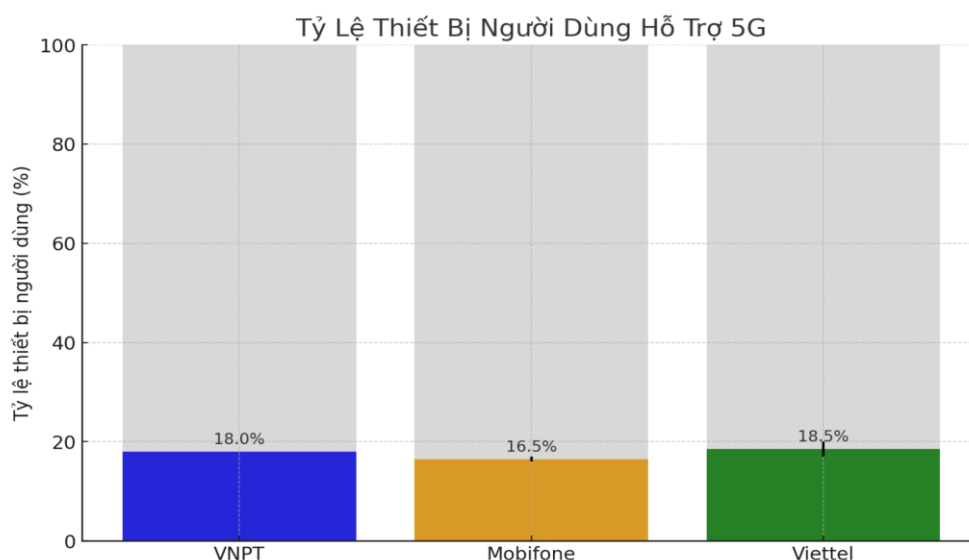
Hình 3-8: Độ trễ xử lý tại phạm vi mạng M1 với sự thay đổi tốc độ tới λ của gói tin

Nhận xét: Mật độ xe, tức số lượng xe trong một khu vực nhất định, ảnh hưởng trực tiếp đến tốc độ đến trung bình (λ) trong hệ thống mạng. Khi mật độ xe tăng, số lượng gói tin gửi đi cũng tăng, dẫn đến việc tăng tốc độ đến của gói tin. Theo lý thuyết hàng đợi M/M/1, thời gian chờ đợi và xử lý gói tin trong hệ thống sẽ tăng lên khi tốc độ đến của gói tin tiến gần tới tốc độ phục vụ μ . Cụ thể, khi λ tiệm cận μ , thời gian xếp hàng tăng lên, kéo theo truyền tải tăng dần lên, gây ra hiện tượng nghẽn mạng và tăng độ trễ trong truyền thông V2X. Điều này có thể dẫn đến suy giảm hiệu suất của các ứng dụng yêu cầu độ trễ thấp và độ tin cậy cao như cảnh báo an toàn, điều khiển tự động và truyền thông giữa các phương tiện. Do đó, các nhà cung cấp dịch vụ phải phát triển các kỹ thuật và giải pháp để cân bằng tải, cải thiện khả năng phục vụ, đảm bảo hệ thống mạng hoạt động hiệu quả ngay cả khi mật độ xe tăng cao.

3.5 Các thách thức trong triển khai 5G tại Việt Nam

Tại Việt Nam, các nhà mạng VNPT, Viettel, và MobiFone đã bắt đầu thí điểm dịch vụ 5G từ cuối năm 2019 tại 55 tỉnh thành phố, các khu công nghiệp, khu công nghệ cao và trường đại học. Sau hơn 4 năm thử nghiệm, các nhà mạng

đã thu thập đủ dữ liệu để đánh giá chất lượng, mạng lưới hạ tầng, và xây dựng chiến lược kinh doanh. Công nghệ 5G đã được thương mại hóa trên thế giới trong nhiều năm qua và trở thành xu hướng tất yếu. Việt Nam không thể đứng ngoài xu thế này, và việc Chính phủ chọn năm 2024 để triển khai là quyết định hợp lý, không quá muộn nhưng cũng không quá sớm.



