

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC BÁCH KHOA**



## **BÁO CÁO ĐỒ ÁN 2**

**NÂNG CAO CHẤT LƯỢNG NGƯỜI DÙNG TRONG MẠNG  
MẬT ĐỘ CAO KẾT HỢP VỚI NỀN TẢNG TRUY CẬP VÔ  
TUYÊN ĐÁM MÂY (CRAP-i-UDN)**

**Sinh viên thực hiện:**

**NGUYỄN VĂN THẮNG**

**MSSV: 1910543**

**Cán bộ hướng dẫn:**

**TS. NGUYỄN ĐÌNH LONG**

**TP. HỒ CHÍ MINH, ngày 30 tháng 05 năm 2023**

## LỜI MỞ ĐẦU

Mạng di động không dây đã trở thành một phần không thể thiếu trong đời sống hằng ngày của chúng ta, nhu cầu trao đổi thông tin tốc độ cao và dữ liệu lớn ngày càng tăng. Các công nghệ mạng thế hệ trước không thể đáp ứng nổi dung lượng dữ liệu với mật độ người dùng khổng lồ. Thế hệ mạng di động thứ 5 (5G) với công nghệ nền tảng là mạng mật độ siêu dày (UDN) được đề xuất, cung cấp tốc độ dữ liệu cực cao, độ trễ thấp với mật độ người dùng dày đặc trên một đơn vị diện tích. Tuy nhiên, mạng UDN phải đối mặt với một số thách thức, chẳng hạn như nhiễu, khả năng hạn chế và các vấn đề về khả năng mở rộng. Để giải quyết những thách thức này, kiến trúc mạng truy cập vô tuyến đám mây (C-RAN) đã được đề xuất như một giải pháp tiềm năng. Kiến trúc C-RAN phân tách các chức năng xử lý cơ sở, cho phép sử dụng hiệu quả hơn các tài nguyên mạng và khả năng mở rộng được cải thiện. Kiến trúc này có khả năng nâng cao trải nghiệm người dùng trong các mạng UDN bằng cách cung cấp hiệu suất mạng tốt hơn, tăng công suất và độ tin cậy được cải thiện.

Đề án này nghiên cứu về việc sử dụng kiến trúc C-RAN trong việc nâng cao trải nghiệm người dùng trong các mạng UDN. Mục đích xác định các yếu tố chính tác động đến trải nghiệm người dùng và nghiên cứu hiệu suất của các cấu hình C-RAN khác nhau. Đồng thời cũng nghiên cứu cách tối ưu hóa kiến trúc mạng để đảm bảo rằng nó có thể đáp ứng các yêu cầu của nhiều các trường hợp sử dụng khác nhau của người dùng.

## **KẾ HOẠCH THỰC HIỆN**

Bố cục của đồ án sẽ đảm bảo bao hàm các khía cạnh của nghiên cứu và được chia ra các phần như sau:

### **Lời mở đầu**

Giới thiệu tổng quan về đề tài, mục đích, kết quả đạt được và hướng nghiên cứu phát triển.

### **Phần 1: Mô hình nghiên cứu: Ultra Dense Network (UDN) kết hợp nền tảng C-RAN (CRAP-i-UDN)**

1. Xu hướng phát triển và sự cần thiết triển khai kiến trúc C-RAN cho mạng UDN (CRAP-i-UDN)?
2. Những thách thức và hướng giải quyết vấn đề trong khi triển khai CRAP-i-UDN
3. Mục tiêu của đồ án

### **Phần 2: Cơ sở lý thuyết của CRAP-i-UDN**

1. Tổng quan và thách thức trong mạng mật độ siêu dày UDN
2. Tổng quan kiến trúc C-RAN và tiềm năng triển khai trong mạng UDN
3. Tính khả thi khi phát triển UDN kết hợp với C-RAN
4. Phát triển và triển khai mô hình mạng CRAP-i-UDN
5. Xu hướng tập trung vào chất lượng người dùng và mục tiêu nâng cao trải nghiệm người dùng trong CRAP-i-UDN

### **Phần 3: Đặt vấn đề và xây dựng mô hình hệ thống CRAP-i-UDN**

1. Kỹ thuật phân cụm để giảm tính phức tạp trong mạng UDN, kiến trúc C-RAN
2. Kỹ thuật phân cụm: Disjoint clustering và User-centric clustering

### **Phần 4: Nghiên cứu cách tối ưu hóa kiến trúc mạng để đáp ứng phù hợp nhiều trường hợp sử dụng của người dùng.**

Nghiên cứu các vấn đề tối ưu hóa trong kiến trúc mạng C-RAN, phân tích các bài toán tối ưu hóa và các thuật toán đề xuất để giải quyết vấn đề

### **Phần 5: Kết luận và hướng phát triển.**

Ở nội dung này, tôi đưa ra các kết luận về mặt kỹ thuật cho các thuật toán đề xuất trên, những ưu điểm và nhược điểm của hai phương pháp phân cụm trên. Sau đó, các

đề xuất giải pháp giúp cải thiện khuyết điểm cũng như phát triển những ưu điểm của các vấn đề đạt được cũng được trình bày.

## **Danh mục các từ viết tắt**

<b>Từ viết tắt</b>	<b>Nghĩa tiếng Anh</b>	<b>Nghĩa tiếng Việt</b>
AI	Artificial Intelligence	Trí tuệ nhân tạo
AP	Access Point	Điểm truy cập
AR/VR	Augmented Reality/Virtual Reality	hai công nghệ liên quan đến trải nghiệm tương tác và thực tế ảo
BBU	Baseband Unit	Đơn vị cơ sở dữ liệu
BS	Base station	Trạm phát sóng
CAPEX	Capital Expenditure	Chi phí vốn
CN	Core Network	Mạng lõi
CoMP	Coordinated Multipoint	Đa điểm phối hợp
CPRI	Common Public Radio Interface	Giao diện vô tuyến chung
C-RAN	Cloud Radio Access Network	Mạng vô tuyến đám mây
CSI	Channel State Information	Thông tin trạng thái kênh truyền
DL	Downlink	Truyền dẫn từ trạm gốc đến thiết bị người dùng
eMBB	Enhanced Mobile Broadband	Dịch vụ với cải tiến và mở rộng khả năng truyền dẫn dữ liệu di động rộng rãi, đạt tốc độ cao và dung lượng lớn
HDTV	High-Definition Television	Công nghệ truyền hình độ phân giải cao
HeNB	Home eNodeB	Trạm cơ sở di động gia đình LTE
HetNet	Heterogeneous Network	Mạng không đồng nhất
ICIC	Inter-Cell Interference Coordination	Kỹ thuật giảm nhiễu và tăng cường hiệu suất giữa các cell trong mạng
IoT	Internet of Thing	Internet vạn vật
LA	Location Area	Khu vực địa lý
LTE	Long-Term Evolution	Tiến hóa dài hạn
MIMO	Multiple-Input Multiple-Output	Đa nhập đa xuất

MME	Mobility Management Entity	Thực thể quản lý di động
mMTC	Massive Machine Type Communications	Dịch vụ với kết nối hàng loạt các thiết bị máy móc, đặc biệt là các thiết bị IoT
NFV	Network Function Virtualization	Ảo hóa chức năng mạng
OFDMA	Orthogonal Frequency Division Multiple Access	Đa truy cập tần số trực giao
OPEX	Operating Expenditure	Chi phí hoạt động
OTT	Over-The-Top	truyền thông qua Internet
PDN	Packet Data Network	Mạng dữ liệu gói
QoS	Quality of Service	Chất lượng dịch vụ
RAT	Radio Access Techonoly	Công nghệ truy cập sóng vô tuyến
RRH	Remote Radio Head	Đầu vô tuyến từ xa
RRM	Radio Resource Management	Quản lý tài nguyên sóng radio
SON	Self-Organizing Network	Mạng tự tổ chức
UDN	Ultra Dense Network	Mạng mật độ siêu cao
UE	User Equipment	Thiết bị người dùng
UL	Uplink	Truyền dẫn từ thiết bị người dùng lên trạm gốc.
URLLC	Ultra-Reliable Low Latency Communications	Dịch vụ với kết nối vô tuyến có độ tin cậy cao và độ trễ thấp
WLAN	Wireless Local Area Network	Mạng không dây địa phương
4G	Fourth Generation Wireless Networks	Mạng vô tuyến thế hệ thứ 4
5G	Fifth Generation Wireless Networks	Mạng vô tuyến thế hệ thứ 5

# MỤC LỤC

<b>Chương 1: Mô hình nghiên cứu: Ultra Dense Network (UDN) kết hợp nền tảng C-RAN (CRAP-i-UDN) .....</b>	<b>1</b>
1.1. Xu hướng phát triển và sự cần thiết triển khai kiến trúc C-RAN cho mạng UDN (CRAP-i-UDN)? .....	1
1.2. Những vấn đề thách thức và hướng giải quyết trong khi triển khai CRAP-i-UDN... 5	
1.3. Định hướng nghiên cứu và mục tiêu đề án .....	9
<b>Chương 2: Cơ sở lý thuyết của CRAP-i-UDN .....</b>	<b>10</b>
2.1. Tổng quan và thách thức trong mạng mật độ siêu dày UDN.....	10
2.2. Tổng quan kiến trúc C-RAN và tiềm năng triển khai trong mạng UDN.....	16
2.3. Phát triển và triển khai mô hình mạng CRAP-i-UDN.....	20
2.4. Xu hướng tập trung vào chất lượng người dùng và mục tiêu nâng cao trải nghiệm người dùng trong CRAP-i-UDN.....	23
<b>Chương 3: Đặt vấn đề và xây dựng mô hình hệ thống CRAP-i-UDN.....</b>	<b>25</b>
2.1. Phân cụm để giảm tính phức tạp trong UDN, kiến trúc C-RAN.....	25
2.2. Kỹ thuật phân cụm Disjoint clustering và User-centric clustering trong mô hình CRAP-i-UDN .....	26
2.2.1. Kỹ thuật phân cụm Disjoint clustering .....	26
2.2.2. Kỹ thuật phân cụm User-centric clustering.....	27
<b>Chương 4: Nghiên cứu các vấn đề tối ưu hóa kiến trúc mạng.....</b>	<b>29</b>
4.1. Phân bổ tài nguyên trong C-RAN.....	29
4.1.1. Mô hình hệ thống .....	30
4.1.2. Bài toán tối ưu và giải thuật .....	31
4.2. Tối ưu hóa độ phủ sóng mạng trong C-RAN.....	36
4.2.1. Mô hình hệ thống .....	37
4.2.2. Bài toán tối ưu và giải thuật .....	38
4.3. Đặc vị trí và phân chia chức năng BBU .....	41
4.3.1. Mô hình phân chia chức năng BBU.....	42
4.3.2. Mô tả cấu trúc mạng .....	43
4.3.3. Mô hình hệ thống .....	44
4.3.4. Bài toán tối ưu và giải thuật .....	44
<b>Chương 5 : Kết luận và hướng phát triển.....</b>	<b>48</b>
<b>Tài liệu tham khảo .....</b>	<b>50</b>

# **Chương 1: Mô hình nghiên cứu: Ultra Dense Network (UDN)**

## **kết hợp nền tảng C-RAN (CRAP-i-UDN)**

### **1.1. Xu hướng phát triển và sự cần thiết triển khai kiến trúc C-RAN cho mạng UDN (CRAP-i-UDN)?**

Xu hướng đang phát triển của mạng di động hiện tại và trong tương lai là hướng đến mạng di động 5G. Mạng di động 5G được thiết kế để cung cấp tốc độ cao hơn, độ trễ thấp hơn và khả năng kết nối đồng thời cho nhiều thiết bị hơn. Có hai động lực chính cho các mạng di động 5G.

*Thứ nhất*, mạng di động trong tương lai sẽ xử lý các dịch vụ mới như giao tiếp cảm biến khối lượng lớn và giao tiếp từ phương tiện đến mọi thứ. yêu cầu thời gian thiết lập và độ trễ ngắn hơn, cũng như giảm lưu lượng thông tin điều khiển và năng lượng tiêu thụ. Ngoài ra, nó cũng có nhiều trường hợp sử dụng khác nhau, bao gồm quản lý đa tế bào (multi-cell) và đa người dùng (multi-user), triển khai mạng với đa công nghệ truy cập vô tuyến (multi-RAT hoặc mạng đa tầng (multi-layer networks), còn được gọi là Heterogeneous Networks (HetNets). Các kịch bản triển khai và dịch vụ phức tạp đòi hỏi các hệ thống 5G trong tương lai phải có khả năng i) hỗ trợ tốc độ chuyển đổi gấp 10-100 lần so với hiện nay; ii) đảm bảo độ trễ thấp (trong phạm vi vài mili giây); iii) hỗ trợ lên đến 10-100 lần số lượng thiết bị hơn so với hiện tại; iv) cung cấp mật độ lưu lượng truyền tải gấp 1000 lần; v) hỗ trợ tốc độ di chuyển nhanh lên đến 500km/h của thiết bị người dùng (UE). Đồng thời, mong muốn có phủ sóng 99,999%, trong khi tiêu thụ năng lượng và chi phí cho cơ sở hạ tầng không được tăng.

*Thứ hai*, các nhà điều hành mong muốn giảm chi phí vận hành mạng và cải thiện hiệu suất phổ trong một khu vực, để duy trì hiệu suất tối ưu trong các hệ thống di động trong tương lai bằng cách sử dụng tài nguyên linh hoạt hơn và các chức năng tự tổ chức tiên tiến hơn.

Để đảm bảo đáp ứng trên, các tiêu chuẩn liên quan đến 5G được đề xuất như URLLC mMTC và eMBB.



**URLLC** trong mạng 5G yêu cầu độ tin cậy cao và độ trễ thấp, chẳng hạn như điều khiển các thiết bị tự động như xe tự lái, robot, hoặc mô phỏng thực tế ảo (AR/VR). Độ trễ càng thấp, kết nối truyền thông không dây càng tương tác thời gian thực và chính xác hơn. Độ tin cậy cao là đảm bảo truyền tải dữ liệu một cách đáng tin cậy và không bị mất mát dữ liệu.

Các ứng dụng của URLLC trong mạng 5G bao gồm:

- Truyền tải video chất lượng cao
- truyền tải âm thanh đa kênh
- Truyền tải dữ liệu trong thời gian thực và điều khiển các thiết bị tự động
- Giám sát y tế từ xa, hỗ trợ cho các thủ tục phẫu thuật, và y tế di động
- Giải trí trực tuyến, AR/VR và các trò chơi trực tuyến

Mục tiêu của URLLC là tạo ra một mạng truyền thông không dây đáng tin cậy, có thể đáp ứng được yêu cầu về độ trễ và tin cậy cho các ứng dụng tương tác thời gian thực.

**mMTC** trong mạng 5G dành cho việc truyền tải dữ liệu từ hàng triệu thiết bị IoT hoặc các cảm biến, đòi hỏi khả năng kết nối đồng thời và sức mạnh tính toán. Các thiết bị IoT có thể là các thiết bị y tế, đèn chiếu sáng, các thiết bị trong gia đình thông minh, các thiết bị đo lường và giám sát môi trường và vật liệu

Các ứng dụng của mMTC trong mạng 5G bao gồm:

- Truyền tải dữ liệu cảm biến và dữ liệu IoT
- Giám sát và điều khiển các thiết bị từ xa
- Thu thập dữ liệu để phân tích và tối ưu hóa hệ thống
- Cải thiện quản lý năng lượng và tài nguyên

Mục tiêu của mMTC trong mạng 5G là tạo ra một mạng truyền thông không dây có khả năng kết nối với hàng triệu thiết bị IoT cùng một lúc, đồng thời đảm bảo độ ổn định và hiệu suất của hệ thống.

**eMBB** trong mạng 5G tập trung vào việc cải thiện tốc độ và hiệu suất truyền tải dữ liệu di động, bao gồm các ứng dụng yêu cầu băng thông cao và tốc độ truyền tải

nhanh hơn, chẳng hạn như truyền tải video 4K/8K, streaming trực tuyến, trò chơi đa người chơi trực tuyến và thực tế ảo.

Các ứng dụng của eMBB trong mạng 5G bao gồm:

- Truyền tải video chất lượng cao (4K/8K)
- Truyền tải âm thanh đa kênh
- Streaming trực tuyến và chia sẻ nội dung đa phương tiện
- Trò chơi đa người chơi trực tuyến
- Thực tế ảo và thực tế tăng cường

Mục tiêu của eMBB trong mạng 5G là tạo ra một mạng truyền thông không dây có khả năng truyền tải dữ liệu nhanh hơn, ổn định hơn và có băng thông rộng hơn, đảm bảo các ứng dụng di động yêu cầu băng thông cao và tốc độ truyền tải nhanh hơn được sử dụng một cách hiệu quả.

