



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Πολυτεχνική Σχολή
Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Διπλωματική Εργασία

Τεχνικές μη εποπτευόμενης μηχανικής μάθησης για τη βελτίωση της τεχνολογίας DUDe

Χίμου Ελένη

A.M. 1067073

Επιβλέπων

Χρήστος Μπούρας, Καθηγητής

Μέλη Επιτροπής Αξιολόγησης

Ιωάννης Γαροφαλάκης, Καθηγητής
Εύη Παπαϊωάννου, Επίκουρη Καθηγήτρια

Πάτρα, 2026

© Copyright συγγραφέας Χίμου Ελένη 2026

© Copyright θέματος Μπούρας Χρήστος

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής του Πανεπιστημίου Πατρών δεν υποδηλώνει απαραίτητα και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά, τον πρότανη του Πανεπιστημίου Πατρών και υπεύθυνο της διπλωματικής μου εργασίας, κύριο Μπούρα Χρήστο, για την εμπιστοσύνη του και την ευκαιρία που μου έδωσε να αναλάβω την παρούσα διπλωματική, καθώς και για την βοήθεια που μου παρείχε ώστε να ολοκληρώσω αυτή την προσπάθεια.

Ένα μεγάλο ευχαριστώ και στα μέλη της τριμελούς επιτροπής μου, κ. Γαροφαλάκη Ιωάννη και κ. Παπαϊωάννου Εύη, για τον χρόνο και την προθυμία τους, καθώς και για τη συμβολή τους στην ολοκλήρωση της παρούσας εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω επίσης, τον Δρ. Κόκκινο Βασίλειο, για την άψογη συνεργασία, καθώς και την στήριξη και την καθοδήγηση του, κατά τη διάρκεια εκπόνησης αυτής της εργασίας.

Περίληψη

Στα σύγχρονα δίκτυα, η παραδοσιακή συζευγμένη σύνδεση uplink–downlink, όπου η επιλογή της κυψέλης σύνδεσης τόσο για το downlink όσο και για το uplink βασίζεται κυρίως στην ισχύ του λαμβανόμενου σήματος στο downlink, οδηγεί συχνά σε αναποτελεσματική κατανομή φορτίου, ιδιαίτερα στο uplink, όπου οι περιορισμοί ισχύος των χρηστών καθιστούν τα macro cells σημείο συμφόρησης. Η παρούσα διπλωματική εργασία μελετά το πρόβλημα της σύνδεσης χρηστών στο uplink σε ετερογενή κυψελωτά δίκτυα, με έμφαση στη μέθοδο Decoupled Uplink and Downlink Access (DUDe). Αρχικά, υλοποιείται ένα ρεαλιστικό σενάριο ετερογενούς δικτύου με έναν macro σταθμό βάσης και πολλαπλούς pico σταθμούς, και αξιολογείται η απόδοση της συζευγμένης πολιτικής σύνδεσης σε σύγκριση με το DUDe. Τα αποτελέσματα δείχνουν ότι το DUDe οδηγεί σε σημαντική αποσυμφόρηση του macro cell στο uplink και σε μετρήσιμη βελτίωση της συνολικής χωρητικότητας, χωρίς να απαιτείται αυθαίρετη ανακατανομή χρηστών. Στη συνέχεια, διερευνάται η δυνατότητα προσέγγισης της πολιτικής DUDe μέσω τεχνικών μη εποπτευόμενης μηχανικής μάθησης. Χρησιμοποιώντας χαρακτηριστικά που σχετίζονται με την απόσταση, το SINR και τη χωρητικότητα στο uplink, εφαρμόζονται μέθοδοι ομαδοποίησης χρηστών και αναλύεται η συμπεριφορά διαφορετικών κατηγοριών χρηστών. Τα αποτελέσματα δείχνουν ότι οι μέθοδοι μηχανικής μάθησης μπορούν να προσεγγίσουν τη συμπεριφορά του DUDe, παράγοντας πιο συντηρητικές αλλά σταθερές αποφάσεις σύνδεσης.