Hình 3-9: Tỷ lệ thiết bị 5G tại 3 nhà mạng

Dù đã sẵn sàng về mặt kỹ thuật, đại diện các nhà mạng lớn tại Việt Nam cho biết số lượng thiết bị hỗ trợ 5G hiện còn thấp. Cụ thể, chỉ khoảng 18% thiết bị của người dùng VNPT, 16%-17% thiết bị của MobiFone và 17%-20% thiết bị của Viettel hỗ trợ công nghệ này. Vì vậy, các nhà mạng ưu tiên triển khai 5G tại những khu vực có nhu cầu cao và tỷ lệ thiết bị hỗ trợ 5G lớn.

Thách thức lớn nhất trong việc triển khai 5G không nằm ở công nghệ mà ở bài toán kinh doanh và quản trị hệ thống hiệu quả. Khó khăn trước hết là cho các doanh nghiệp, bao gồm vấn đề đấu giá băng tần phức tạp và sự chần chừ trong đầu tư. Đầu tư cho 5G đòi hỏi chi phí lớn, và thị trường vẫn chưa rõ ràng về mức độ chấp nhận dịch vụ này, dù cơ hội tiềm năng là rất lớn. Ngoài ra, giá các dịch vụ viễn thông truyền thống đang giảm, làm tăng thêm thách thức cho các nhà mạng trong việc đầu tư vào hệ thống mới. Doanh nghiệp tư nhân có thể chấp nhận lỗ 2-3 năm trước khi đạt điểm hòa vốn và có lợi nhuận, nhưng doanh nghiệp nhà nước lại gặp khó khăn trong việc đầu tư ồ ạt do phải cân nhắc kỹ lưỡng về cơ chế đầu tư và hoạt động. Đây là thách thức không nhỏ mà các nhà mạng Việt Nam

cần vượt qua để triển khai thành công 5G.

3.6 Kết luận chương 3

Mạng 5G có thể cung cấp các dịch vụ V2X phụ thuộc vào việc nó có thể đáp ứng các yêu cầu về độ trễ cực thấp theo yêu cầu đề ra cho dịch vụ này. Các giải pháp, cấu hình triển khai mạng và kích bản số lượng, vị trí của thiết bị đầu cuối sẽ ảnh hưởng mạnh đến độ trễ E2E cho dịch vụ V2X. Nghiên cứu trong chương thực hiện các tính toán thực nghiệm để khảo sát các yếu tố tác động đến trễ ở các phạm vi mạng khác nhau khi triển khai các dịch vụ V2X – một dịch vụ mang tính thách thức về hiệu năng thực thi yêu cầu về trễ cực thấp trong mạng 5G.

Độ trễ vô tuyến trong mạng 5G có thể giảm bằng cách sử dụng kỹ thuật truyền dựa trên khoảng cách sóng mang con hoặc kỹ thuật tiên tiến hơn. Kết quả tính toán chỉ ra rằng trong phạm vi mạng vô tuyến, các yếu tố về lựa chọn các tham số cấu hình truyền dẫn, chẳng hạn như số lượng, khoảng cách sóng mang con, độ rộng băng thông ảnh hưởng nhiều đến độ trễ truyền tải dịch vụ. Tuy nhiên, điều này đòi hỏi băng thông lớn hơn khi tải lưu lượng tăng lên. Độ trễ sinh ra trong mạng truyền tải backhaul phụ thuộc rất nhiều vào việc triển khai mạng và tỷ lệ dung lượng liên kết được phân bổ để hỗ trợ lưu lượng V2X. Phân tích và tính toán cũng cho thấy thấy số lượng người sử dụng đầu cuối (mật độ xe), tỷ lệ phục vụ (năng lực xử lý tính toán của các nút mạng), yêu cầu đảm bảo độ tin cậy đều ảnh hưởng đáng kể đến độ trễ trong mạng 5G cho các ứng dụng V2X. Mật độ xe cao làm tăng tải lưu lượng, dẫn đến gia tăng độ trễ. Tỷ lệ phục vụ cao giúp giảm thiểu độ trễ bằng cách xử lý nhanh các yêu cầu từ xe. Đảm bảo độ tin cậy trong truyền tải dữ liệu giữa các xe và cơ sở hạ tầng cũng góp phần giảm độ trễ và tăng tính ổn định của hệ thống. Do đó, việc cân nhắc đồng thời các yếu tố này là cần thiết để tối ưu hóa hiệu suất mạng 5G cho V2X, đặt nền tảng cho các giải pháp mạng 5G thông minh, hiệu quả và linh hoạt hơn trong định cấu hình, cài đặt tài nguyên mạng thích ứng với trạng thái lưu lượng, dịch vụ tại các nút và phạm vi mạng cụ thể.