Để đáp ứng các tiêu chuẩn URLLC, mMTC và eMBB trong mạng di động 5G, cần phải sử dụng mô hình mạng đáp ứng được các yêu cầu đặc biệt của từng tiêu chuẩn. Một trong những mô hình mạng được đề xuất để đáp ứng các yêu cầu của mạng di động 5G là mô hình mạng với mật độ siêu cao UDN. Mô hình mạng UDN cho phép triển khai các trạm thu phát sóng và các thiết bị kết nối IoT với mật độ cao hơn, cung cấp khả năng truyền tải dữ liệu với độ trễ thấp và độ tin cậy cao hơn. Hơn nữa, mô hình UDN cũng cho phép triển khai các anten thông minh và hệ thống học máy để tối ưu hóa hoạt động mạng.

Theo nghiên cứu của IMT-2020, Bảng 1 liệt kê một số kịch bản UDN điển hình và các chỉ số hiệu suất chính cần thiết của hệ thống 5G, cho thấy không chỉ có tốc độ dữ liệu cao mà còn đem lại trải nghiệm người dùng tuyệt vời và thời gian trễ nghiêm ngặt được mong đợi trong các kịch bản UDN.

Bảng 1.1: Các kịch bản của mô hình mạng UDN và các chỉ số đánh giá

Chỉ số đánh giá	Mật độ lưu lượng truyền tải (bps/km <sup>2</sup> , DL/UL)	Thời gian trễ (ms)	Trải nghiệm người dùng (Mbps, DL/UL)	Tính di động (Km/h)

Ngoại ô	1T/130G	15 - 40	1024/512	0-3 cho người đi bộ, 50 cho ô tô
Khu tập trung thương mại cao	1102G/150G	5- 10	1024/512	0-3
Đường sắt, tàu ngầm	10T/-	15 - 40	1024/512	110

Tuy nhiên, mạng UDN phải đối mặt với một số thách thức, chẳng hạn như nhiễu, khả năng hạn chế và các vấn đề về khả năng mở rộng. Để giải quyết những thách thức này, kiến trúc mạng truy cập vô tuyến đám mây (C-RAN) đã được đề xuất như một giải pháp tiềm năng và cần thiết cho mô hình mạng UDN.

Kiến trúc C-RAN (Cloud Radio Access Network) là một lựa chọn tiềm năng cho mô hình mạng UDN (Ultra-Dense Network) vì nhiều lý do:

1. Sử dụng tài nguyên hiệu quả: Kiến trúc C-RAN cho phép sử dụng tài nguyên mạng một cách hiệu quả bằng cách tập trung chức năng xử lý baseband tại trung tâm dữ liệu từ xa. Điều này giúp giảm yêu cầu phần cứng tại cạnh mạng và cho phép sử dụng tần số có sẵn một cách hiệu quả hơn.
2. Khả năng mở rộng: Mạng UDN yêu cầu khả năng mở rộng cao để hỗ trợ số lượng thiết bị lớn và tốc độ dữ liệu cao. Kiến trúc C-RAN cho phép mở rộng dễ dàng bằng cách thêm các đơn vị xử lý vào trung tâm dữ liệu trung tâm.
3. Độ trễ thấp hơn: Kiến trúc C-RAN giảm độ trễ tổng thể của mạng bằng cách tập trung xử lý baseband, giúp cho việc xử lý dữ liệu nhanh hơn và giảm độ trễ giữa thiết bị di động và trung tâm dữ liệu.
4. Cải thiện phạm vi phủ sóng: Kiến trúc C-RAN cung cấp phạm vi phủ sóng và khả năng cải thiện bằng cách sử dụng các kỹ thuật beamforming tiên tiến và các thuật toán tiên tiến để phân bổ tài nguyên và quản lý nhiễu.
5. Chi phí thấp hơn: Bằng cách tập trung xử lý baseband tại trung tâm dữ liệu từ xa, kiến trúc C-RAN giảm yêu cầu phần cứng và chi phí bảo trì tại cạnh mạng.

Nó cũng cho phép sử dụng tần số có sẵn một cách hiệu quả hơn, có thể giảm chi phí thêm nữa.

Tổng thể, kiến trúc C-RAN có tiềm năng cung cấp các giải pháp hiệu quả, có khả năng mở rộng, độ trễ thấp và tiết kiệm chi phí cho các mô hình mạng UDN.

## **1.2. Những vấn đề thách thức và hướng giải quyết trong khi triển khai CRAP-i-UDN**

Mô hình mạng UDN là một kịch bản quan trọng cho công nghệ mạng 5G trong tương lai. Trong kịch bản này, chúng ta sẽ bàn về những thách thức điển hình của việc triển khai kiến trúc C-RAN trong mô hình mạng này và những định hướng giải quyết cho các vấn đề đó.

### *Độ phức tạp tính toán cao (High Computational Complexity)*

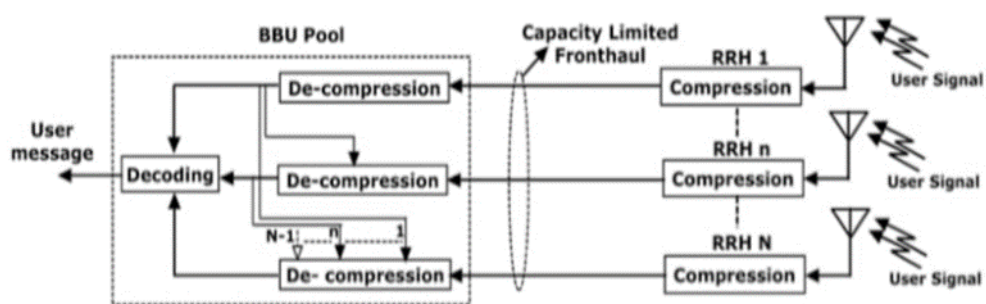
Trong C-RAN dày đặc (CRAP-i-UDN), bể BBU thường hỗ trợ một lượng lớn các RRH, và số lượng biến cần tối ưu, chẳng hạn như các vector beamforming sẽ trở nên quá nhiều, ngay cả trong bối cảnh tính toán đám mây. Kỹ thuật phổ biến nhất để giảm độ phức tạp là sử dụng kỹ thuật gom cụm. Cơ bản, có hai loại kỹ thuật nhóm trong mạng C-RAN là nhóm không giao nhau (disjoint clustering) và nhóm tập trung vào người dùng (user-centric clustering). Với disjoint clustering, tất cả các RRH trong mạng được chia thành một số nhóm không giao nhau, và các RRH trong mỗi nhóm sử dụng kỹ thuật CoMP để phục vụ các người dùng trong nhóm đó. Tuy nhiên, người dùng ở vùng biên của cụm vẫn gặp phải nhiễu từ các cụm lân cận.

Ngược lại, đối với user-centric clustering, mỗi người dùng được phục vụ riêng bởi các trạm thu phát gần đó. Người dùng được sắp xếp là trung tâm của cụm. Các cụm khác nhau có thể chồng lên nhau, điều này sẽ loại bỏ hiệu ứng nhiễu giữa các cụm. Hai kỹ thuật phân cụm này sẽ được phân tích kỹ hơn và được phô phỏng ở các phần kế tiếp.

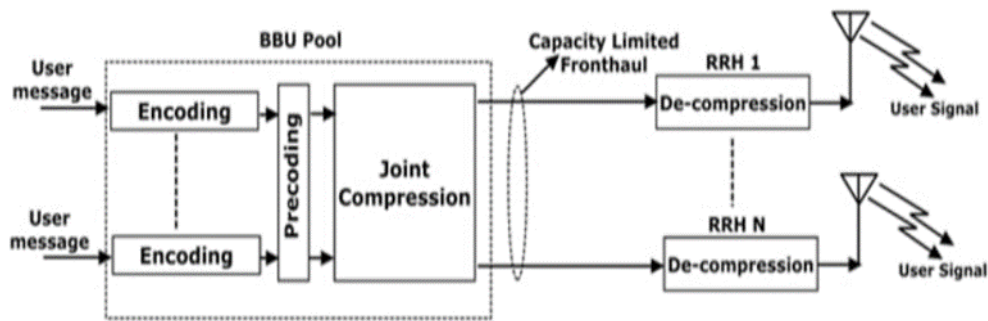
### *Yêu cầu nghiêm ngặt về khả năng của Fronthaul (Stringent Fronthaul Capacity Requirement)*

Trong C-RAN thông thường, các liên kết fronthaul thường là các liên kết dây, chẳng hạn như sợi quang. Tuy nhiên, trong C-RAN siêu mật độ, một lượng lớn các liên kết fronthaul được yêu cầu. Việc đặt các liên kết dây yêu cầu chi phí vận hành và bảo trì cao. Một lựa chọn hấp dẫn là sử dụng các liên kết fronthaul không dây, chẳng hạn như liên lạc sóng siêu cao tần (mmWave) được coi là có khả năng mở rộng và hiệu quả chi phí hơn so với các liên kết dây. Tuy nhiên, băng thông của các liên kết không dây thấp hơn rất nhiều so với các liên kết dây, điều này có nghĩa là số lượng người dùng được hỗ trợ bởi mỗi liên kết không dây ít hơn nhiều. Ràng buộc năng lực fronthaul đã được nghiên cứu một cách rộng rãi và có thể chia thành hai loại: chiến lược nén và chiến lược chia sẻ dữ liệu.

Trong chiến lược nén, các tín hiệu nhận được tại mỗi RRH được nén và truyền đến BBU pool trên một công suất giới hạn Fronthaul ở đường uplink và ngược lại các tín hiệu được nén lại và được truyền từ BBU pool đến các RRH ở đường downlink. Khả năng truyền tải của fronthaul liên quan đến mức độ tốt của độ phân giải của tín hiệu nén: độ phân giải càng cao thì yêu cầu khả năng truyền tải fronthaul càng cao. Do đó, độ phân giải nén nên được tối ưu hóa dưới ràng buộc về khả năng truyền tải của fronthaul. Đã có các đề xuất sử dụng nén kết hợp các tín hiệu từ các RRH khác nhau để tối ưu hóa độ phân giải nén tốt hơn.



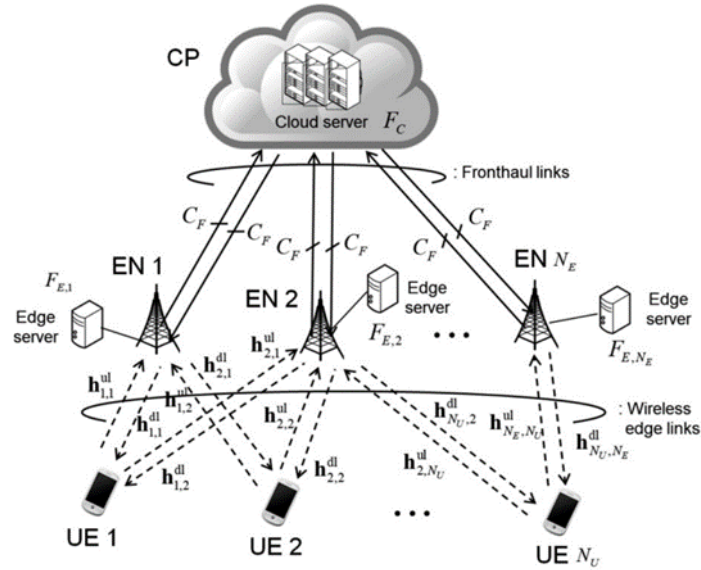
Hình 1.1: Tín hiệu được nén truyền từ RRH đến BBU ở đường uplink [8]



Hình 1.2: Tín hiệu được nén truyền từ BBU pool đến RRH ở đường downlink [8]

Trong chiến lược thứ hai, các vector beamforming được tính toán tại BBU pool được gửi trực tiếp đến các RRH tương ứng. Sau đó, BBU pool chia sẻ dữ liệu của mỗi người dùng trực tiếp với nhóm phục vụ của nó. Tín hiệu được beamformed được tạo ra tại mỗi RRH. Trong chiến lược này, khả năng của fronthaul phụ thuộc vào số lượng người dùng được phục vụ bởi mỗi liên kết. Do đó, các liên kết người dùng-RRH nên được tối ưu hóa dưới các ràng buộc về khả năng truyền tải nội địa phía trước.

Ngoài ra, kết hợp với edge computing cũng là một giải pháp cho việc quá tải của fronthaul, Edge computing có thể giúp giảm yêu cầu về băng thông của đường truyền fronthaul trong kiến trúc C-RAN bằng cách chuyển phần tính toán và xử lý dữ liệu tới các thiết bị và máy chủ edge nằm gần các RRH. Điều này giúp giảm lượng dữ liệu cần truyền tải qua fronthaul, giảm yêu cầu về băng thông của fronthaul. Hơn nữa, edge computing có thể cung cấp khả năng xử lý thời gian thực, điều quan trọng đối với các ứng dụng đòi hỏi xử lý và phản hồi thời gian thực, chẳng hạn như lái xe tự động, tự động hóa công nghiệp và chăm sóc sức khỏe. Bằng cách đưa xử lý gần hơn với nguồn dữ liệu, edge computing có thể cung cấp thời gian phản hồi nhanh hơn và giảm độ trễ của mạng, dẫn đến hiệu suất mạng chung tốt hơn.



Hình 1.3: Kết hợp kiến trúc C-RAN với kỹ thuật Edge computing [9]

#### *Huấn luyện để ước tính CSI quá lớn (Huge Training Overhead for CSI Estimation)*

CSI (Channel State Information) là thông tin về trạng thái kênh liên lạc giữa các nút trong mạng, thông tin này rất quan trọng để các nút trong mạng có thể thiết lập các kết nối tối ưu. Tuy nhiên, để ước tính CSI, các nút cần phải tiến hành việc luyện tập, đây là một quá trình tốn kém và tốn nhiều thời gian. Vấn đề là với mạng C-RAN dày đặc, lượng việc luyện tập này trở nên quá lớn, vượt quá khả năng xử lý của mạng.

Vì vậy, để giảm bớt khối lượng việc luyện tập cần thiết, các nhà nghiên cứu đang tìm cách sử dụng phương pháp CSI một phần (Partial CSI), chỉ cần ước tính CSI cho các liên kết trong cụm của mỗi người dùng (gọi là Intra-cluster CSI) và theo dõi đường mất tín hiệu và sự che khuất ngoài cụm của người dùng đó (gọi là Inter-cluster CSI), rồi gửi các thông tin này trở lại trung tâm BBU pool để thiết kế các kết nối tối ưu. Những tham số này là cần thiết cho thiết kế CoMP tại bể BBU. Thực tế, sự suy giảm tỷ lệ lớn có thể được theo dõi bởi vì nó thay đổi chậm so với CSI tức thời. Thiết kế các vector beamforming dưới trường hợp CSI một phần này là một nhiệm vụ thách thức. Do đó, có ít đóng góp dựa trên CSI một phần.

Để làm rõ, phương pháp thu thập CSI nén được đề xuất để xác định tập hợp CSI tức thời và các thông số mức độ suy hao sóng truyền (Large-scale fading gains). Tuy nhiên, độ phức tạp của nó là cao, nên nó không thể được thực hiện ngay lập tức trong môi trường mật độ cao C-RAN. Gần đây, đã có khung thiết kế để thiết kế truyền tải xanh trong trường hợp CSI một phần. Tuy nhiên, CSI trong cụm được giả định là hoàn hảo trong điều này khó thực hiện trong thực tế. Do đó cần những khung thiết kế cho trường hợp CSI trong cụm không hoàn hảo gồm cả quy trình ước lượng kênh cho CSI trong cụm và thiết kế tối ưu vector beamforming.

### 1.3. Định hướng nghiên cứu và mục tiêu đề án

Ngoài các vấn đề thách thức điển hình đã được đề cập, mô hình Ultra Dense Network (UDN) kết hợp nền tảng C-RAN (CRAP-i-UDN) sẽ còn nảy sinh rất nhiều những thách thức khác cần được giải quyết và tối ưu hóa để giúp nâng cao trải nghiệm người dùng. Một số thách thức khác có thể bao gồm việc cô lập các tài nguyên bị lỗi quản lý di động và mạng, và phân bổ tài nguyên đa mục tiêu, .... Những thách thức này có thể làm nổi bật một số hướng nghiên cứu trước khi đưa vào thực thi. Định hướng nghiên cứu của đề án này tiếp theo sẽ  *nghiên cứu về các cách tối ưu hóa kiến trúc mạng C-RAN*

Trong đề án này, mục tiêu sẽ trình bày về:

1. Xu hướng, tổng quan về mô hình mạng UDN,
2. Tổng quan về kiến trúc C-RAN, triển khai mô hình mạng UDN kết hợp kiến trúc C-RAN (CRAP-i-UDN).
3. Sau đó đặt vấn đề trong mô hình này để giải quyết các thách thức, cụ thể ở đây là dùng kỹ thuật phân cụm để trước hết giảm được tính phức tạp tính toán cao của mô hình mạng này hướng tới việc nâng cao hơn trải nghiệm cho người dùng.
4. Nghiên cứu về các vấn đề tối ưu hóa trong kiến trúc C-RAN

Phân phân tích và mô phỏng sẽ được trình bày ở các phần tiếp.