Abstract

In modern cellular networks, the traditional coupled uplink–downlink association, where the selection of the serving cell for both uplink and downlink is primarily based on the received signal power in the downlink, often leads to inefficient load distribution, particularly in the uplink, where user power limitations render macro cells a congestion point. This diploma thesis investigates the problem of uplink user association in heterogeneous cellular networks, with emphasis on the Decoupled Uplink and Downlink Access (DUDe) approach.

Initially, a realistic heterogeneous network scenario consisting of one macro base station and multiple pico base stations is implemented, and the performance of the coupled association policy is evaluated in comparison with DUDe. The results demonstrate that DUDe leads to significant uplink load relief at the macro cell and to a measurable improvement in overall capacity, without requiring arbitrary user reassignment.

Subsequently, the feasibility of approximating the DUDe policy through unsupervised machine learning techniques is investigated. Using features related to distance, uplink SINR, and uplink capacity, user clustering methods are applied and the behavior of different user categories is analyzed. The results indicate that machine learning techniques can approximate the behavior of DUDe, producing more conservative but stable association decisions.

Περιεχόμενα

1	6
Εισαγωγή	6
1.1 Σημασία του προβλήματος	6
1.2 Στόχοι της Εργασίας	7
1.3 Διάρθρωση της Διπλωματικής Εργασίας	8
2	9
Θεωρητικό υπόβαθρο	9
2.1 Ετερογενή Κυψελωτά Δίκτυα στα Συστήματα 5G	9
2.2 Διαφορές Uplink και Downlink στα Δίκτυα 5G	10
2.3 Μέτρα Απόδοσης: SINR και Χωρητικότητα.....	11
2.4 Κατανομή Φορτίου και Πρόβλημα Σύνδεσης Χρηστών	12
2.5 Decoupled Uplink and Downlink Access (DUDe)	13
3	14
Μοντέλο Συστήματος και Μεθοδολογία	14
3.1 Περιγραφή Σεναρίου και Τοπολογία Δικτύου	15
3.1.1 Ορισμός Απόστασης, Pathloss, Λαμβανόμενης Ισχύος και Uplink SINR	15
3.1.2 Μοντέλο Διάδοσης και Παρεμβολών	16
3.1.3 Υπολογισμός Uplink Χωρητικότητας.....	17
3.2 Πολιτικές Σύνδεσης Χρηστών	17
3.2.1 Coupled Baseline (DL-based Association με CRE bias)	18
3.2.2 DUDe (Decoupled Uplink and Downlink Access).....	18
3.3 Δημιουργία Συνόλου Δεδομένων και Επιλογή Χαρακτηριστικών	18
3.4 Πίνακας Μεταβλητών Υλοποίησης	19
4	22
Μη Εποπτευόμενη Μάθηση & Clustering	22
4.1 Μη Εποπτευόμενη Μάθηση – Θεωρητικό Υπόβαθρο	23
4.2 Μέθοδοι Ομαδοποίησης (Clustering)	24
4.2.1 K-Means και παραλλαγές	26
4.2.2 Ιεραρχική Ομαδοποίηση (Hierarchical Clustering)	28
4.2.3 Gaussian Mixture Models (GMM)	29
4.2.4 DBSCAN.....	31
4.2.5 Spectral Clustering.....	32
4.2.6 Birch (Balanced Iterative Reducing and Clustering using Hierarchies)	34
4.2.7 OPTICS (Ordering Points To Identify the Clustering Structure)	35

4.2.8 Isolation Forest (Ανίχνευση ανωμαλιών)	35
4.3 Μετρικές Αξιολόγησης Clustering	36
4.3.1 Εσωτερικές Μετρικές	37
4.3.2 Εξωτερικές Μετρικές	38
4.3.3 Ανάλυση Κύριων Συνιστωσών (PCA)	39
4.4 Συγκριτική Ανάλυση Μετρικών Clustering	39
4.4.1 Σύνολο δεομένων και διαδικασία ομαδοποίησης	41
4.4.2 Συγκριτική αξιολόγηση αλγορίθμων clustering και τελική επιλογή	42
4.4.3 Ποσοτική αξιολόγηση ποιότητας clustering	43
4.4.4 Τελική επιλογή αλγορίθμου clustering	44
5	46
Πειραματικά Αποτελέσματα	46
5.1 Απόδοση uplink στο βασικό σενάριο (Baseline)	48
5.2 Βελτίωση uplink μέσω DUDe	49
5.3 Απόδοση μη εποπτευόμενης προσέγγισης ML-DUDe	51
5.4 Δομή δεδομένων και ερμηνεία των αποτελεσμάτων	55
5.5 Ερμηνευσιμότητα των αποτελεσμάτων	58
5.6 Συνοπτική συζήτηση	59
6	61
Συμπεράσματα & Μελλοντική Εργασία	61
6.1 Συμπεράσματα	61
6.2 Μελλοντική Εργασία	63
Βιβλιογραφία	65
Παράρτημα: Κώδικας της Εργασίας	67