Nội dung nghiên cứu của chương cũng phân tích hiện trạng, những thách thức về triển khai mạng 5G tại Việt Nam. Tại Việt Nam, triển khai 5G là xu hướng tất yếu. Tuy nhiên việc triển khai 5G tại Việt Nam hiện đang trong giai đoạn thử

nghiệm. Triển khai thương mại mạng 5G để cung cấp các dịch vụ tiên tiến, yêu cầu độ trễ thấp cần phải cân nhắc, xem xét dựa trên nhiều yếu tố, không chỉ là các yếu tố thực thi kỹ thuật, công nghệ, mà còn cả các yếu tố đầu tư, kinh doanh để từ đó một lộ trình triển khai phù hợp.

KẾT LUẬN CỦA LUẬN VĂN

Mục tiêu của luận văn thực hiện nghiên cứu về trễ trong mạng 5G. Theo đó, nội dung của các chương thể hiện những kết quả nghiên cứu tìm hiểu, phân tích, đánh giá hiệu quả thực thi trễ tại các phân đoạn mạng 5G để từ đó đưa ra các nhận xét, kết luận và đề xuất về tối ưu độ trễ trong mạng 5G. Cụ thể như sau:

Chương 1 cung cấp một cái nhìn tổng quan về các giai đoạn phát triển của mạng 5G thông qua các phiên bản được cải tiến của nó. Tại mỗi một phiên bản phát triển, các kỹ thuật, công nghệ mới được xem xét áp dụng để từng bước hoàn thiện công nghệ này triển khai trong thực tiễn đáp ứng mục tiêu đề ra. Một số kỹ thuật, công nghệ mới nổi cho các mạng truyền thông di động 5G như công nghệ như MIMO lớn, công nghệ sóng cực ngắn mmWave, truyền sóng dạng búp ở thời điểm hiện tại cũng như trong tương lai gần sẽ được ứng dụng triển khai. Điều này sẽ tác động, ảnh hưởng lớn tới phương pháp tiếp cận thiết kế và phát triển nhằm đạt được các mục tiêu thực thi mạng 5G, đặc biệt là truyền tải dữ liệu với yêu cầu độ trễ cực thấp cho một số dịch vụ của mạng 5G. Các giải pháp điển hình được trình bày là kiến trúc phân tách mặt phẳng trễ, hệ thống điều chế và mã hóa tiên , phương pháp tiếp cận thực thi kênh điều khiển, mã hóa kênh, RAN đám mây và các khía cạnh thực thi kỹ thuật khác để tăng cường hiệu năng thực thi của mạng cho mục tiêu đạt được độ trễ truyền tải thấp.

Nội dung nghiên cứu chương 2 giới thiệu, mô tả, phân tích các tham số, chỉ tiêu, yêu cầu cũng như các thành phần tác động đến trễ trong mạng 5G. Các giải pháp khác nhau có thể được áp dụng phù hợp với chiến lược triển khai trong các phạm vi như RAN, mạng lõi và bộ nhớ đệm để đạt được mục tiêu thực thi các yêu cầu về trễ. Các giải pháp điển hình như kiến trúc phân tách mặt phẳng trễ, hệ thống điều chế và mã hóa, phương pháp tiếp cận thực thi kênh điều khiển, mã hóa kênh, RAN đám mây và các khía cạnh thực thi kỹ thuật khác được triển khai để tăng cường hiệu năng thực thi của mạng cho mục tiêu đạt được độ trễ truyền tải thấp. Mạng 5G ngoài việc cung cấp băng thông cực rộng, các yêu cầu về độ trễ cực nhỏ của một số dịch vụ sẽ tác động lớn đến các khía cạnh thiết kế, đáp ứng thực thi của hệ thống trong một phạm vi phổ sóng rộng. Một số khía cạnh thực thi ở lớp vật lý cần được nghiên cứu cải tiến, chẳng hạn như kỹ thuật điều khiển

tạo chùm tia tương thích với thành phần dữ liệu hoặc kỹ thuật tái cấu hình từng phần dựa trên giao thức URLLC. Tuy nhiên, nghiên cứu cũng chỉ ra rằng, những vấn đề về công nghệ, kỹ thuật, giải pháp được nêu hầu hết có tính mở, cần phải có những nghiên cứu bổ sung, thử nghiệm với những kịch bản triển khai cụ thể để hoàn thiện.