## Chương 2: Cơ sở lý thuyết của CRAP-i-UDN

### 2.1. Tổng quan và thách thức trong mạng mật độ siêu dày UDN

Khái niệm mạng mật độ siêu cao (UDN)

Để phát triển 5G, những nỗ lực đã được thực hiện trong 3 vấn đề sau: mở rộng băng tần, nâng cao hiệu quả sử dụng băng tần và tăng cường mật độ mạng. Tăng cường mật độ mạng được coi là phương pháp hàng đầu và chiếm ưu thế để giải quyết thách thức dữ liệu. Điều này có thể được đạt được bằng cách triển khai một lượng lớn các cell nhỏ - microcell, picocell, femtocell, nút relay và điểm truy cập WiFi - là các nút truy cập radio có công suất thấp và có vùng phủ nhỏ hơn so với macrocell. Mạng siêu mật độ UDNs là công nghệ chính để đáp ứng các yêu cầu về mật độ lưu lượng truy cập siêu cao trong các kịch bản hiểm nóng. Hệ số tái sử dụng phổ cao có thể được đạt được trong UDN thông qua việc triển khai các thiết bị không dây mật độ cao. Điều này dẫn đến cải thiện khả năng chứa đựng lên đến hàng trăm lần hoặc hơn trong các khu vực điểm nóng



Hình 2.1: UDN với các cell nhỏ được triển khai mật độ cao [2]

5G sẽ cần đáp ứng nhiều loại ứng dụng trong tất cả các khía cạnh của cuộc sống. Các kịch bản điển hình được xác định cho UDN bao gồm văn phòng, khu dân cư dày đặc, khu đô thị dày đặc, khuôn viên trường học, ...

Bảng 2.1: Các kịch bản và dịch vụ ứng dụng.

Kịch bản ứng dụng	Các dịch vụ
Văn Phòng	điện thoại video (video telephony), desk cloud, tải xuống dữ liệu và lưu trữ đám mây.
Khu vực dân cư đông đúc	Điện thoại video, HDTV, thực tế ảo để mua sắm, chơi game trực tuyến, tải xuống dữ liệu, lưu trữ đám mây, tin nhắn OTT và nhà thông minh
Khu đô thị dày đặc	Thực tế ảo, thực tế tăng cường, chơi game trực tuyến, tải xuống dữ liệu, lưu trữ đám mây, tin nhắn OTT, lái xe an toàn giữa các phương tiện.
Khuôn viên trường học	Điện thoại video, HDTV, thực tế ảo để mua sắm, chơi game trực tuyến, tải xuống dữ liệu, lưu trữ đám mây, tin nhắn OTT.
Tụ hợp đông đúc ngoài trời	Phát video, chia sẻ video trực tiếp/ thực tế tăng cường, tải ảnh chất lượng cao lên và tin nhắn OTT.
Sân vận động	Phát video, chia sẻ video trực tiếp/thực tế tăng cường, tải ảnh chất lượng cao lên và tin nhắn OTT.
Khu thương mại	Thực tế tăng cường, tin nhắn OTT và giám sát video.
Ga tàu	Phát video, chơi game trực tuyến và tin nhắn OTT.

Sự khác biệt chính giữa UDN và mạng di động truyền thống nằm ở mật độ điểm truy cập (AP). Trong UDN, bán kính vùng phủ sóng của AP chỉ khoảng 10 m và sẽ có hàng ngàn AP trên 1 km<sup>2</sup>. Nhưng trong mạng di động truyền thống, phạm vi cell là hơn 500 m và thường có ít hơn 3-5 trạm gốc (BS) trong 1 km<sup>2</sup>. Tương ứng, chỉ có một hoặc một số thiết bị đầu cuối được kết nối với một UDN AP, trong khi có hàng

trăm hoặc thậm chí hàng ngàn người dùng đang cư trú và hoạt động trong một macro cell của mạng truyền thống.

Bảng 2.2: So sánh UDN và mạng di động truyền thống

Tiêu chí	UDN	Mạng di động truyền thống
Kịch bản triển khai	Trong nhà, Điểm phát sóng	Vùng phủ rộng
Mật độ AP	Hơn 1000 trạm/km <sup>2</sup>	3~5 trạm/km <sup>2</sup>
Vùng phủ AP	Khoảng 10m	Vài trăm mét hoặc hơn
Loại AP	Small cell, Pico cell, Femto cell, UE relay, Relay	Macro cell
Mạng lõi AP	Lý tưởng/Không lý tưởng Có dây/Không dây	Lý tưởng Có dây
Mật độ người dùng	Siêu cao	Trung bình
Tính di động của người dùng	Thấp	Cao
Mật độ lưu lượng	Cao	Trung bình
Phân bố	Không đồng nhất, không đều	Một lớp, đồng đều
Băng thông	Hàng trăm MHz	Hàng chục MHz
Phổ tần	> 3 GHz cho tới mmWave	< 3 GHz

Một điểm quan trọng khác là chủng loại AP trong UDN khá đa dạng. Trạm di động nhỏ, trạm relay, đầu phát sóng vô tuyến từ xa (RRH) và thiết bị người dùng (UE) đều có thể hoạt động như AP trong UDN. Trong khi đó, trạm macro trong mạng di động truyền thống là điểm truy cập duy nhất. Bên cạnh các tính năng trên, phổ tần số cao hơn và băng thông rộng hơn, triển khai không đồng nhất và bất thường, mạng lõi linh hoạt, tính di động của người dùng thấp hơn cũng là các điểm khác nhau giữa UDN và mạng di động truyền thống.

Nói chung, UDN là một giải pháp mạng không dây mới cho các kịch bản mạng dạng điểm nóng, cần cung cấp thông lượng cao hơn và trải nghiệm người dùng tốt hơn. Trong UDN, mật độ AP có thể ngang bằng hoặc thậm chí cao hơn mật độ người dùng. Các loại AP khác nhau sẽ hợp tác chặt chẽ để đạt được hiệu suất phổ cao hơn, tiêu thụ điện năng thấp hơn và tính di động liên mạch hơn.

Mục tiêu cung cấp mật độ lưu lượng truy cập rất cao và trải nghiệm người dùng tốt hơn trong UDN đặt ra nhiều thách thức mới bao gồm kiến trúc mạng, quản lý tính di động, quản lý nhiễu, v.v. Từ đó, các định hướng kỹ thuật để phát triển hạ tầng mạng cũng được xây dựng để vượt qua những thách thức đó.

#### *Thách thức và định hướng về kiến trúc mạng*

Kiến trúc mạng di động truyền thống được thiết kế cho vùng phủ sóng rộng và liên mạch. Do những sự khác biệt của UDN so với mạng di động truyền thống sẽ có nhiều vấn đề nảy sinh nếu tận dụng kiến trúc mạng di động truyền thống (ví dụ: kiến trúc mạng 4G) cho UDN.

Thứ nhất, quá tải tín hiệu điều khiển và đường truyền dữ liệu phân cấp quá dài: Quá nhiều chức năng như kiểm soát dịch vụ và điều khiển di động được tập trung tại mạng lõi (CN) tại các thực thể của hệ thống mạng 4G như MME - Thực thể quản lý tính di động và PDN-Gateway - Cổng truy nhập mạng dữ liệu gói hay Serving Gateway - Cổng phục vụ. Điều này sẽ là không hiệu quả đối với UDN khi lưu lượng truy cập tăng cao và các AP được triển khai cực kỳ dày đặc vì khi đó số lượng tín hiệu điều khiển sẽ trở nên quá cao, thậm chí quá tải, cũng như đường truyền dữ liệu phải đi qua quá nhiều chặng phân cấp, dẫn tới trễ và giảm độ tin cậy.

Thứ hai, chuyển giao quá thường xuyên: Gắn kết chặt chẽ user plane và control plane trên một giao diện vô tuyến như trong hệ thống mạng 4G sẽ dẫn đến việc chuyển giao thường xuyên khi vùng phủ sóng của AP trong UDN là rất nhỏ. Điều này rất không hiệu quả và kém linh hoạt trong hệ thống mạng có tính không đồng nhất của UDN bao gồm cả vùng phủ sóng của cell vĩ mô và UDN AP.

Thứ ba, phân tán các chức năng mạng: Để hỗ trợ tốt hơn việc quản lý nhiễu và quản lý tài nguyên cho UDN, các chức năng trên mỗi AP phân tán cần được tập trung.

Nói cách khác, các xử lý ở lớp cao hơn như quản lý tài nguyên vô tuyến (RRM), các chức năng quản lý di động được căn phân tán trên mỗi AP một cách độc lập.

Thứ tư, trải nghiệm người dùng tốt hơn: UDN đặt mục tiêu cung cấp quá trình chuyển giao mượt mà với tốc độ dữ liệu rất cao cho mỗi người dùng với mật độ AP cực kỳ dày đặc. Chức năng thu thập và truyền dữ liệu đơn giản của Local Gateway (LGW) trong hệ thống mạng 4G không thể hỗ trợ trải nghiệm người dùng tốt hơn. Cần có nhiều chức năng hơn cho LGW.

Do đó, tìm ra một kiến trúc UDN mới là cần thiết để hỗ trợ triển khai AP với mật độ cao và quản lý mạng một cách linh hoạt. Trong kiến trúc mới này, cần có một trung tâm dịch vụ người dùng tập trung cục bộ để đo lường môi trường vô tuyến của người dùng. Bên cạnh đó, RRM và trung tâm điều khiển dịch vụ người dùng ở gần hơn với người dùng để giải quyết tốt hơn những bước xử lý chung và kiểm soát chất lượng dịch vụ (QoS). Cùng với đó, các chức năng của mạng lõi nên được đơn giản hóa để chỉ cung cấp những dịch vụ ở mức cao cho người dùng.

Một số ví dụ về kiến trúc mạng cho UDN được đưa ra như: Kiến trúc GPP HeNB, kiến trúc tăng cường Small Cell, C-RAN,... Kiến trúc C-RAN sẽ được đề xuất ở đây và phân tích kỹ hơn ở phần tiếp theo.

#### *Thách thức và định hướng quản lý tính di động*

Với phạm vi phủ sóng rất nhỏ và cấu trúc liên kết mạng bất quy tắc, quản lý di động trong UDN khá khác với mạng di động truyền thống. Trong một mạng truyền thống, các location area được cấu hình tĩnh. Và cách chia các LA không liên quan đến bất kỳ yếu tố người dùng nào. Khác với trong UDN, ranh giới giữa các LA trong mạng di động trở nên không rõ ràng. Do đó, chế độ quản lý vị trí sẽ được thay đổi từ quy hoạch AP tĩnh thành kết hợp AP động.

Mạng UDN lấy người dùng làm trung tâm nên sẽ hoạt động dựa theo cách mà người dùng sử dụng dịch vụ, ví dụ các dịch vụ theo yêu cầu hoặc tính di động của người dùng, để có thể cung cấp những dịch vụ tốt nhất trong một môi trường vô tuyến phức tạp. Vì vậy, việc quản lý tính di động cần được tối ưu kết hợp với việc quản lý tài nguyên hay quản lý nhiễu.

Trong các tình huống mà các AP được triển khai cực kỳ dày đặc, phạm vi bao phủ của các AP này cực kỳ nhỏ, tức là chỉ vài mét đến vài chục mét, theo phương pháp quyết định chuyển giao truyền thống, việc chuyển giao sẽ có thể xảy ra thường xuyên và khả năng gián đoạn về trải nghiệm dữ liệu tốc độ cao của người dùng có thể tăng lên. Thêm vào đó, từ quan điểm của nhà mạng, tình huống này sẽ dẫn tới một sự quá tải hơn nữa của tín hiệu điều khiển.

#### *Thách thức và định hướng quản lý nhiễu*

Số lượng lớn các AP có thể mang lại thông lượng cao hơn nhiều và trải nghiệm người dùng tốt hơn, nhưng cũng có thể dẫn đến các vấn đề về nhiễu. Việc quản lý nhiễu có thể ảnh hưởng trực tiếp đến hiệu năng của hệ thống. Cùng với việc ghép kênh tài nguyên để phục vụ được nhiều lượng truy cập hơn, nhiễu cũng tăng lên và trở nên phức tạp hơn đối với mạng di động truyền thống. Chúng ta cần giải quyết các vấn đề sau:

Môi trường vô tuyến siêu dày đặc dẫn đến nhiều nguồn gây nhiễu hơn. Ví dụ, trong đám đông trên tàu điện ngầm, rất nhiều thiết bị đầu cuối và AP cùng tồn tại. Tín hiệu có thể có nhiều đường phản xạ và phân tán hơn. Sau đó, công suất truyền tải sẽ tăng lên, điều này sẽ tạo ra nhiễu nhiều hơn trước đây. Giảm nhiễu và tăng hiệu suất sử dụng tài nguyên trở thành hai phần mâu thuẫn lẫn nhau và chúng ta cần phải tìm điểm cân bằng thích hợp. Các thông số hiện tại để đánh giá tác động của nhiễu như ngưỡng nhiễu có thể không phản ánh hiệu suất tổng thể của mạng. Vì vậy, cần sử dụng các thông số phù hợp hơn để đưa ra chỉ báo tốt hơn giữa kết quả của việc quản lý nhiễu và thông lượng, hiệu suất năng lượng và các thông số khác của hệ thống. Vì vậy, chúng ta nên thiết lập các mô hình can thiệp thích hợp, phân tích kịch bản truyền dẫn không dây điển hình, và sau đó đề xuất phương pháp kiểm soát nhiễu hiệu quả cho UDN.

#### *Thách thức và định hướng về tính linh hoạt của hệ thống mạng*

Với mật độ cực cao và việc triển khai phức tạp mạng không đồng nhất, việc lập kế hoạch và tối ưu hóa mạng sẽ trở nên khó khăn đối với UDN. Điều quan trọng là tăng cường hơn nữa Mạng tự tổ chức (SON) để có được một hệ thống mạng linh hoạt.

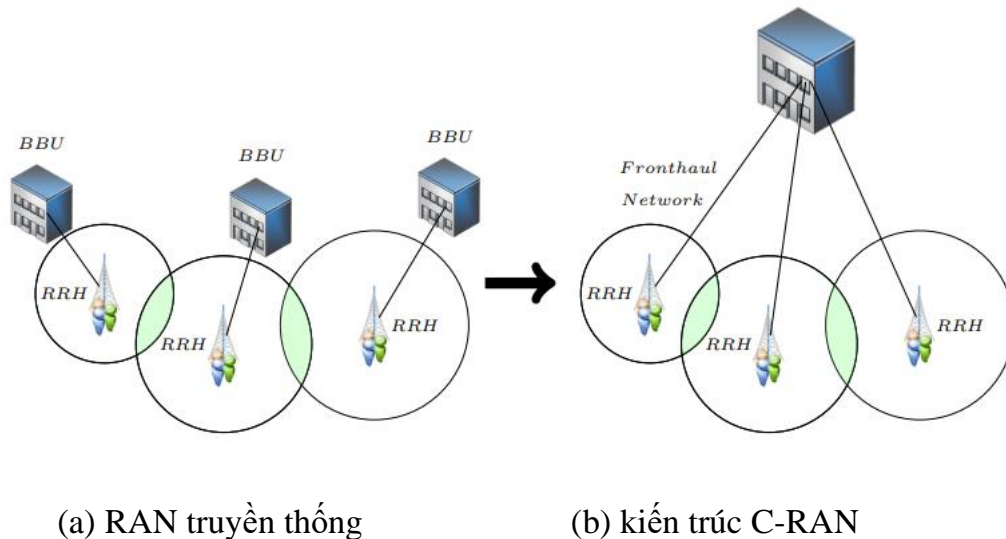
Một số lượng lớn các AP trong UDN làm cho việc tự cấu hình, tự tối ưu và tự sửa chữa trở nên phức tạp hơn. Thông lượng cực cao, độ trễ cực thấp, độ tin cậy cực cao, lượng kết nối lớn cần phải được cung cấp trong UDN. Ở một khía cạnh khác, UDN là một mạng rất phức tạp bao gồm việc hỗ trợ cả các kịch bản trong nhà và ngoài trời, với các mạng trực liên kết các AP được triển khai một cách lý tưởng lẫn không lý tưởng. Từ quan điểm của các công nghệ truy cập vô tuyến, sẽ tồn tại công nghệ truy cập mới trong 5G, công nghệ truy cập LTE, công nghệ truy cập WLAN làm việc cùng nhau. Vì vậy, kiến trúc mạng linh hoạt với cảm biến mạng thông minh và tối ưu hóa mạng là rất quan trọng đối với mạng linh hoạt và tăng hiệu quả phổ trong UDN.

Dựa trên phân tích trên, xây dựng một kiến trúc mạng mới cho UDN là một hướng đi quan trọng. Các công nghệ chủ chốt hướng tới mạng linh hoạt, tự tái cấu trúc, phối hợp nhiều RAT, quản lý nhiễu nâng cao, quản lý di động nâng cao và quản lý tài nguyên vô tuyến cũng rất cần thiết cho UDN.