Λίστα Εικονών

Εικόνα 1: Αποτελέσματα σύγκρισης Silhouette.....	43
Εικόνα 2: Αποτελέσματα σύγκρισης Davies-Bouldin	44
Εικόνα 3: Τοπολογία δικτύου υλοποίησης.....	47
Εικόνα 4: Σύγκριση χωρητικότητας uplink.....	48
Εικόνα 5: CDF της uplink χωρητικότητας για Baseline (Macro UL) και DUDe (Pico UL)	49
Εικόνα 6: Μείωση φορτίου macro σταθμού βάσης στο uplink μέσω DUDe	50
Εικόνα 7: Σύγκριση της uplink χωρητικότητας για τα σενάρια Baseline, DUDe και ML-DUDe	52
Εικόνα 8: CDF της uplink χωρητικότητας για τα σενάρια Baseline, DUDe και ML-DUDe	53
Εικόνα 9: Μείωση φορτίου macro σταθμού βάσης στο uplink για τα σενάρια Baseline, DUDe και ML-DUDe ..	54
Εικόνα 10: Ομαδοποίηση χρηστών στο uplink σε προβολή PCA	55
Εικόνα 11: Προβολή PCA των χρηστών με χρωματική ένδειξη των ομάδων K-Means	56
Εικόνα 12: Μέσες τιμές χαρακτηριστικών ανά ομάδα χρηστών (cluster profiling) για τον αλγόριθμο K-Means..	57
Εικόνα 13: Σχέση πλεονεκτήματος SINR και κέρδους uplink χωρητικότητας.....	58
Εικόνα 14: Σταθερότητα πολιτικής συσχέτισης (flip rate) για DUDe και ML-DUDe.....	59

Λίστα Πινάκων

Πίνακας 1: Μεταβλητές πειραμάτων.....	19
Πίνακας 2: Σύνολο χαρακτηριστικών που χρησιμοποιούνται για την ομαδοποίηση.....	24
Πίνακας 3: Συγκριτική παρουσίαση αλγορίθμων ομαδοποίησης.....	42

Συνομογραφίες

Ακρωνόμιο	Επεξήγηση
5G	Fifth Generation mobile communication systems
BS	Base Station
UE	User Equipment
UL	Uplink
DL	Downlink
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
HetNet	Heterogeneous Network
Macro BS	Macro Base Station
Pico BS	Pico Base Station
CRE	Cell Range Expansion
DUDe	Decoupled Uplink and Downlink Access
ML	Machine Learning
UL-DUDe	Uplink Decoupled Access
K-Means	K-Means Clustering
GMM	Gaussian Mixture Model
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
OPTICS	Ordering Points To Identify the Clustering Structure
BIRCH	Balanced Iterative Reducing and Clustering using Hierarchies
PCA	Principal Component Analysis
QoS	Quality of Service
IoT	Internet of Things
PDF	Probability Density Function
CDF	Cumulative Distribution Function
dB	Decibel
dBm	Decibel referenced to one milliwatt

1

Εισαγωγή

Τα σύγχρονα δίκτυα κινητών επικοινωνιών βρίσκονται σε μια περίοδο ραγδαίας εξέλιξης, με βασικό στόχο την υποστήριξη ολοένα αυξανόμενων απαιτήσεων σε χωρητικότητα, αξιοπιστία και ποιότητα υπηρεσίας. Η ανάπτυξη νέων εφαρμογών και υπηρεσιών, όπως το Internet of Things (IoT), η μηχανική επικοινωνία (machine-type communications) και οι εφαρμογές πραγματικού χρόνου, έχει μεταβάλει ουσιαστικά τα πρότυπα χρήσης των δικτύων, αυξάνοντας σημαντικά την κίνηση δεδομένων που κατευθύνεται από τους χρήστες προς το δίκτυο.