Nghiên cứu trong chương 3 thực hiện các tính toán thực nghiệm để khảo sát các yếu tố tác động đến trễ ở các phạm vi mạng khác nhau khi triển khai các dịch vụ V2X – một dịch vụ mang tính thách thức về hiệu năng thực thi yêu cầu về trễ cực thấp trong mạng 5G. Mạng 5G có thể cung cấp các dịch vụ V2X phụ thuộc vào việc nó có thể đáp ứng các yêu cầu về độ trễ cực thấp theo yêu cầu đề ra cho dịch vụ này. Các giải pháp, cấu hình triển khai mạng và kịch bản số lượng, vị trí của thiết bị đầu cuối sẽ ảnh hưởng mạnh đến độ trễ E2E cho dịch vụ V2X. Kết quả tính toán chỉ ra rằng trong phạm vi mạng vô tuyến, các yếu tố về lựa chọn các tham số cấu hình truyền dẫn, chẳng hạn như số lượng, khoảng cách sóng mang con, độ rộng băng thông ảnh hưởng nhiều đến độ trễ truyền tải dịch vụ. Độ trễ mạng truyền tải backhaul phụ thuộc rất nhiều vào việc cấu hình triển khai mạng và tỷ lệ dung lượng liên kết được phân bổ để hỗ trợ lưu lượng V2X tương ứng với tình trạng biến động của số người sử dụng dịch vụ. Điều này đòi hỏi mạng 5G phải có những chức năng quản trị cấu hình thông minh, hiệu quả và linh hoạt hơn; đảm bảo tài nguyên mạng được sử dụng tối ưu, thích ứng với trạng thái lưu lượng, dịch vụ được cung cấp, truyền tải qua các phạm vi mạng. Ngoài ra nghiên cứu của chương cũng phân tích hiện trạng, những thách thức về triển khai mạng 5G tại Việt Nam để từ đó đưa ra một lộ trình triển khai phù hợp đối với mạng 5G trong nước.

Những thách thức về thực thi mạng để đạt được độ trễ thấp vẫn còn là vấn đề cần phải giải quyết trong mọi phạm vi mạng 5G. Nó đòi hỏi phải tiếp tục nghiên cứu hoàn thiện các kỹ thuật, công nghệ đã có cũng như các kỹ thuật công nghệ mới. Kỹ thuật antenna multiple-input multiple-output (MIMO) hiện nay đang hướng tới giải quyết tối ưu hóa sử dụng tần số và các biện pháp cải thiện cường độ tín hiệu thu phát để từ đó cải thiện tính ổn định và khả năng đáp ứng nhanh của hệ thống. Công nghệ điện toán biên đa truy cập/di động (MEC) với giải pháp đưa dữ liệu và tài nguyên tính toán ra biên đến gần người dùng hiện đang được nghiên cứu; hứa hẹn cải thiện tăng tốc độ xử lý, hiệu quả sử dụng băng

thông để giảm độ trễ truyền tải các dịch vụ. Kỹ thuật điều khiển và quản lý tài nguyên thông minh cũng là một hướng nghiên cứu tiếp theo hứa hẹn mang lại những cải tiến đáng kể về cải thiện độ trễ trong mạng 5G. Các thuật toán điều khiển đơn giản và hiệu quả giúp điều phối thông lượng và tối ưu hóa việc sử dụng tài nguyên mạng, đảm bảo rằng các dịch vụ như V2X, dịch vụ điều khiển các quá trình tự động hóa trong sản xuất công nghiệp thông minh hoạt động một cách hiệu quả và ổn định.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1]. Alex Barakabitze, Md. A. A. (2015). Evolution of LTE and Related Technologies towards IMTAdvanced. Evolution of LTE and Related Technologies towards IMT-Advanced, 1 4.
- [2]. Holma, H., & Toskala, A. (2012). LTE Advanced. Hoboken, NJ, États-Unis : Wiley.
- [3]. Vaezi, M., Ding, Z., & Poor, H. V. (Éd.). (2019). Multiple Access Techniques for 5G Wireless Networks and Beyond.
- [4]. Hoymann, C., Astely, D., Stattin, M., Wikstrom, G., Cheng, J.-F., Høglund, A., Gunnarsson, F. (2016). LTE release 14 outlook. IEEE Communications Magazine, 54(6), 44 49.
<https://doi.org/10.1109/mcom.2016.7497765>
- [5]. 3GPP Release 15 Overview. (s. d.). Consulté le 15 juillet 2020, à l'adresse <https://spectrum.ieee.org/telecom/wireless/3gpp-release-15-overview>
- [6]. Holma, H., Toskala, A., & Nakamura, T. (2019). 5G Technology. Hoboken, NJ, États-Unis, Wiley
- [7]. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Release 17 Description; Summary of Release 17 Work Items (Release 17), 2021