Ngoài ra, do các kịch bản vô tuyến khác nhau và tần số cao hơn, băng thông rộng hơn, các công nghệ truyền dẫn mới và thiết kế giao diện vô tuyến mới cũng rất cần thiết để nâng cao hơn nữa hiệu năng hệ thống cho UDN. Truyền thông sóng milimet là một hướng rất hấp dẫn đối với UDN khi có thể cung cấp trải nghiệm người dùng lên tới hàng Gbps và thông lượng hàng chục Gbps trên mỗi AP. Và nó cũng rất thích hợp cho việc cải tiến mạng trực không dây cho các UDN AP.

## **2.2. Tổng quan kiến trúc C-RAN và tiềm năng triển khai trong mạng UDN**

Các mạng di động thế hệ hiện tại đang sử dụng mạng truy cập sóng vô tuyến (RAN) truyền thống, trong đó các chức năng xử lý sóng radio và baseband được đặt trong cùng một trạm cơ sở. Trong thực tế, như được mô tả trong Hình 5, trạm cơ sở truyền thống bao gồm hai thành phần, ăng-ten (RRH) và BBU (trung tâm dữ liệu), được đặt cùng vị trí tại trang trại lớn (eNodeB). Tuy nhiên, các mạng này không còn có thể cung cấp tốc độ truyền dữ liệu cao, đáp ứng yêu cầu thời gian trễ mạnh và đảm bảo QoS cao cho các yêu cầu của người dùng cuối. Để đạt được những mục tiêu này, C-RAN đã được đề xuất như một kiến trúc tiềm năng cho thế hệ tiếp theo của mạng di động (5G) để xử lý các yêu cầu dịch vụ đa dạng. Khái niệm chính của C-RAN là tách BBU khỏi các ăng-ten và gom tài nguyên tính toán vào trung tâm dữ liệu tập trung, tức là các nhóm BBU. Các nhóm tính toán BBU sẽ được chia sẻ giữa nhiều trạm cơ sở để đạt được lợi ích về sử dụng tài nguyên và tiết kiệm chi phí mạng. Hình 5 mô tả sự khác biệt chính giữa RAN truyền thống và kiến trúc C-RAN về khái niệm tập trung tài nguyên và tập trung BBU.



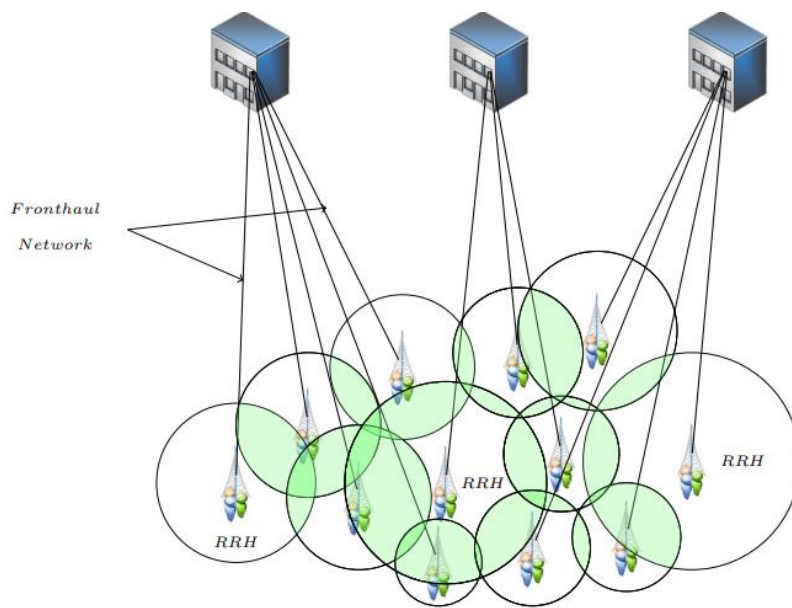
Hình 2.2: Kiến trúc RAN truyền thống và kiến trúc C-RAN [4]

Như đã đề cập ở trên, kiến trúc C-RAN được đề xuất bao gồm việc tập trung xử lý baseband của các yêu cầu ăng-ten từ những điểm truy cập di động khác nhau vào các



nhóm BBU khi các RRH được kết nối với trung tâm dữ liệu tập trung thông qua mạng fronthaul.

- **RRHs (Remote Radio Heads):** được đặt tại các điểm truy cập di động và chúng chuyển tiếp các tín hiệu baseband nhận được từ các thiết bị User Equipment (UEs) đến các nhóm BBU để xử lý tập trung trong uplink, đồng thời truyền các tín hiệu tần số vô tuyến (RF) đến các UEs trong downlink. RRHs thực hiện các chức năng radio bao gồm chuyển đổi RF, khuếch đại, lọc, chuyển đổi tương tự sang số và chuyển đổi số sang tương tự.
- **BBU pool hoặc trung tâm dữ liệu tập trung:** là một vị trí tập trung các tài nguyên tính toán và xử lý được chia sẻ giữa nhiều điểm truy cập di động. Thực tế, mỗi BBU pool có thể phục vụ từ 10 đến 1000 RRHs và nó bao gồm việc tập trung xử lý các yêu cầu baseband của các ăng-ten và tối ưu hóa phân bổ tài nguyên tính toán
- **Fronthaul Networks:** là một tập hợp các liên kết truyền thông giữa RRHs và BBU pool. Lưu lượng fronthaul trao đổi giữa các ăng-ten và trung tâm dữ liệu tập trung có thể được truyền qua các giao thức thông thường bao gồm OBSAI và CPRI, đây là giao thức được sử dụng rộng rãi nhất trong các mạng di động. Fronthaul có thể được triển khai bằng các công nghệ khác nhau như truyền thông sợi quang, truyền thông không dây tiêu chuẩn hoặc truyền thông sóng millimeter (mmWave)



Hình 2.3: Kiến trúc C-RAN và thành phần [4]

*Lợi ích của kiến trúc C-RAN và tiềm năng cho mạng UDN*

- **Tiết kiệm chi phí trong CAPEX và OPEX:** việc tập trung tài nguyên tính toán trong C-RAN cho phép sử dụng hiệu quả tài nguyên BBU bằng cách giảm tổng số BBUs cần thiết để đáp ứng nhu cầu của người dùng cuối. Điều này dẫn đến tiêu thụ năng lượng thấp hơn và tiềm năng giảm chi phí trong CAPEX và OPEX. C-RAN có thể giảm CAPEX 15% và OPEX 50% so với một mạng di động truyền thống.
- **Cải thiện khả năng và phủ sóng:** trong kiến trúc C-RAN, các trung tâm dữ liệu tập trung (BBU pools) dự kiến sẽ chứa nhiều yêu cầu của người dùng cuối đến từ các điểm truy cập di động khác nhau. Mỗi BBU pool nên hỗ trợ 10 đến 1000 trạm cơ sở, giúp cải thiện khả năng mạng bằng cách bao phủ một khu vực lớn hơn và phục vụ nhiều người dùng hơn so với các trạm cơ sở truyền thống khi đảm bảo chất lượng dịch vụ cao.
- **Tăng cường sử dụng tài nguyên:** vì trong C-RAN, xử lý băng thông của nhiều tế bào được thực hiện trong các BBU pool tập trung, việc chia sẻ tài nguyên trở nên khả thi và do đó phân bổ tài nguyên có thể linh hoạt hơn và theo yêu cầu, khác với các mạng truyền thống. Điều này cho phép cải thiện hiệu quả sử dụng tài nguyên mạng

- **Giảm nhiễu giữa các tế bào:** trong C-RAN, việc xử lý tập trung cho phép thực hiện dễ dàng các thuật toán xử lý và lập lịch phối hợp, giúp giảm nhiễu giữa các tế bào và cải thiện hiệu suất phổ. Trong thực tế, các kỹ thuật quản lý nhiễu hiệu quả như Coordinated Multi-Point (CoMP) và Inter-cell Interference Coordination (ICIC) có thể được thực hiện dễ dàng trong BBU pool để tối ưu hóa truyền thông từ nhiều cell đến nhiều BBU.
- **Linh hoạt và có khả năng mở rộng:** Với việc tập trung các nguồn lực tính toán tại một địa điểm chung, các nhà điều hành mạng sẽ nhanh chóng triển khai các anten mới (RRHs) và kết nối chúng với BBU pool để mở rộng vùng phục vụ và nâng cấp hạ tầng mạng. Điều này sẽ cải thiện tính mở rộng và tính linh hoạt của mạng và sẽ thuận tiện cho việc bảo trì mạng.

### 2.3. Phát triển và triển khai mô hình mạng CRAP-i-UDN

Mô hình mạng CRAP-i-UDN (Cloud Radio Access Platform - inspired Ultra Dense Networks) là một giải pháp kết hợp giữa hai công nghệ mạng, đó là CRAN (Cloud Radio Access Network) và UDN (Ultra Dense Network). Mô hình này có thể cung cấp khả năng truy cập mạng tốt hơn và đáp ứng được nhu cầu tăng trưởng số lượng người dùng và dữ liệu trong tương lai. Để phát triển mô hình này, cần liệt kê qua những yêu cầu chính trong từng thành phần của mô hình và các khó khăn khi triển khai.

#### 2.3.1 Tài nguyên sợi quang do việc tập trung hóa

Front-haul được định nghĩa là liên kết giữa BBUs và RRUs. Các ví dụ về giao thức front-haul bao gồm Common Public Radio Interface (CPRI) và Open Base Station Architecture Initiative (OBSAI). Tập trung hóa (centralization) là bước đầu tiên quan trọng để thực hiện tất cả các tính năng khác của C-RAN. Centralization tổng hợp các BBU khác nhau (thường là vài chục hoặc vài trăm) vào một trung tâm với các cơ sở

chung. Thách thức chính đối với centralization là nó yêu cầu một số lượng lớn tài nguyên sợi quang nếu sử dụng giải pháp sợi quang tối, tức là kết nối sợi trực tiếp.

Để vượt qua bất lợi về sợi quang trong việc triển khai centralization, các giải pháp khác nhau được đưa ra. Một số giải pháp đã phù hợp để triển khai trong khi những giải pháp khác vẫn đang ở giai đoạn phát triển ban đầu. Các kỹ thuật nén là bước đầu tiên được thực hiện để giảm tiêu thụ sợi quang. Với ý tưởng liên kết không dây, các kỹ thuật Edge computing, Fog computing cũng được đề xuất để giải quyết sự quá tải ở Front haul.

### **2.3.2 Ứng dụng điện toán đám mây và công nghệ ảo**

Tính năng cốt lõi của C-RAN là đám mây hóa tài nguyên, trong đó tài nguyên xử lý có thể được phân bổ động, liên tục để tạo thành thực thể BBU mềm. Với các nền tảng độc quyền và đóng của các nhà cung cấp hiện tại, việc phát triển một nền tảng BBU mới dựa trên công nghệ ảo hóa được tìm thấy trong các trung tâm dữ liệu hiện đại là thuận lợi.

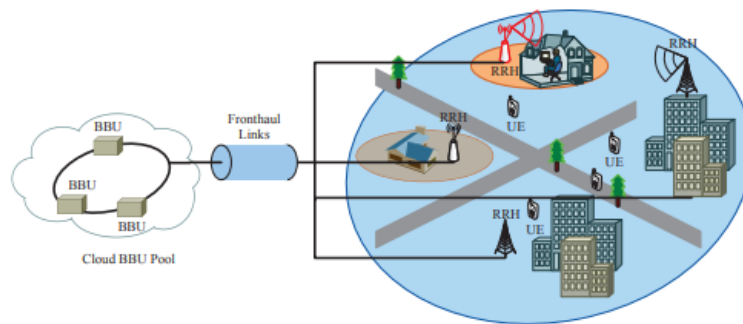
Một phương pháp phù hợp của ảo hóa mạng là sử dụng ảo hóa chức năng mạng NFV (Network Function Virtualization), tức là tập trung nhiều chức năng mạng khác nhau vào các máy chủ, switch và bộ nhớ tiêu chuẩn, được đặt tại các trung tâm dữ liệu, các nút mạng và tại nơi sử dụng cuối. Việc sử dụng công nghệ ảo hóa giúp cho việc triển khai và quản lý mạng trở nên linh hoạt và hiệu quả hơn, từ đó giảm chi phí đầu tư và vận hành mạng cho các nhà cung cấp dịch vụ viễn thông.

### **2.3.3 Thách thức triển khai C-RAN dưới kịch bản UDN**

Triển khai small cell là một trong những xu hướng công nghệ triển vọng nhất để đối phó với nhu cầu tăng cao của tốc độ truyền dữ liệu trong các mạng di động trong tương lai, bao gồm các loại small cell khác nhau như picocell, relay, Machine-to-Machine và cả Device-to-Device.

Do sự tăng cường của small cell, khoảng cách giữa UE và điểm truy cập giảm đáng kể, tức chỉ còn vài mươi mét, do đó một UE sẽ rơi vào một số vùng phủ của small cell, và tất cả những BS small cell đó có thể nhận tín hiệu UL từ UE hoặc phục vụ UE DL. Dựa trên tính chất này của UDN, một số cơ chế hợp tác có thể cung cấp trải nghiệm tốt hơn cho người dùng.

Trong mô hình UDN, tần suất thực hiện handover sẽ tăng lên, gây tải nặng tín hiệu và làm ảnh hưởng đến trải nghiệm người dùng. Ngoài ra, sự giảm kích thước của small cell và tăng cường nhiễu giữa các cell cũng có thể gây ra các vấn đề liên quan đến việc handover không thành công và hiện tượng ping-pong. Để giải quyết vấn đề này, trong trường hợp handover giữa các eNB, C-RAN có thể giảm thiểu độ trễ handover và giảm thiểu nguy cơ mất kết nối với cell đang phục vụ khi UE vẫn đang chờ lệnh handover, từ đó giảm tỷ lệ thất bại handover.



Hình 2.4: Kiến trúc C-RAN dưới kịch bản triển khai UDN [7]

Trong triển khai mạng UDN, có một hiện tượng rõ rệt là lưu lượng dữ liệu biến đổi theo thời gian. Trong kiến trúc C-RAN, bể BBU có thể đánh giá tình hình hiện tại của toàn bộ mạng với các yếu tố như nhu cầu lưu lượng, hiệu suất hệ thống, mức hiệu suất mục tiêu, tính khẩn cấp của người dùng, tình trạng năng lượng vào thời điểm của ngày và nhiều yếu tố tương tự khác để quyết định xem có cần kích hoạt small cell RRH phù hợp hay không. Ví dụ, nếu một khu vực cụ thể của Marco cell chứa một số lượng lớn người dùng và phần còn lại của khu vực cell có yêu cầu nhu cầu lưu lượng dữ liệu thấp, bể BBU có thể giảm tải người dùng trong khu vực tắc nghẽn bằng cách kích hoạt các RRH tương ứng và xử lý phần còn lại của người dùng bằng Marco RRH.

Để phát triển và triển khai mô hình mạng CRAP-i-UDN, nhìn chung cần thực hiện các bước quan trọng sau:

1) *Đánh giá và phân tích yêu cầu*

Cần đánh giá yêu cầu của hệ thống và cải tiến phù hợp với mô hình mạng CRAP-i-UDN. Phân tích các yếu tố khác nhau, chẳng hạn như độ phủ sóng, lưu lượng truy cập, sự thay đổi trong tài nguyên mạng, và độ trễ, sẽ giúp xác định yêu cầu thiết kế cho mô hình mạng. Đối với mô hình mạng CRAP-i-UDN, việc đánh giá yêu cầu còn bao gồm xác định số lượng trạm cơ sở cần thiết, địa điểm lắp đặt trạm và các yếu tố khác liên quan đến kiến trúc mạng.

2) *Thiết kế hệ thống*

Sau khi đã đánh giá các yêu cầu thiết kế hệ thống dựa trên các yêu cầu đã xác định. Điều này bao gồm thiết kế kiến trúc hệ thống, lựa chọn công nghệ mạng và các giao thức liên kết, phân phối tài nguyên và bố trí trạm cơ sở. Khi thiết kế kiến trúc, cần xác định các khu vực phủ sóng, tần số sóng, số lượng anten và các khu vực có mật độ truy cập mạng cao. Ngoài ra, cần thiết kế các kết nối giữa trạm cơ sở và các máy chủ điều khiển, và thiết kế các thuật toán quản lý tài nguyên.

3) *Kiểm tra và thử nghiệm*

Sau khi hoàn thành thiết kế, cần kiểm tra và thử nghiệm hệ thống để đảm bảo tính khả thi và hiệu quả của nó. Việc kiểm tra bao gồm kiểm tra tính ổn định, khả năng truy cập mạng, độ trễ, và các yếu tố khác liên quan đến hiệu suất mạng.

4) *Triển khai hệ thống*

Sau khi đã hoàn tất kiểm tra và thử nghiệm, cần triển khai hệ thống mạng CRAP-i-UDN. Việc triển khai bao gồm cài đặt và cấu hình hệ thống, bảo trì và vận hành, cũng như giám sát hiệu suất mạng.