Παράλληλα, η αρχιτεκτονική των κυψελωτών δικτύων εξελίσσεται μέσω της υιοθέτησης ετερογενών δομών, όπου συνυπάρχουν σταθμοί βάσης διαφορετικής ισχύος και εμβέλειας, με σκοπό την αποδοτικότερη εκμετάλλευση των διαθέσιμων ραδιοπόρων. Στο πλαίσιο αυτό, η αποτελεσματική διαχείριση της σύνδεσης των χρηστών με το δίκτυο αναδεικνύεται σε κρίσιμο ζήτημα για τη συνολική απόδοση του συστήματος.

Η παρούσα διπλωματική εργασία εστιάζει στο πρόβλημα της σύνδεσης χρηστών στο uplink σε ετερογενή κυψελωτά δίκτυα, εξετάζοντας τόσο παραδοσιακές όσο και σύγχρονες προσεγγίσεις, καθώς και τη δυνατότητα αξιοποίησης τεχνικών μηχανικής μάθησης για τη βελτίωση της αποδοτικότητας και της σταθερότητας των σχετικών αποφάσεων

1.1 Σημασία του προβλήματος

Η ραγδαία αύξηση της κίνησης δεδομένων στα σύγχρονα κυψελωτά δίκτυα, σε συνδυασμό με την εξάπλωση υπηρεσιών που βασίζονται σε συσκευές περιορισμένης ισχύος, όπως εφαρμογές Internet of Things (IoT) και γενικότερα uplink-intensive υπηρεσίες, έχει αναδείξει το uplink ως κρίσιμο παράγοντα απόδοσης των δικτύων κινητών επικοινωνιών [1], [2]. Σε αντίθεση με το

downlink, όπου η ισχύς εκπομπής των σταθμών βάσης είναι σημαντικά μεγαλύτερη, το uplink περιορίζεται από τη διαθέσιμη ισχύ των τερματικών χρηστών, γεγονός που οδηγεί συχνά σε φαινόμενα συμφόρησης και ανισορροπίας φορτίου [3].

Στα παραδοσιακά κυψελωτά δίκτυα, η σύνδεση των χρηστών στο uplink και στο downlink πραγματοποιείται με συζευγμένο τρόπο, βασιζόμενη κυρίως στην ισχύ του λαμβανόμενου σήματος στο downlink. Η προσέγγιση αυτή, αν και απλοποιεί τη διαχείριση του δικτύου, αγνοεί τις ιδιαιτερότητες του uplink καναλιού και μπορεί να οδηγήσει σε μη αποδοτική αξιοποίηση των διαθέσιμων πόρων. Το πρόβλημα αυτό καθίσταται ιδιαίτερα έντονο σε ετερογενή κυψελωτά δίκτυα, όπου συνυπάρχουν μακροκυψέλες και μικρότερης εμβέλειας σταθμοί βάσης [4], [5].

Η εισαγωγή ετερογενών δικτύων με macro και small cells στοχεύει στη βελτίωση της χωρητικότητας και της κάλυψης του συστήματος. Ωστόσο, χωρίς κατάλληλες πολιτικές σύνδεσης χρηστών στο uplink, τα αναμενόμενα οφέλη δεν αξιοποιούνται πλήρως, καθώς μεγάλος αριθμός χρηστών συνεχίζει να εξυπηρετείται από macro σταθμούς βάσης, παρά την ύπαρξη πλησιέστερων small cells με ευνοϊκότερες συνθήκες uplink [5]. Ως εκ τούτου, το ζήτημα της αποδοτικής σύνδεσης χρηστών στο uplink αναδεικνύεται σε κεντρικό πρόβλημα για την επίτευξη υψηλής απόδοσης και ισορροπημένης κατανομής φορτίου στα σύγχρονα δίκτυα κινητών επικοινωνιών.