- [8]. Ma, Z., Zhang, Z., Ding, Z., Fan, P., & Li, H. (2015). Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives. *Science China Information Sciences*, 58(4), 1 20. <https://doi.org/10.1007/s11432-015-5293-y>
- [9]. Bennis, M., Debbah, M., & Poor, H. V. (2018). Ultrareliable and Low-Latency Wireless Communication : Tail, Risk, and Scale. *Proceedings of the IEEE*, 106(10), 1834 1853.
- [10]. Abuu B. Kihero, Muhammad, Sohaib J. Solaijam, Husayin Arslan, Inter-Numerology Interference for Beyond 5G, *IEEE Access* Vol.7, 2019
- [11]. Radiocommunication sector of ITU. (2017). Minimum requirements related to technical performance for IMT-2020 radio interface(s) (Report ITU-R M.2410-0)
- [12]. Parvez, I., Rahmati, A., Guvenc, I., Sarwat, A. I., & Dai, H. (2018). A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions. *IEEE Communications Surveys & Tutorials*, 20(4), 3098 3130.
- [13]. Hyounghu Ji, Sunho Park. (2018). Ultra Reliable and Low Latency Communications in 5G Downlink: Physical Layer Aspects. Samsung Electronics, Korea. Dept. of Electrical and Computer Engineering, Seoul National University, Korea Samsung Electronics, Korea.
- [14]. Lei, Wan & Soong, Anthony & Jianghua, Liu & Yong, Wu & Classon, Brian & Xiao, Weimin & Mazzaresse, David & Yang, Zhao & Saboorian, Tony. (2020). 5G System Design: An End to End Perspective. 10.1007/978-3-030-22236-9.
- [15]. She, C., Yang, C., & Quek, T. Q. S. (2017). Radio Resource Management for Ultra-Reliable and Low-Latency Communications. *IEEE Communications Magazine*, 55(6), 72 78.
<https://doi.org/10.1109/mcom.2017.1601092>
- [16]. I.-G. Jang and G.-D. Jo, "Study on the latency efficient IFFT design

- method for low latency communication systems,” in Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS), Oct. 2016
- [17]. Dutta, Sourjya & Mezzavilla, Marco & Ford, Russell & Zhang, Menglei & Rangan, Sundeep & Zorzi, Michele. (2016). MAC layer frame design for millimeter wave cellular system. 117-121. 10.1109/EuCNC.2016.7561016.
- [18]. "5G New Radio Evaluation Against IMT-2020 Key Performance Indicators" Digital Object Identifier 10.1109/ACCESS.2020.3001641, June. 2020.
- [19]. NR; Physical Layer Procedures for Data (Release 15), document 3GPP TS 38.214 v15.9.0, 3GPP, Apr. 2020.
- [20]. NR; Physical Channels and Modulation (Release 15), document 3GPP TS 38.211 v15.8.0, 3GPP, Jan. 2020.
- [21]. EB. Coll-Perales, M.C. Lucas-Estañ, T. Shimizu, J. Gozalvez, Senior, T. Higuchi, S. Avedisov, O. Altintas, M. Sepulcre, "End-to-End V2X Latency Modeling and Analysis in 5G Networks", IEEE Transactions on Vehicular Technology, Nov. 2022.
- [22]. O. Al-Saadeh, et al., “End-to-End Latency and Reliability Performance of 5G in London”, Proc. IEEE GLOBECOM, pp. 1-6, Dec. 2018.
- [23]. D. Bertsekas and R. Gallager, “Data Networks (Chapter 3. Delay Models in Data Networks)”, Prentice Hall, 1992.
- [24]. L. Cominardi, et al., "Understanding QoS Applicability in 5G Transport Networks", Proc. IEEE BMSB, pp. 1-5, Jun. 2018.
- [25]. Q. Ye, et al., "End-to-End Delay Modeling for Embedded VNF Chains in 5G Core Networks", IEEE Internet Things J., vol. 6(1), pp. 692-704, Feb. 2019.