## **2.4. Xu hướng tập trung vào chất lượng người dùng và mục tiêu nâng cao trải nghiệm người dùng trong CRAP-i-UDN**

Trong mạng di động ngày nay, xu hướng tập trung vào chất lượng người dùng và mục tiêu nâng cao trải nghiệm người dùng đang trở thành một vấn đề quan trọng và được đánh giá cao bởi các nhà cung cấp dịch vụ viễn thông. Điều này đặc biệt quan trọng trong bối cảnh mạng di động đang phát triển với tốc độ nhanh chóng và có sự cạnh tranh khốc liệt giữa các nhà cung cấp dịch vụ. Mô hình mạng CRAP-i-UDN được đề xuất triển khai nhằm mục tiêu tập trung hơn vào từng người dùng và nâng cao hơn về trải nghiệm sử dụng, bằng những cách sau đây:

1) *Tối ưu hóa mạng để đạt được độ phủ sóng tốt hơn*

Để đảm bảo trải nghiệm người dùng tốt nhất, mạng CRAP-i-UDN cần đạt được độ phủ sóng tốt hơn. Điều này có thể được đạt được bằng cách tối ưu hóa cấu trúc mạng và địa điểm đặt các thiết bị UDN để đảm bảo phủ sóng tốt hơn.

2) *Tăng cường khả năng xử lý của C-RAN*

C-RAN được sử dụng để xử lý dữ liệu từ các thiết bị UDN. Do đó, nâng cao khả năng xử lý của C-RAN là một yếu tố quan trọng để đảm bảo trải nghiệm người dùng tốt nhất. Có thể sử dụng các công nghệ mới như phân phối tính toán, 5G Core, IoT... để nâng cao khả năng xử lý và giảm độ trễ trong quá trình truyền dữ liệu.

3) *Tăng tốc độ truyền dữ liệu*

Tốc độ truyền dữ liệu là một yếu tố quan trọng ảnh hưởng đến trải nghiệm người dùng. Cần tối ưu hóa mạng để đạt được tốc độ truyền dữ liệu cao hơn và đáp ứng được yêu cầu của người dùng. Các nhà điều hành mạng có thể tối ưu hóa mạng bằng cách sử dụng các công nghệ mới như OFDMA, MIMO, MU-MIMO, Massive MIMO, Beamforming... để tăng tốc độ truyền dữ liệu và đáp ứng được yêu cầu của người dùng.

4) *Cải thiện độ trễ*

Độ trễ là thời gian mà dữ liệu mất để đi từ thiết bị của người dùng đến đích. Độ trễ thấp là yếu tố quan trọng để đảm bảo trải nghiệm người dùng tốt nhất. Cần tối ưu hóa mạng để giảm độ trễ và tăng tốc độ phản hồi của hệ thống. Để cải thiện độ trễ, các nhà điều hành mạng có thể sử dụng các công nghệ mới như Network Slicing, edge

computing, IoT... để giảm thời gian truyền dữ liệu và tăng tốc độ phản hồi của hệ thống.

#### 5) *Sử dụng công nghệ mới*

Sử dụng các công nghệ mới như AI, machine learning, edge computing, IoT... có thể giúp tối ưu hóa mạng và đảm bảo trải nghiệm người dùng tốt nhất.

## **Chương 3: Đặt vấn đề và xây dựng mô hình hệ thống CRAP-i-UDN**

### **2.1. Phân cụm để giảm tính phức tạp trong UDN, kiến trúc C-RAN**

Trong một mạng C-RAN dày đặc, bể BBU thường phải hỗ trợ một số lượng lớn các RRH. Do đó, có thể có một số lượng đáng kể các biến cần tối ưu, chẳng hạn như vector beamforming, có thể trở nên quá phức tạp ngay cả trong bối cảnh tính toán đám mây. Để giải quyết sự phức tạp này, kỹ thuật phổ biến nhất được sử dụng là sử dụng kỹ thuật phân cụm.

Kỹ thuật phân cụm bao gồm chia bể BBU thành các cụm nhỏ hơn, với mỗi cụm phục vụ một tập con của các RRH. Có thể áp dụng kỹ thuật phân cụm không chỉ ở phía BBU pool mà còn ở phía RRH để giảm độ phức tạp của hệ thống UDN. Theo phương pháp này, các RRH được phân thành các cụm nhỏ hơn để giảm số lượng biến cần tối ưu hóa và cải thiện khả năng xử lý của hệ thống. Hai kỹ thuật phân cụm cho các RRH là disjoint clustering và user-centric clustering sẽ được trình bày ở phần tiếp theo.

Có nhiều thuật toán khác nhau để phân cụm [10], một số phổ biến bao gồm:

- **K-means clustering:** là một trong những thuật toán phân cụm phổ biến nhất trong machine learning. Thuật toán này tìm ra k trung tâm ban đầu và sau đó cố gắng phân chia các điểm dữ liệu thành k cụm dựa trên khoảng cách đến các trung tâm đó.



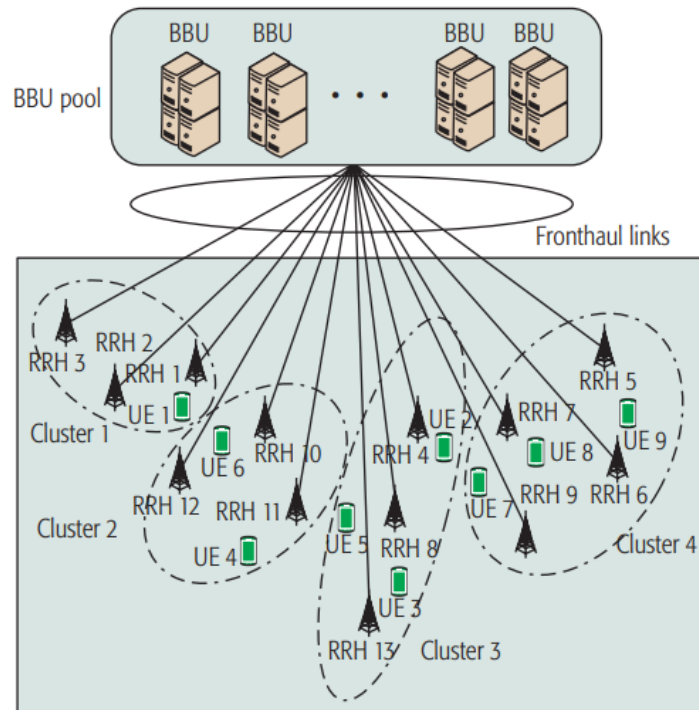
- **Spectral clustering:** là một phương pháp phân cụm dựa trên phân tích giá trị riêng của ma trận đồ thị. Thuật toán này giúp phân cụm các nút dựa trên sự tương đồng về mặt địa lý hoặc các thông số khác.
- **Hierarchical clustering:** là một thuật toán phân cụm dựa trên việc xây dựng một cấu trúc cây, với các nhánh biểu thị cho các cụm con của cụm cha. Thuật toán này có thể được thực hiện dưới hai hình thức: phân cấp hoặc gộp cụm.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** là một thuật toán phân cụm dựa trên mật độ. Nó tìm kiếm các vùng trong không gian đa chiều có mật độ dày đặc của các điểm dữ liệu và phân chia chúng thành các cụm.

Các thuật toán trên đều có ưu điểm và hạn chế riêng. Việc chọn thuật toán phù hợp để phân cụm trong mạng UDN với kiến trúc C-RAN phụ thuộc vào nhiều yếu tố, bao gồm kích thước của mạng, số lượng thiết bị, khoảng cách địa lý và các thông số kỹ thuật khác.

## 2.2. Kỹ thuật phân cụm Disjoint clustering và User-centric clustering trong mô hình CRAP-i-UDN

### 2.2.1. Kỹ thuật phân cụm Disjoint clustering

Disjoint clustering (phân cụm không giao nhau) là phương pháp chia mạng UDN thành các cụm không chồng lên nhau. Mỗi cụm được phục vụ bởi một tập con của các Remote Radio Head (RRH), và các cụm không chia sẻ RRH. Vì vậy, khi một RRH được phục vụ bởi một trung tâm xử lý cụm, nó không thể được phục vụ bởi bất kỳ trung tâm xử lý cụm nào khác.



Hình 3.1: Disjoint clustering [1]

*Ưu điểm:*

Giảm độ phức tạp tính toán của hệ thống bằng cách giới hạn số lượng biến tối ưu hóa vào bất kỳ thời điểm nào.

Phân phối công việc tính toán hiệu quả hơn bằng cách phân phối tải công việc trên nhiều cụm.

Đễ dàng triển khai và thực hiện trong môi trường mạng.

*Hạn chế:*

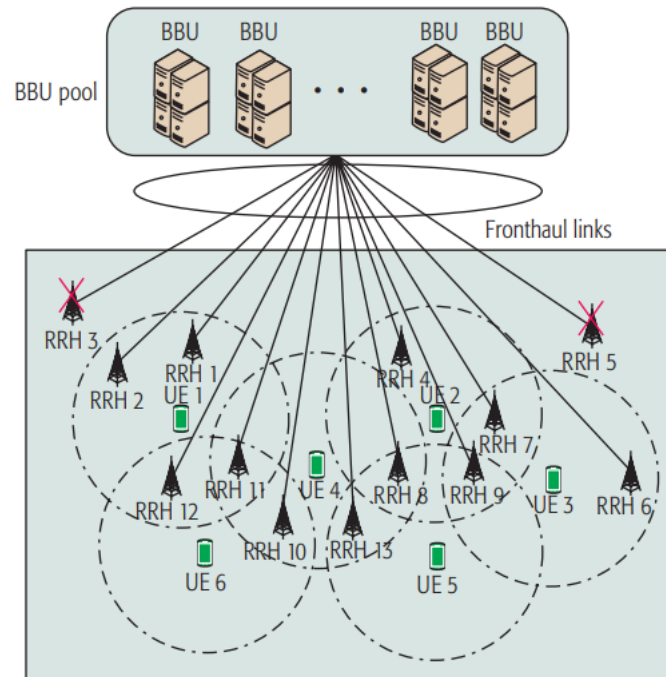
Các người dùng ở rìa của cụm vẫn phải chịu đựng nhiều can tác liên cụm cao

Không tối ưu hóa hiệu quả cho các thiết bị gần trung tâm xử lý hơn, do các thiết bị này cũng được kết nối với các trung tâm xử lý xa hơn.

Không hiệu quả trong các trường hợp có nhiều thiết bị RRH một cách ngẫu nhiên.

### 2.2.2. Kỹ thuật phân cụm User-centric clustering

Là phương pháp phân chia mạng UDN dựa trên sự tập trung vào người dùng. Trong phương pháp này, mỗi cụm được hình thành bởi các RRH phục vụ cùng một nhóm người dùng, bất kể vị trí của các RRH. Vì vậy, các cụm có thể chồng lên nhau và chia sẻ một số RRH.



Hình 3.2: user-centric clustering [1]

#### *Ưu điểm:*

- Các vấn đề về can nhiễu của người dùng ở rìa cụm được giảm.
- Tối ưu hóa hiệu quả cho các thiết bị gần trung tâm xử lý hơn.
- Tối ưu hóa hiệu quả cho các ứng dụng cụ thể của người dùng.
- Giảm độ trễ trong quá trình truyền dữ liệu và tăng độ ổn định của hệ thống.

#### *Hạn chế:*

- Đòi hỏi phải có thông tin cụ thể về các yêu cầu của người dùng và các ứng dụng của họ để phân cụm hiệu quả.
- Khó khăn trong việc triển khai và thực hiện trong môi trường mạng.
- Có thể làm tăng độ phức tạp tính toán của hệ thống.

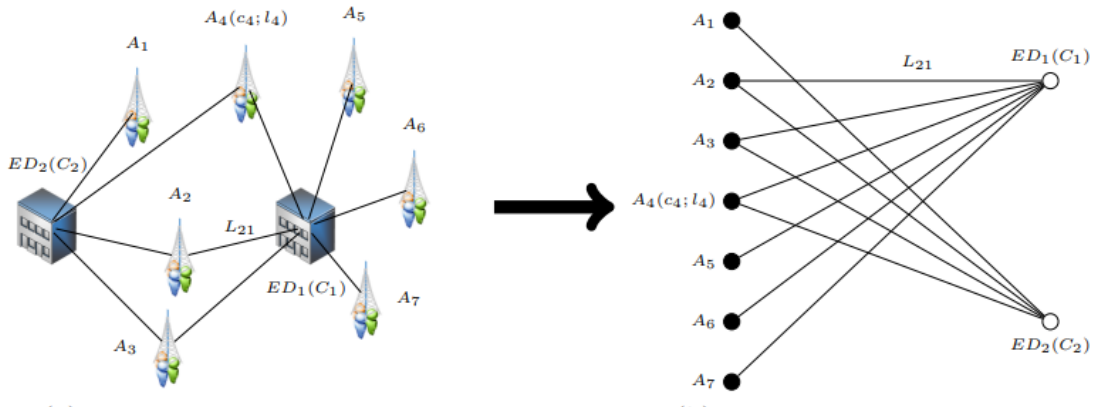
## **Chương 4: Nghiên cứu các vấn đề tối ưu hóa kiến trúc mạng**

### **4.1. Phân bổ tài nguyên trong C-RAN**

Việc triển khai kiến trúc C-RAN (Cloud Radio Access Network), trong đó hạ tầng được chia sẻ trên nhiều trạm thu phát sóng, được kỳ vọng sẽ giảm cả vốn đầu tư và chi phí vận hành (CAPEX và OPEX) cũng như tăng cường hiệu quả sử dụng tài nguyên. Những lợi ích này chỉ có thể đạt được bằng cách phân bổ các yêu cầu anten (RRHs) một cách hiệu quả đến các trung tâm dữ liệu tập trung (BBU pool) khi yêu cầu về độ trễ và xử lý được đáp ứng. Do đó, để giải quyết vấn đề phân bổ tài nguyên bị ràng buộc này, các nhà điều hành mạng đang nghiên cứu các thuật toán mới để xác định các chiến lược tốt nhất để phân bổ RRHs cho BBUs (còn được gọi là vấn đề phân bổ RRH-BBU). Các thuật toán đề xuất sẽ phân bổ đồng thời tài nguyên xử lý và tài nguyên vô tuyến cho các yêu cầu anten, tận dụng việc tập trung các tài nguyên tính toán trong các trung tâm dữ liệu đầu mút chung. Ánh xạ tối ưu giữa RRHs và BBUs, tức là phân bổ RRH-BBU tối ưu, được đặt khi đồng thời tối thiểu hóa độ trễ truyền trên mạng fronthaul và tiêu thụ tài nguyên tính toán.

Trong kịch bản này, để giải quyết vấn đề phân bổ RRH-BBU khi đồng thời đáp ứng các yêu cầu độ trễ lớn trên mạng fronthaul và ràng buộc khả năng giới hạn của các trung tâm dữ liệu đầu mút. Công việc tối ưu hóa được biểu diễn bằng một công thức chính xác trước khi nghiên cứu các thuật toán heuristics mà hội tụ đến các giải pháp gần tối ưu trong thời gian chấp nhận được.

#### 4.1.1. Mô hình hệ thống



a) Ví dụ về một lưới mạng

b) xây dựng sơ đồ đồ bipartite

Hình 4.1: Mô hình hệ thống cho vấn đề phân bổ tài nguyên [4]

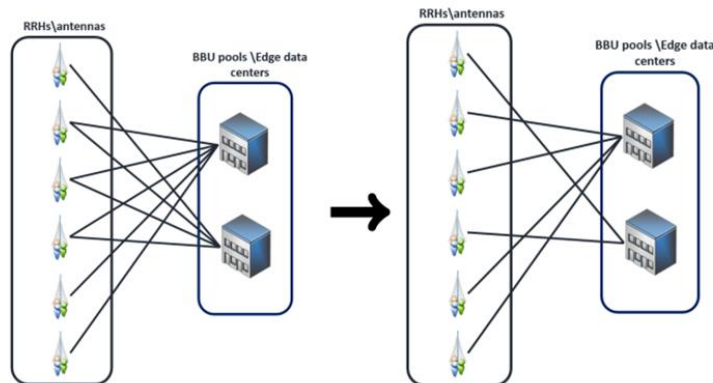
$C1$  (tương ứng với  $C2$ ): tổng số lõi CPU có sẵn trong bể BBU  $ED1$  (tương ứng với  $ED2$ )

$c4$ : số lõi CPU được yêu cầu để xử lý các yêu cầu của ăng-ten  $A4$ .

$l4$ : độ trễ kỳ vọng khi xử lý các yêu cầu của ăng-ten  $A4$ .

$L21$ : độ trễ trong truyền trên kênh fronthaul giữa ăng-ten  $A2$  và pool BBU  $ED1$ .

Đây là một ví dụ về mạng C-RAN bao gồm 6 RRH (antennas), 2 trung tâm dữ liệu (BBU pools) và một mạng fronthaul được đại diện bởi một tập các liên kết truyền thông. Vấn đề phân bổ tài nguyên bị ràng buộc bao gồm việc xác định các chiến lược tối ưu để gán các yêu cầu của các anten cho các trung tâm dữ liệu có sẵn với các yêu cầu xử lý và độ trễ nghiêm ngặt.