1.2 Στόχοι της Εργασίας

Βασικός στόχος της παρούσας διπλωματικής εργασίας είναι η μελέτη και αξιολόγηση πολιτικών σύνδεσης χρηστών στο uplink σε ετερογενή κυψελωτά δίκτυα, με έμφαση στη μέθοδο Decoupled Uplink and Downlink Access (DUDe) [6], [7]. Στο πλαίσιο αυτό, επιδιώκεται η κατανόηση του τρόπου με τον οποίο η αποσύνδεση της διαδικασίας σύνδεσης uplink και downlink μπορεί να συμβάλει στη βελτίωση της απόδοσης και στη μείωση της συμφόρησης στο uplink.

Ειδικότερα, η εργασία στοχεύει στην υλοποίηση ενός ρεαλιστικού σεναρίου ετερογενούς κυψελωτού δικτύου που περιλαμβάνει macro και pico σταθμούς βάσης, με σκοπό τη μοντελοποίηση της συμπεριφοράς του uplink σε περιβάλλοντα πολλαπλών επιπέδων κάλυψης.

Παράλληλα, αξιολογείται η απόδοση της παραδοσιακής συζευγμένης πολιτικής σύνδεσης uplink–downlink και πραγματοποιείται σύγκριση με τη μέθοδο DUDe ως προς τη χωρητικότητα και την κατανομή φορτίου στο uplink.

Επιπλέον, διερευνάται η δυνατότητα αξιοποίησης τεχνικών μη εποπτευόμενης μηχανικής μάθησης για την προσέγγιση της πολιτικής DUDe, μέσω της ανάλυσης χαρακτηριστικών που σχετίζονται με την απόσταση, το SINR και τη χωρητικότητα στο uplink [8]. Στο πλαίσιο αυτό, εξετάζεται η συμπεριφορά διαφορετικών κατηγοριών χρηστών και αξιολογείται η σταθερότητα και η πρακτική εφαρμόσιμότητα των αποφάσεων σύνδεσης που προκύπτουν από τις τεχνικές μηχανικής μάθησης.

Τέλος, στόχος της εργασίας είναι η εξαγωγή συμπερασμάτων σχετικά με τα οφέλη και τους περιορισμούς της αποσυνδεδεμένης σύνδεσης στο uplink, καθώς και η ανάδειξη πιθανών κατευθύνσεων για μελλοντική έρευνα στον τομέα της αποδοτικής διαχείρισης πόρων σε ετερογενή κυψελωτά δίκτυα [9].

1.3 Διάρθρωση της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία οργανώνεται ως εξής. Στο Κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο που σχετίζεται με τα ετερογενή κυψελωτά δίκτυα και το πρόβλημα της σύνδεσης χρηστών στο uplink, με ανάλυση βασικών εννοιών όπως ο λόγος σήματος προς παρεμβολή και θόρυβο (SINR), η κατανομή φορτίου και η μέθοδος Decoupled Uplink and Downlink Access (DUDe). Στο Κεφάλαιο 3 περιγράφεται η μεθοδολογία που ακολουθείται, καθώς και το μοντέλο δικτύου και οι παράμετροι προσομοίωσης που χρησιμοποιούνται για την αξιολόγηση των εξεταζόμενων πολιτικών σύνδεσης. Στο Κεφάλαιο 4 εξετάζεται αναλυτικά η χρήση αλγορίθμων μη εποπτευόμενης μηχανικής μάθησης για την υποστήριξη της διαδικασίας σύνδεσης χρηστών στο uplink και τεκμηριώνεται η επιλογή της καταλληλότερης μεθόδου clustering για την υλοποίηση της πολιτικής ML-DUDe. Στο Κεφάλαιο 5 παρουσιάζονται και αναλύονται τα αποτελέσματα της προσομοίωσης για τη συζευγμένη πολιτική σύνδεσης, το DUDe και τις προσεγγίσεις μηχανικής μάθησης. Τέλος, στο Κεφάλαιο 6 συνοψίζονται τα βασικά συμπεράσματα της εργασίας και παρουσιάζονται προτάσεις για μελλοντική έρευνα στον τομέα της αποδοτικής διαχείρισης πόρων σε ετερογενή κυψελωτά δίκτυα.