Hình 4.2: Ví dụ giải pháp của bài toán phân bổ tài nguyên có ràng buộc [4]

Mục đích của bài toán phân bổ RRH-BBU là tìm một phép ánh xạ tối ưu của tất cả các yêu cầu về các ăng-ten trên các trung tâm dữ liệu khi đáp ứng các yêu cầu xử lý và độ trễ. Để đạt được mục tiêu này, chúng tôi sẽ thảo luận về một phương pháp chính xác và các thuật toán heuristic để đạt được các giải pháp tối ưu và gần tối ưu, tương ứng.

#### 4.1.2. Bài toán tối ưu và giải thuật

Để làm rõ, dưới đây là tổng hợp các biến và tham số sẽ được sử dụng để mô hình hóa vấn đề phân bổ tài nguyên bị ràng buộc.

Vấn đề về phân bổ RRH-BBU: biến và thông số

$G = (I \cup J, E)$	: weighted bipartite graph
$I$	: Tập các ăng-ten/RRHs
$J$	: Tập các trung tâm dữ liệu/BBU pools
$E$	: Tập đường link giao tiếp giữa $I$ và $J$
$d_{ij}$	: Khoảng cách giữa một ăng-ten $i (x_i, y_i)$ và trung tâm dữ liệu $j (x_j, y_j)$
$c_i$	: Số lõi CPU được yêu cầu để xử lý các yêu cầu tổng hợp của anten $i$
$C_j$	: Tài nguyên tính toán (lõi CPU) tại cái trung tâm dữ liệu $J$
$l_i$	: Thời gian trễ kỳ vọng để xử lý các yêu cầu tổng hợp của anten $i$
$L_{ij}$	: Độ trễ truyền trên liên lạc giữa một anten $i$ và một trung tâm dữ liệu $j$
$x_{ij}$	: Biến quyết định (biến nhị phân)

	$x_{ij} = \begin{cases} 1, & \text{nếu ăn – ten } i \in I \text{ được phân bổ} \\ & \text{cho trung tâm dữ liệu } j \in J \\ 0, & \text{ngược lại} \end{cases}$
$y_j$	<p>: Biến quyết định (biến nhị phân)</p> $y_j = \begin{cases} 1, & \text{nếu trung tâm dữ liệu } j \text{ được dùng} \\ 0, & \text{ngược lại} \end{cases}$

*Bài toán tối ưu:*

Để đạt được những lợi ích trong việc sử dụng tài nguyên khi số lượng trung tâm dữ liệu sử dụng và độ trễ fronthaul được tối thiểu hóa.

$$\min f = \sum_{j \in J} \sum_{i \in I} L_{ij} \times x_{ij} + \sum_{j \in J} y_j$$

S.T

$$\sum_{j \in J} x_{ij} = 1, \forall i \in I$$

$$\sum_{i \in I} c_i \times x_{ij} \leq C_j \times y_j, \forall j \in J$$

$$L_{ij} \times x_{ij} \leq l_i, \forall i \in I, \forall j \in J$$

$$y_i \leq \sum_{i \in I} x_{ij}, \forall j \in J$$

$$x_{ij}, y_j \in \{0,1\}, \forall i \in I, \forall j \in J$$

Vì vấn đề phân bổ tài nguyên của chúng ta là NP-Hard trong tối ưu hóa, thời gian hội tụ cần thiết để đạt được các giải pháp tối ưu bằng cách sử dụng phương pháp chính xác dựa trên công thức ILP (Integer Linear Programming) tăng theo hàm số mũ với sự tăng của số lượng yêu cầu ăng-ten. Do đó, chúng ta cần nghiên cứu các thuật toán xấp xỉ mới mà hội tụ nhanh chóng và cung cấp các giải pháp tối ưu hoặc gần tối ưu cho các trường hợp vấn đề lớn hơn.

Sau đây là các thuật toán heuristic được đề xuất:

- *Matroid-based approach*

Vì giải pháp tối ưu được cung cấp bởi công thức ILP đang hiệu quả trong việc tối ưu hóa độ trễ và phân bổ tài nguyên một cách đồng thời, chúng tôi đề xuất một thuật toán mới dựa trên lý thuyết matroid với các thuộc tính và tiêu chí tương tự.

Việc đánh giá độ phức tạp của thuật toán Matroid-based được đề xuất là rất quan trọng. Lưu ý rằng vấn đề được giải quyết là NP-Hard, nên cần các phương pháp nhanh chóng và hiệu quả để đối phó với độ phức tạp này. Trong thực tế, thuật toán Matroid-based có độ phức tạp toàn cục (trong trường hợp xấu nhất) là  $O(m \ln(m) + m)$ , trong đó giá trị thứ nhất  $m \ln(m)$  là độ phức tạp của việc sắp xếp một tập hợp  $m$  cạnh theo trọng số của chúng (độ trễ trong trường hợp này), và giá trị thứ hai  $m$  là số lần vòng lặp "For" được thực thi theo thuật toán.

---

**Algorithm 3** Matroid-based algorithm for RRH-BBU assignment problem

---

```

Put  $A = \emptyset$ ;
 $l_{e_1} \leq l_{e_2} \leq \dots \leq l_{e_m}$ ;
for  $i = 1$  to  $m$  do
  if  $A \cup \{e_i\} \in \mathcal{F}$  then
    if  $c_{I(e_i)} \leq C_{T(e_i)}$  then
       $A := A \cup \{e_i\}$ 
       $C_{T(e_i)} - = c_{I(e_i)}$ 
    end if
  end if
end for

```

$l_{e_i}$  is the communication latency on the edge  $e_i$ ;  
 $I(e_i)$  (resp.  $T(e_i)$ ) represents the initial (resp. terminal) extremity of the edge  $e_i$ ;  
 $c_{I(e_i)}$  represents the number of CPU cores requested for processing the antenna demand  $I(e_i)$ ;  
 $C_{T(e_i)}$  represents the available amount of CPU in an edge data center  $T(e_i)$ .

---

- *B-matching formulation*

Một phương pháp tiếp cận heuristics mới dựa trên lý thuyết b-matching để giải quyết các trường hợp vấn đề lớn hơn và tìm các giải pháp tối ưu hoặc gần tối ưu trong thời gian ngắn. Thuật toán b-matching sử dụng bipartite graph, để tìm b-matching có trọng số nhỏ nhất giữa các ăng-ten và trung tâm dữ liệu khi đồng thời giảm số lượng trung tâm dữ liệu được sử dụng và độ trễ được yêu cầu để xử lý các yêu cầu được tổng hợp. Thuật toán này sẽ đồng thời xem xét các yêu cầu về độ trễ mạnh mẽ của các yêu cầu từ các ăng-ten và các ràng buộc về khả năng xử lý giới hạn của các trung tâm dữ liệu.

Với  $G = (I \cup J, E)$  là đồ thị weighted bipartite. Vấn đề phân bổ tài nguyên có ràng buộc được định nghĩa ở trên có thể được giải quyết bằng cách tìm b-matching có trọng số nhỏ nhất trong khi xem xét các thông số sau đây:



The integral edge capacities:  $u = I$

$b(i) = 1, \forall i \in I$  ( $I$  là một tập các ăng-ten).

$b(j) = \min \left\{ |I_j|, \left\lfloor \frac{C_j}{\overline{c(j)}} \right\rfloor \right\}, \forall j \in J$  ( $J$  là một tập trung tâm dữ liệu)

$I_j$ : là một tập con ăng-ten được phân bổ cho trung tâm dữ liệu  $j \in J$  trong khi đảm bảo đáp ứng được yêu cầu về thời gian trễ mong đợi và số lượng lõi CPU cho yêu cầu mỗi ăng-ten:  $I_j = \{i \in I \mid l_i \geq L_{ij} \wedge (c_i \leq C_j)\}$

$\overline{c(j)}$ : Số lõi CPU trung bình của các yêu cầu ăng-ten có thể được gán cho trung tâm dữ liệu  $j \in J$ :  $\overline{c(j)} = \frac{\sum_{i \in I_j} c_i}{|I_j|}$

Ngoài ra, để giúp vấn đề tối ưu hóa của chúng ta tìm được giải pháp tối ưu với các biến nguyên, chúng ta thêm các bất đẳng thức blossom dưới đây vào công thức:

$$\sum_{e \in E(G[X])} x_e + \sum_{e \in E} x_e \leq \left\lfloor \frac{1}{2} (\sum_{v \in X} b(v) + |F|) \right\rfloor, \forall X \subseteq I \cup J, F \subseteq \delta(X)$$

Ở đây,  $E(G(X))$  đại diện cho một tập con các edges trong đồ thị con  $G(X)$  được tạo ra bởi một tập con các đỉnh  $X$ , và  $\delta(X)$  là một tập các edges liền kề của tập đỉnh  $X$ .

Bài toán tối ưu:

$$\min \mathcal{F} = \sum_{e \in E} L_e \times x_e$$

*S.T.:*

$$\sum_{e \in \delta(i)} x_e = 1, \forall i \in I;$$

$$\sum_{e \in \delta(j)} x_e \leq \min \left\{ |I_j|, \left\lfloor \frac{C_j}{\overline{c(j)}} \right\rfloor \right\} \forall j \in J;$$

$$\sum_{e \in E(G[X])} x_e + \sum_{e \in E} x_e \leq \left\lfloor \frac{1}{2} (\sum_{v \in X} b(v) + |F|) \right\rfloor, \forall X \subseteq I \cup J, F \subseteq \delta(X)$$

$$x_e \in \mathbb{R}^+, \forall e \in E;$$

- *Multiple knapsack-based algorithm*

Ngoài các công thức toán học chính xác và thuật toán heuristics được đề xuất ở trên, một thuật toán mới sử dụng công thức multiple knapsack được nhiều người biết đến để giải quyết vấn đề phân bổ RRH-BBU khi xét đến các năng lực hạn chế của trung tâm dữ liệu. Thực tế, công thức multiple knapsack là một sự tổng quát hóa của vấn đề knapsack kinh điển từ một knapsack đến  $m$  knapsacks với các dung lượng khác nhau. Mục tiêu của thuật toán multiple knapsack là phân bổ mỗi mặt hàng vào

tối đa một trong các knapsacks sao cho không có ràng buộc dung lượng nào bị vi phạm và tổng lợi nhuận của các mặt hàng đặt vào các knapsack được tối đa hóa.

Bằng cách biểu diễn vấn đề phân bổ tài nguyên giới hạn dưới dạng bipartite graph  $G = (I \cup J, E)$ , trong đó  $I$  là tập hợp các RRH và  $J$  là tập hợp các BBU, và  $E$  biểu thị các kết nối có thể giữa RRH và BBU, chúng ta có thể thiết lập một sự tương đương giữa vấn đề phân bổ tài nguyên và công thức multiple knapsack.

- The knapsack là những trung tâm dữ liệu ( $j \in J$ )
- Các yêu cầu của ăng-ten ( $i \in I$ ) được xem như là những mặt hàng để vào knapsack (trung tâm dữ liệu)
- Trọng số  $w_j$  là số lượng lõi CPU  $c_i$  được yêu cầu để xử lý yêu cầu của ăng-ten  $i$
- Giá trị lợi nhuận  $p_j$  không khác nhau giữa các yêu cầu của các anten và có thể được thiết lập là 1 ( $p_j = 1$ ).

Công thức trước giải quyết vấn đề phân bổ tài nguyên bị ràng buộc bằng cách chỉ tập trung vào khả năng xử lý của các trung tâm dữ liệu, khi giảm yêu cầu về độ trễ của các yêu cầu từ anten. Sự giảm bớt này ảnh hưởng đến việc lựa chọn trung tâm dữ liệu nào sẽ đáp ứng yêu cầu từ anten. Do đó, để xem xét các ràng buộc này trong giải pháp cuối cùng, một sửa đổi đơn giản trong thuật toán multiple knapsack [4], bao gồm kiểm tra xem độ trễ dự kiến có được đảm bảo trước khi gán yêu cầu từ anten cho trung tâm dữ liệu.

Thuật toán:

---

**Algorithm 4** Modified Multiple Knapsack Algorithm

---

**Input:**  $G = (I \cup J, E)$ , Antenna demands, Edge data centers.

**Output:** A joint mapping (CPU, Latency) of all antennas demands on the available edge data centers.

This is summarized formally in steps:

**Step 1:** Sort the edge data centers ( $j \in J$ ) in increasing order of their CPU capacities  $C_j$ ;

**Step 2:** Select the antennas demands that can be assigned to the selected edge data center  $j$  by checking if :

- The expected latency of the antenna demand is provided by the communication link joining it to the selected edge data center  $j$ ;
- The available computing resources in the selected edge data center  $j$  are greater than the number of CPU cores requested by the antenna demand;

**Step 3:** Pick as many antennas demands as possible to the selected edge data center using the dynamic programming approach (see Algorithm 2);

**Step 4:** Update the total number of available CPU cores in the selected edge data center;

**Step 5:** Repeat Steps 2, 3 and 4 until all considered antennas demands are assigned to the edge data centers;

---

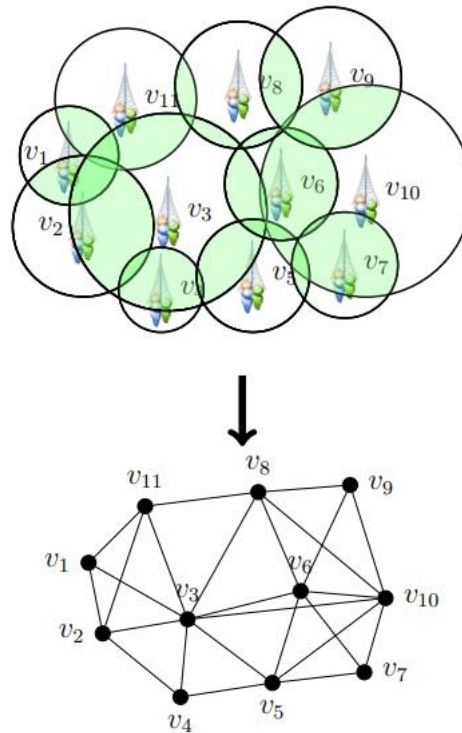
## 4.2. Tối ưu hóa độ phủ sóng mạng trong C-RAN

Sự tối ưu đáng kể về độ trễ, tận dụng tài nguyên và tiết kiệm chi phí, đạt được thông qua việc tối ưu phân bổ RRH-BBU (được trình bày ở phần trên), chịu sự hạn chế lớn từ sự gia tăng nhiễu nội tuyến, gây giảm đáng kể chất lượng dịch vụ của người dùng cuối. Thật vậy, các nhà điều hành mạng đang tìm kiếm các giải pháp mới để tăng mật độ các cell hiện có bằng cách triển khai thêm các ăng-ten để mở rộng phổ mạng và đáp ứng các yêu cầu của người dùng cuối. Sự tăng mật độ mạng được coi là một phương pháp chính trong kiến trúc C-RAN để nâng cao khả năng phủ sóng và dung lượng mạng và đạt được tốc độ truyền dữ liệu cao hơn. Tuy nhiên, điều này mang lại nhiều thách thức bao gồm quản lý và giảm thiểu nhiễu nội tuyến, do triển khai đặc biệt dày đặc các ăng-ten, và tối ưu hóa phủ sóng mạng bằng cách phát hiện

và loại bỏ các lỗ không phủ sóng. Do đó, việc nghiên cứu các phương pháp mới cho phép tìm thấy một sự cân bằng tốt giữa việc loại bỏ/giảm thiểu nhiễu nội tuyến và tối ưu hóa phủ sóng mạng là rất quan trọng đối với các nhà điều hành mạng.

#### 4.2.1. Mô hình hệ thống

Xem xét một mạng di động triển khai trên một khu vực lớn được đại diện bởi một tập hợp các anten. Khu vực phủ sóng của mỗi ăng-ten được mô hình bằng các vòng tròn có bán kính phủ sóng có thể thay đổi.



Hình 4.3: xây dựng đồ thị dựa trên vị trí của các ăng-ten và nhiễu [4]

Như được miêu tả trong Hình 10, mô hình mạng được mô tả bằng một đồ thị vô hướng được ký hiệu là  $G = (\mathcal{A}, \mathcal{E})$  trong đó  $\mathcal{A}$  và  $\mathcal{E}$  lần lượt là các tập hợp của các nút và cạnh có sẵn. Có một cạnh  $(i, j)$  giữa hai ăng-ten  $i$  và  $j$  nếu điều kiện sau đây được đáp ứng.

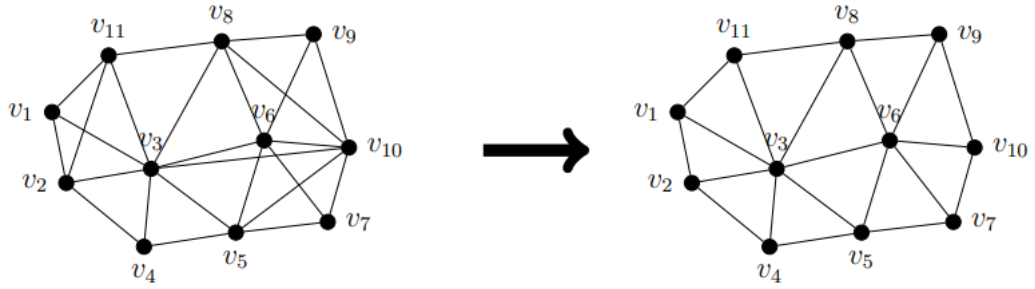
$$r_i + r_j \geq d_{ij}$$

trong đó  $r_i$  và  $r_j$  lần lượt là bán kính của ăng-ten  $i$  và  $j$ , và  $d_{ij}$  là khoảng cách Euclid giữa  $i$  (với tọa độ  $(x_i, y_i)$ ) và  $j$  (với tọa độ  $(x_j, y_j)$ ):

$$d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

Hãy để  $\delta_{ij}$  là sự trùng lấp (inter-cell interference) gây ra bởi hai ăng-ten  $i$  và  $j$ .

$$\delta_{ij} = r_i + r_j - d_{ij}$$



Hình 4.4: Một ví dụ giải pháp đồ thị của vấn đề mạng phủ sóng đầy đủ [4]

#### 4.2.2. Bài toán tối ưu và giải thuật

Trong phần này, mục đích tìm kiếm các phương pháp mới cho phép tập trung và tối ưu lại bán kính của các ăng-ten nhằm giảm nhiễu nội tuyến trong khi vẫn duy trì phủ sóng toàn bộ mạng trong kiến trúc C-RAN. Một mô hình toán học mới dựa trên công thức ILP để mô tả convex hull của bài toán tối ưu phủ sóng toàn bộ mạng. Sau đó, mở rộng mô tả này bằng cách thêm các bất đẳng thức và các mặt cắt để tăng tốc thời gian hội tụ và đạt được các giải pháp tối ưu ngay cả đối với số lượng lớn các ăng-ten. Cuối cùng, sẽ đề xuất một thuật toán Branch-and-Cut dựa trên các mặt cắt này để giải quyết hiệu quả các trường hợp về vấn đề quy mô lớn và đánh giá tính hiệu quả, hữu ích của phương pháp đề xuất.

Trước khi giới thiệu bài toán, các biến và tham số được sử dụng được liệt kê dưới đây:

- $G = (\mathcal{A}, \mathcal{E}, E)$  biểu diễn cấu trúc mạng như được minh họa trong hình 10.  $\mathcal{A}$  là tập hợp các ăng-ten và  $\mathcal{E}$  là tập hợp các cạnh giữa các ăng-ten

- Mỗi anten  $i \in \mathcal{A}$  có thể hoạt động với bán kính phủ sóng  $r_i$  của riêng nó, biến thiên trong khoảng  $[r_i^{min}, r_i^{max}]$
- Cho  $x_{ij}$  là một biến nhị phân cho biết liệu cạnh  $(i, j)$  của  $G$  có được xem xét trong giải pháp cuối cùng ( $x_{ij} = 1$ ), hoặc không ( $x_{ij} = 0$ ).
- $\mathcal{N}(i)$  là tập hợp các nút/ăng ten lân cận của  $i$ . Một nút  $j$  là lân cận của  $i$  chỉ nếu điều kiện  $r_i + r_j \geq d_{ij}$  được sử dụng với giá trị bán kính tối đa được đảm bảo.
- $\mathcal{J}(i, j)$  là tập hợp tất cả các cạnh  $(i, j)$  không giao với bất kỳ cạnh nào khác  $(k, l)$ , trong đó  $(i, j)$  và  $(k, l)$  không có điểm đầu và điểm cuối chung nào.

Mục tiêu của bài toán bao phủ toàn mạng là nhanh chóng phát hiện ra các điểm trống trong mạng và giảm đáng kể nhiễu tín hiệu giữa các cell được đo lường toán học bằng công thức  $\delta_{ij} = r_i + r_j - d_{ij}$ . Những mục tiêu này sẽ được đạt được bằng cách tối ưu hóa giá trị bán kính của các ăng-ten. Điều này tương đương với việc chỉ chọn các cặp ăng-ten có khoảng cách Euclide tối thiểu, đảm bảo tính kết nối đồ thị và bao phủ toàn mạng, giúp giảm nhiễu tín hiệu giữa các cell một cách trực quan. Ta có hàm mục tiêu như sau:

$$\min \Gamma = \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{N}(i)} d_{ij} \times x_{ij}$$

Ràng buộc đảm bảo rằng mỗi nút  $i$  có ít nhất hai lân cận trong đồ thị (mục tiêu là thu được một đa giác lồi đáp ứng tính kết nối và giảm nhiễu giữa các cell).

$$\sum_{j \in \mathcal{N}(i)} x_{ij} \geq 2, \quad \forall i \in \mathcal{A}$$

Ràng buộc đòi hỏi nếu một cạnh  $(i, j)$  không có bất kỳ giao điểm nào trong đồ thị ban đầu, thì  $x_{ij} = 1$  trong đồ thị cuối cùng (đồ thị giải pháp). Các ràng buộc này được cung cấp theo công thức toán học sau:

$$x_{ij} = 1, \forall i \in \mathcal{A}, \forall j \in \mathcal{N}(i), \forall (i, j) \in \mathcal{J}(i, j)$$

Để tránh sự giao nhau giữa hai cạnh  $(i, j)$  và  $(k, l)$  trong đồ thị cuối cùng. Bất đẳng thức phi tuyến được đề xuất.

$$x_{ij} \times x_{il} \leq x_{jl} + \sum_{k \in \mathcal{N}(i)} x_{ik}, \quad \forall i \in \mathcal{A}, \quad \forall j \in \mathcal{N}(i), \quad \forall l \in \mathcal{N}(i) \cap \mathcal{N}(i)$$

Ràng buộc trên là phi tuyến vì chúng ta sử dụng tích của hai biến quyết định. Chúng ta sẽ thay thế chúng bằng một họ bất đẳng thức tuyến tính mới khi giới thiệu một biến nhị phân mới  $z_{ijl}$ , sao cho  $z_{ijl} = x_{ij} \times x_{il}$ ,  $\forall i \in \mathcal{A}, \forall j \in \mathcal{N}(i), \forall l \in \mathcal{N}(i) \cap \mathcal{N}(i)$ . Do đó, chúng ta thu được các ràng buộc như sau:

$$z_{ijl} \leq x_{ij}$$

$$z_{ijl} \leq x_{il}$$

$$z_{ijl} \leq x_{ij} + x_{il} - 1$$

Mô hình toán học được xác định bởi Integer Linear Programming:

$$\min \Gamma = \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{N}(i)} d_{ij} \times x_{ij}$$

*S.T.* :

$$\sum_{j \in \mathcal{N}(i)} x_{ij} \geq 2, \quad \forall i \in \mathcal{A}$$

$$x_{ij} = 1, \quad \forall i \in \mathcal{A}, \forall j \in \mathcal{N}(i), \forall (i, j) \in \mathcal{J}(i, j)$$

$$z_{ijl} \leq x_{jl} + \sum_{k \in \mathcal{N}(i)} x_{ik}, \quad \forall i \in \mathcal{A}, \quad \forall j \in \mathcal{N}(i), \quad \forall l \in \mathcal{N}(i) \cap \mathcal{N}(i)$$

$$z_{ijl} \leq x_{ij} + x_{il} - 1, \quad \forall i \in \mathcal{A}, \forall j \in \mathcal{N}(i), \forall l \in \mathcal{N}(i) \cap \mathcal{N}(i)$$

$$z_{ijl} \leq x_{ij}, \quad \forall i \in \mathcal{A}, \forall j \in \mathcal{N}(i), \forall l \in \mathcal{N}(i) \cap \mathcal{N}(i)$$

$$z_{ijl} \leq x_{il}, \quad \forall i \in \mathcal{A}, \forall j \in \mathcal{N}(i), \forall l \in \mathcal{N}(i) \cap \mathcal{N}(i)$$

$$x_{ij}, z_{ijl} \in \{0, 1\}, \quad \forall i \in \mathcal{A}, \forall j \in \mathcal{N}(i), \forall l \in \mathcal{N}(i) \cap \mathcal{N}(i)$$

Cuối cùng, cung cấp một thuật toán bên dưới, một bản tóm tắt của thuật toán đề xuất dựa trên hình thức Branch-and-Cut [4] để xử lý vấn đề bảo đảm độ phủ mạng.

---

**Algorithm 5 Full network coverage algorithm : Branch-and-Cut approach**

---

**Input:** A real telecommunications network (cells with antennas) with given interference and coverage holes

**Output:** A full coverage network (no holes) with no interference

- Graph transformation of the real network  $G$ : each antenna is a node
  - There is an edge between two nodes (antennas)  $i$  and  $j$  if  $r_i + r_j \geq d_{ij}$
  - An interference is represented by an intersection of two edges
  - Run the Branch and Cut optimization model (4.17)
  - The optimized/obtained network has no holes and no interference
- 

### 4.3. Đặc vị trí và phân chia chức năng BBU

Với sự tăng trưởng về lưu lượng dữ liệu di động, các nhà điều hành mạng sẽ phải thêm một lượng lớn tần số cùng việc tăng mật độ của các cell bằng cách triển khai thêm nhiều ăng-ten hơn. Điều này sẽ không chỉ tăng mức độ nhiễu giữa các cell (đã thảo luận trong phần trước), mà còn sẽ tăng đáng kể lượng xử lý cơ sở ban đầu cần thiết để đáp ứng số lượng tăng của các yêu cầu ăng-ten và do đó tăng lượng lưu lượng dữ liệu yêu cầu, với các ràng buộc về độ trễ và băng thông nghiêm ngặt, giữa các ăng-ten và trung tâm dữ liệu tập trung. Trong thực tế, trong khi điều tra trong phần trước các giải pháp mới để giảm nhiễu giữa các cell khi đảm bảo phủ sóng toàn bộ mạng, phần này thảo luận về cách vượt qua sự đánh đổi giữa lợi ích của việc tập trung xử lý cơ sở ban đầu trong BBU pools và yêu cầu độ trễ trên các liên kết fronthaul.

Trong phần này, chúng tôi nghiên cứu các phương pháp mới để tìm các sự đánh đổi tốt nhất giữa việc tập trung hóa và yêu cầu vận chuyển trên mạng fronthaul. Thực tế, để đáp ứng các yêu cầu này, phân phối linh hoạt hơn (hoặc phân chia) các chức

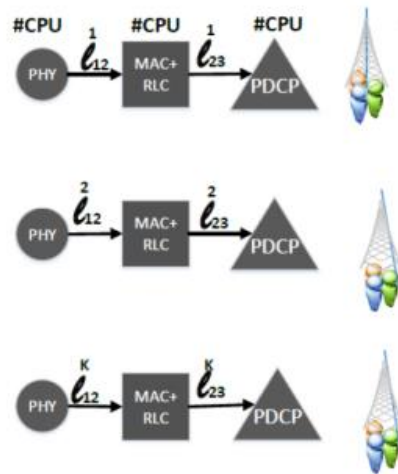


năng xử lý cơ sở sẽ được xem xét giữa các RRH và các trung tâm dữ liệu tập trung.. Trong ngữ cảnh này, một loạt các phân chia chức năng BBU đang được giới thiệu và nghiên cứu, mỗi lựa chọn trong số đó đều đòi hỏi các yêu cầu khác nhau về khả năng chịu tải, độ trễ và băng thông trên mạng fronthaul. Xem xét tùy chọn chia RAN 3GPP để mô hình hóa các yêu cầu thu gom của anten và tìm cách xác định một cách hiệu quả vị trí tối ưu của các chức năng BBU trong các trung tâm dữ liệu tập trung dựa trên cấu hình phân chia được xem xét và các đặc điểm của mạng vận chuyển. Các lợi ích về sử dụng tài nguyên có thể được đạt được bằng cách xác định vị trí tối ưu của các chức năng BBU khi đáp ứng đồng thời các yêu cầu xử lý và độ trễ của các yêu cầu của anten.

Để giải quyết vấn đề này, đề xuất một mô hình toán học dựa trên phương pháp ILP để triển khai hàm xử lý baseband một cách tối ưu trong mạng khi đồng thời giảm thiểu tiêu thụ tài nguyên và độ trễ dự kiến. Sau đó, với mục đích mở rộng, nghiên cứu các thuật toán heuristic mới dựa trên sự xây dựng một *multi-stage graph* để đạt được các giải pháp tốt cho kích thước mạng lớn trong thời gian chấp nhận được.

#### **4.3.1. Mô hình phân chia chức năng BBU**

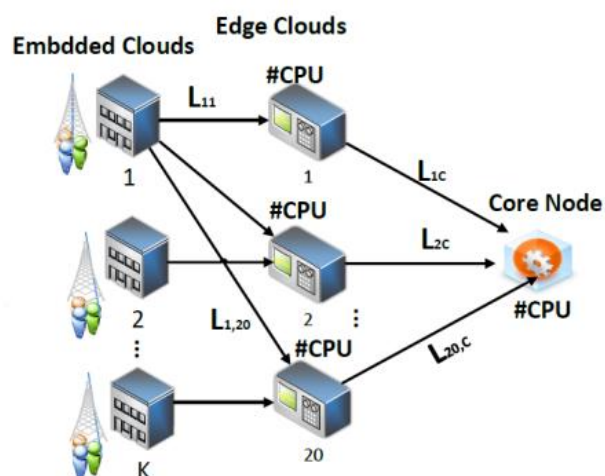
Sử dụng 3GPP RAN split để mô hình hóa vấn đề đặt chức năng phân chia BBU. Cấu hình chia này được đề xuất là tùy chọn tốt nhất trong các tiêu chuẩn 3GPP. Tùy chọn phân chia RAN 3GPP được sử dụng để mô hình hóa các yêu cầu tổng hợp của các anten được đại diện bởi một tập hợp các chuỗi có hướng, như được miêu tả trong Hình 12 Mỗi chuỗi chức năng được tạo thành từ ba nút đại diện cho bốn lớp được kết nối với nhau: PHY, MAC + RLC và PDCP. Để đơn giản, mỗi nút có các yêu cầu tài nguyên tính toán, được thể hiện dưới dạng lõi CPU. Các nút này được kết nối bởi hai vòng cung có trọng số cho thấy trình tự giữa các hàm BBU với độ trễ giao tiếp dự kiến.



Hình 4.5: Mô hình phân chia chức năng BBU cho mỗi yêu cầu của anten [4]

#### 4.3.2. Mô tả cấu trúc mạng

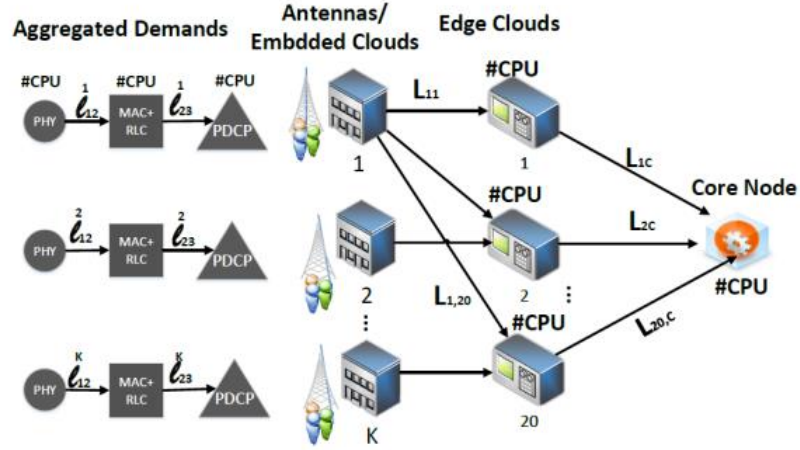
Hình 4.5 cho thấy kiến trúc mạng vật lý được xem xét trong vấn đề chia sẻ chức năng BBU. Trong thực tế, mạng vật lý được mô hình hóa như một kiến trúc phân cấp, đây là một phương pháp tiềm năng để đạt được các chiến lược triển khai linh hoạt của các chức năng BBU trên các trung tâm dữ liệu chia sẻ.



Hình 4.6: Kiến trúc mạng vật lý [4]

### 4.3.3. Mô hình hệ thống

Mục tiêu của vấn đề phân chia BBU là triển khai tối ưu các chuỗi yêu cầu (Hình 4.5) trên kiến trúc mạng (Hình 4.6) khi đáp ứng đồng thời yêu cầu CPU và độ trễ. Hình 4.7 minh họa mô hình hệ thống của cho vấn đề trên.



Hình 4.7: Mô hình hệ thống cho BBU function split placement [4]

### 4.3.4. Bài toán tối ưu và giải thuật

Các biến và các thông số để định nghĩa mô hình được liệt kê trong bảng dưới đây:

$G = (\mathcal{J}, \mathcal{E})$	: weighted directed graph.
$\mathcal{J}$	: tập các nút vật lý $j$ bao gồm các đám mây nhúng, đám mây biên và một đám mây trung tâm.
$\mathcal{E}$	: tập các link liên kết giữa các node vật lý $j$
$C_j$	: Số lõi CPU có sẵn trên nút vật lý $j$ (embedded, edge and core clouds)
$L_{(j,j')}$	: độ trễ trên liên kết giữa hai nút vật lý $j$ và $j'$
$j' \in \mathcal{P}(j)$	: tập tất cả các nút vật lý liên kết hai nút vật lý $j$ và $j'$
$\mathcal{K}$	: tập các chuỗi có hướng $k$ biểu thị các yêu cầu tổng hợp.

$i \in \{1, 2, 3\}$	: tập các nút ảo $i$ của mỗi chuỗi $k$ trong đó 1 đại diện cho lớp PHY, 2 đại diện cho nút thứ hai chứa các lớp MAC và RLC và 3 đại diện cho lớp PDCP
$c_i^k$	: Số lõi CPU được yêu cầu để xử lý nút ảo $i$ (các lớp PHY, MAC + RLC và PDCP) của chuỗi $k$ .
$l_{(i,i+1)}^k$	: Độ trễ dự kiến giữa hai nút ảo liên tiếp $i$ và $i + 1$ của cùng một chuỗi $k$ .

### *Biến quyết định (Decision variables)*

Bắt đầu việc mô hình hóa của vấn đề bằng cách giới thiệu ba biến quyết định như sau:

- $x_{i,j}^k$ : Là một biến nhị phân, giá trị của nó là 1 nếu nút ảo, tức là chức năng BBU,  $i$  của ăng-ten  $k$  được đặt trên nút vật lý  $j$ , và 0 nếu ngược lại.
- $y_{(i,i+1);(j,j')}^k$ : Là Một biến nhị phân, giá trị của nó là 1 nếu một cạnh ảo (virtual edge)  $(i, i+1)$  cho một liên kết giữa hai chức năng BBU của yêu cầu ăng-ten  $k$  được đặt trên một đường dẫn vật lý nối hai nút vật lý  $j$  và  $j'$ , và 0 nếu ngược lại.
- $z_j$ : Là một biến nhị phân, giá trị của nó là 1 nếu một nút vật lý  $j$  được sử dụng và 0 nếu ngược lại.

### *Hàm mục tiêu:*

Mục tiêu của việc đặt chức năng BBU là ánh xạ đồng thời tất cả các chuỗi được kết nối vào mạng vật lý, đồng thời giảm thiểu tổng lượng tiêu thụ lõi CPU và độ trễ end-to-end. Nó được xác định bởi:

$$\min \mathcal{F} = - \sum_{j \in \mathcal{J}} (C_j z_j - \sum_{k \in \mathcal{K}} \sum_{i \in \{1,2,3\}} c_i^k x_{i,j}^k) + \sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{J}} \sum_{j' \in \mathcal{P}(j)} \sum_{i \in \{1,2\}} L_{(j,j')} y_{(i,i+1);(j,j')}^k$$

phần đầu của công thức trên chỉ định tài nguyên tính toán còn dư lại (CPU) trong các trung tâm dữ liệu khác nhau, tức là các trung tâm dữ liệu nhúng, cạnh và trung tâm lõi đám mây (embedded, edge and core cloud data centers). Trong khi phần thứ hai

đại diện cho tổng chi phí về độ trễ được cung cấp để xử lý các chuỗi chức năng BBU (yêu cầu ăng-ten tổng hợp). Giải pháp tối ưu của vấn đề này sẽ giảm thiểu độ trễ kỳ vọng trên mạng fronthaul và tối đa hóa tài nguyên tính toán còn lại theo số lõi CPU trên mỗi nút của mạng vật lý. Giải pháp này phải tuân thủ một tập hợp các ràng buộc của vấn đề được đề cập sẽ được trình bày chi tiết trong phần tiếp theo.

*Các ràng buộc (Constraints):*

Vấn đề tối ưu hóa cho việc đặt chức năng BBU có một tập hợp các ràng buộc nhất định như sau:

Ràng buộc dưới đây ngăn chặn việc đặt lớp vật lý của chuỗi  $k$  trên các embedded nodes khác không được chỉ định cho ăng-ten  $k$ :

$$\sum_{j \in \mathcal{J}_1} \mathbf{1}_{(k,j)} x_{1,j}^k = 0, \forall k \in \mathcal{K}$$

trong đó  $\mathbf{1}_{(k,j)}$  bằng 1 nếu ăng-ten  $k$  không được chỉ định cho nút nhúng  $j$ , và bằng 0 nếu ngược lại và  $\mathcal{J}_1$  đại diện cho tập các đám mây nhúng.

Ràng buộc dưới đây đảm bảo rằng mỗi nút ảo, tức là một chức năng BBU, được triển khai trên chính xác một nút vật lý.

$$\sum_{j \in \mathcal{J}} x_{i,j}^k = 1, \forall k \in \mathcal{K}, \forall i \in \{1,2,3\}$$

Ràng buộc dưới đây đảm bảo rằng việc đặt các chức năng BBU không thể tiêu thụ nhiều tài nguyên hơn tài nguyên có sẵn trên nút vật lý được chọn.

$$\sum_{k \in \mathcal{K}} \sum_{i \in \{1,2,3\}} x_{i,j}^k \times c_i^k \leq C_j, \forall j \in \mathcal{J}$$

Ràng buộc dưới đây được sử dụng để đảm bảo việc xếp chồng các chức năng BBU. Ví dụ, nếu lớp PHY được triển khai trên nút vật lý  $i$ , thì nút ảo chứa các lớp MAC và RLC phải được triển khai trên  $j$  sao cho có một đường dẫn vật lý  $P(j)$  bắt đầu từ nút  $i$  đến nút  $j$ .

$$x_{i,j}^k \leq \sum_{j' \in \mathcal{P}(j)} x_{i+1,j'}^k, \forall k \in \mathcal{K}, \forall i \in \{1,2\}, \forall j \in \mathcal{J}$$

Hai ràng buộc dưới đây đảm bảo rằng nếu một nút ảo  $i$  được triển khai trên một nút vật lý  $j$  tức là  $x_{i,j}^k = 1$ , thì nút ảo  $i + 1$  phải nằm trên một nút vật lý  $j'$  tức là  $x_{i+1,j'}^k = 1$  mà có liên kết vật lý với  $j$  (tức là có đường đi vật lý từ  $j$  đến  $j'$ ).

$$\sum_{j \in \mathcal{V}} y_{(i,i+1);(j,j')}^k = x_{i+1,j'}^k, \forall k \in \mathcal{K}, \forall i \in \{1,2\}, \forall j' \in \mathcal{P}(j)$$

$$\sum_{j' \in \mathcal{P}(j)} y_{(i,i+1);(j,j')}^k = x_{i,j}^k, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}, \quad \forall j \in \mathcal{J}$$

Ràng buộc dưới đây đảm bảo rằng mỗi virtual arc  $(i, i+1)$  được triển khai trên chính xác một đường dẫn vật lý.

$$\sum_{j \in \mathcal{J}} \sum_{j' \in \mathcal{P}(j)} y_{(i,i+1);(j,j')}^k = 1, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}$$

Ràng buộc dưới đây yêu cầu rằng độ trễ trên đường truyền fronthaul, trên đường đã chọn trong mạng vật lý không được vượt quá yêu cầu về độ trễ của chuỗi chức năng BBU.

$$L_{(j,j')} \times y_{(i,i+1);(j,j')}^k \leq l_{(i,i+1)}^k, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}, \quad \forall j \in \mathcal{J}, \forall j' \in \mathcal{P}(j)$$

Ràng buộc dưới đây cho thấy nếu tồn tại ít nhất một chức năng BBU được triển khai trên một nút vật lý  $j$ , thì chức năng đó nên được sử dụng để chứa các nút ảo khác nếu cần thiết.

$$x_{i,j}^k \leq z_j, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2,3\}, \quad \forall j \in \mathcal{J}$$

Bài toán tối ưu về phân chia chức năng BBU: Integer Linear Programming

$$\begin{aligned} \min \mathcal{F} = & - \sum_{j \in \mathcal{J}} (C_j z_j - \sum_{k \in \mathcal{K}} \sum_{i \in \{1,2,3\}} c_i^k x_{i,j}^k) + \\ & \sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{J}} \sum_{j' \in \mathcal{P}(j)} \sum_{i \in \{1,2\}} L_{(j,j')} y_{(i,i+1);(j,j')}^k \end{aligned}$$

S.T.:

$$\sum_{j \in \mathcal{J}_1} \mathbf{1}_{(k,j)} x_{1,j}^k = 0, \quad \forall k \in \mathcal{K}$$

$$\sum_{j \in \mathcal{J}} x_{i,j}^k = 1, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2,3\}$$

$$\sum_{k \in \mathcal{K}} \sum_{i \in \{1,2,3\}} x_{i,j}^k \times c_i^k \leq C_j, \quad \forall j \in \mathcal{J}$$

$$x_{i,j}^k \leq \sum_{j' \in \mathcal{P}(j)} x_{i+1,j'}^k, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}, \quad \forall j \in \mathcal{J}$$

$$\sum_{j \in \mathcal{V}} y_{(i,i+1);(j,j')}^k = x_{i+1,j'}^k, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}, \quad \forall j' \in \mathcal{P}(j)$$

$$\sum_{j' \in \mathcal{P}(j)} y_{(i,i+1);(j,j')}^k = x_{i,j}^k, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}, \quad \forall j \in \mathcal{J}$$

$$\sum_{j \in \mathcal{J}} \sum_{j' \in \mathcal{P}(j)} y_{(i,i+1);(j,j')}^k = 1, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}$$

$$L_{(j,j')} \times y_{(i,i+1);(j,j')}^k \leq l_{(i,i+1)}^k, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2\}, \quad \forall j \in \mathcal{J}, \forall j' \in \mathcal{P}(j)$$

$$x_{i,j}^k \leq z_j, \quad \forall k \in \mathcal{K}, \quad \forall i \in \{1,2,3\}, \quad \forall j \in \mathcal{J}$$

$$x_{i,j}^k \in \{0,1\}, \quad \forall i \in \{1,2,3\}, \quad \forall j \in \mathcal{J}$$

$$y_{(i,i+1);(j,j')} \in \{0,1\}, \forall i \in \{1,2\}, \forall j \in \mathcal{J}$$

$$z_j \in \{0,1\}, \forall i \in \{1,2\}, \forall j \in \mathcal{J}$$

Để giải quyết kích thước bài toán lớn, chúng tôi đề xuất các thuật toán heuristic hiệu quả mà hội tụ đến các giải pháp tối ưu hoặc gần tối ưu (Multi-stage graph algorithm) [4]

---

**Algorithm 6** Multi-stage graph algorithm

---

**Input:** Aggregated BBU functions chains, Physical network.

**Output:** A joint mapping of all the requested chains on the physical network.

**Step 1:** Select a strategy (MIN-MIN, MIN-MAX, MAX-MIN or MAX-MAX) to find the optimal or near-optimal mapping of BBU function chains;

**Step 2:** Sort all the requested chains according to the total amount of requested CPU for each chain  $c$ ;

**Step 3:** Create the multi-stage graph  $\mathcal{G}_m$  according to the description given above;

**Step 4:** If the selected chain  $c$  is deployed : repeat until all chains are deployed, otherwise : the problem has no **complete\*** solution.

\*There is a solution when all the BBU functions chains are deployed successfully.

---

Việc đánh giá độ phức tạp của thuật toán Multi-stage graph được đề xuất là rất quan trọng. Vấn đề được giải quyết là NP-Hard, nên cần các phương pháp nhanh chóng và hiệu quả về chi phí để đối phó với độ phức tạp này.

## Chương 5 : Kết luận và hướng phát triển

Trong bối cảnh phát triển nhanh chóng của công nghệ mạng, việc nâng cao chất lượng người dùng (QoE) đã trở thành một ưu tiên quan trọng. Hai công nghệ đóng vai trò quan trọng trong việc đạt được mục tiêu này là Mạng Mật độ Cao (UDN) và nền tảng truy cập vô tuyến đám mây (C-RAN). UDN tập trung vào việc xử lý các vấn đề như tải trọng mật độ cao và khả năng truy cập tốt trong các môi trường mật độ dân số cao. C-RAN sử dụng cơ sở hạ tầng đám mây để tập trung các tài nguyên và cung cấp hiệu suất cao hơn cho người dùng cuối. UDN là một kiến trúc mạng tiên tiến, cho

phép triển khai các trạm cơ sở có mật độ cao hơn và cung cấp phạm vi phủ sóng rộng hơn. Nó giúp giảm sự chồng chéo tần số và tăng khả năng xử lý dữ liệu của mạng. C-RAN là một kiến trúc mạng truy cập mở rộng và mạnh mẽ, trong đó các chức năng xử lý tách rời và được tập trung trong trung tâm dữ liệu truy cập vô tuyến đám mây. Việc kết hợp UDN và C-RAN mang lại nhiều lợi ích cho người dùng, như giảm độ trễ, tăng băng thông, cải thiện khả năng phục hồi sau sự cố và tăng tính sẵn sàng của mạng.

Để phát triển chủ đề này, có một số hướng đi tiềm năng:

Tăng cường khả năng kết nối: Nghiên cứu và phát triển các công nghệ truyền dẫn không dây và anten thông minh để tăng cường khả năng kết nối trong UDN và C-RAN. Việc nâng cao khả năng kết nối giữa các trạm cơ sở và trung tâm dữ liệu sẽ cải thiện hiệu suất mạng và trải nghiệm người dùng.

Tối ưu hóa quản lý tài nguyên: Nghiên cứu và phát triển các thuật toán và giao thức quản lý tài nguyên hiệu quả để đảm bảo sự phân bổ tài nguyên hợp lý trong UDN và C-RAN. Quản lý tài nguyên thông minh giúp tối ưu hóa sử dụng băng thông và năng lượng, đồng thời tăng khả năng mở rộng của mạng.

Bảo mật và quản lý mạng: Tăng cường các biện pháp bảo mật để đảm bảo an toàn thông tin trong UDN và C-RAN. Nghiên cứu các phương pháp quản lý mạng linh hoạt để giám sát và điều khiển mạng một cách hiệu quả.

Tích hợp công nghệ mới: Tiếp tục nghiên cứu và tích hợp các công nghệ mới như trí tuệ nhân tạo (AI), Internet of Things (IoT) và blockchain vào UDN và C-RAN. Việc áp dụng các công nghệ mới này có thể đem lại nhiều cơ hội mới và tăng cường khả năng của hệ thống.

Hướng phát triển này đòi hỏi sự cộng tác giữa các nhà nghiên cứu, nhà cung cấp dịch vụ viễn thông và các tổ chức liên quan. Việc đầu tư vào nghiên cứu và phát triển trong lĩnh vực này có thể mang lại lợi ích lớn cho người dùng và góp phần nâng cao chất lượng mạng trong tương lai.



## Tài liệu tham khảo

- [1] Cunhua Pan, Maged ElKashlan, Jiangzhou Wang, Jinhong Yuan, and Lajos Hanzo, “*User-Centric C-RAN Architecture for Ultra-Dense 5G Networks: Challenges and Methodologies*”, IEEE Communications Magazine, tháng 6 2018.
- [2] Jianchi Zhu, Xiaoming She and Peng Chen, “*Ultra Dense Networks: General Introduction and Design Overview*”, 2016 John Wiley & Sons.
- [3] China Mobile Research Institute -White Paper, “*C-RAN The Road Towards Green RAN*”, bản 2.5 tháng 10, 2011.
- [4] Universite Paris-Saclay, “*Cloud-Radio Access Networks : Design, Optimization and Algorithms*”, 10/10/2019.
- [5] Aleksandra Checko, Henrik L. Christiansen, Ying Yan, Lara Scolari, Georgios Kardaras, Michael S. Berger, and Lars Dittmann. “*Cloud RAN for Mobile Networks—A Technology Overview*”, IEEE communication surveys & tutorials, vol. 17, 2015.
- [6] Checko, Aleksandra; Berger, Michael Stübert; Kardaras, Georgios; Dittmann, Lars; Christiansen, Henrik Lehrmann. “*Cloud Radio Access Network architecture. Towards 5G mobile networks*”, 2016.
- [7] Rui Wang, Honglin Hu, Senior Member, IEEE and Xiumei Yang. “*Potentials and Challenges of C-RAN Supporting Multi-RATs towards 5G Mobile Networks*”, 2014.
- [8] Venu Balaji Vinnakota, Kirtan Gopal Panda, Debarati Sen and Sandip chakraborty. “*Fronthaul Design in Cloud Radio Access Networks:A Survey*”. Indian Institute of Technology Kharagpur, Kharagpur, India 721302.
- [9] Seok-Hwan Park , Member IEEE, Seongah Jeong , Member IEEE, Jinyeop Na, Student Member, IEEE, Osvaldo Simeone , Fellow IEEE, and Shlomo Shamai , Life Fellow IEEE. “*Collaborative Cloud and Edge Mobile Computing in C-RAN Systems With Minimal End-to-End Latency*”. IEEE transactions on signal and information processing over networks, vol. 7, 2021.
- [10] Milecia McGregor. “*8 Clustering Algorithms in Machine Learning that All Data Scientists Should Know*”. Freecodecamp september 21, 2020.