2

Θεωρητικό υπόβαθρο

Στο παρόν κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο που απαιτείται για την κατανόηση του προβλήματος της σύνδεσης χρηστών στο uplink σε ετερογενή κυψελωτά δίκτυα. Αρχικά, εισάγονται βασικές έννοιες που αφορούν την αρχιτεκτονική των ετερογενών δικτύων και τη συνύπαρξη macro και μικρότερων σταθμών βάσης. Στη συνέχεια, αναλύονται οι διαφορές μεταξύ uplink και downlink, καθώς και τα βασικά μεγέθη απόδοσης, όπως το SINR και η χωρητικότητα, με έμφαση στους περιορισμούς του uplink. Τέλος, παρουσιάζεται η μέθοδος Decoupled Uplink and Downlink Access (DUDe), η οποία αποτελεί τη βάση της μελέτης που ακολουθεί.

2.1 Ετερογενή Κυψελωτά Δίκτυα στα Συστήματα 5G

Τα ετερογενή κυψελωτά δίκτυα αποτελούν βασικό δομικό στοιχείο των σύγχρονων δικτύων πέμπτης γενιάς (5G), με στόχο την αντιμετώπιση της αυξανόμενης ζήτησης για υψηλότερη χωρητικότητα, χαμηλότερη καθυστέρηση και βελτιωμένη ποιότητα υπηρεσίας. Σε αντίθεση με τα παραδοσιακά ομοιογενή δίκτυα, τα 5G ετερογενή δίκτυα βασίζονται στη συνύπαρξη σταθμών βάσης διαφορετικών χαρακτηριστικών, όπως macro και μικρότερης εμβέλειας σταθμών βάσης (small cells), οι οποίοι λειτουργούν συμπληρωματικά στο ίδιο δίκτυο πρόσβασης [4].

Οι macro σταθμοί βάσης παρέχουν ευρεία κάλυψη και εξασφαλίζουν τη βασική συνδεσιμότητα του δικτύου, ωστόσο η μεγάλη περιοχή κάλυψής τους οδηγεί συχνά σε αυξημένο φόρτο,

ιδιαίτερα σε περιοχές με υψηλή πυκνότητα χρηστών. Στο πλαίσιο των δικτύων 5G, η πυκνή ανάπτυξη small cells, όπως οι pico σταθμοί βάσης, χρησιμοποιείται για την αύξηση της χωρητικότητας και την εξυπηρέτηση τοπικών απαιτήσεων κίνησης. Η αρχιτεκτονική αυτή επιτρέπει την αποφόρτιση του macro επιπέδου και τη βελτίωση της συνολικής απόδοσης του δικτύου.

Παρά τα πλεονεκτήματα που προσφέρουν τα ετερογενή δίκτυα 5G, η ταυτόχρονη λειτουργία πολλαπλών επιπέδων κάλυψης δημιουργεί νέες προκλήσεις, κυρίως όσον αφορά τη διαχείριση παρεμβολών και την αποδοτική σύνδεση χρηστών. Οι παραδοσιακές πολιτικές σύνδεσης, οι οποίες βασίζονται κυρίως σε μετρήσεις του downlink, δεν λαμβάνουν υπόψη τις ιδιαιτερότητες του uplink, γεγονός που μπορεί να οδηγήσει σε μη αποδοτική κατανομή φορτίου. Το ζήτημα αυτό αποκτά ιδιαίτερη σημασία στα 5G δίκτυα, όπου η υποστήριξη εφαρμογών με υψηλές απαιτήσεις στο uplink, όπως το Internet of Things και υπηρεσίες χαμηλής καθυστέρησης, καθιστά την αποδοτική διαχείριση του uplink κρίσιμη.

Η κατανόηση της αρχιτεκτονικής και των προκλήσεων των ετερογενών κυψελωτών δικτύων στα συστήματα 5G αποτελεί βασική προϋπόθεση για τη μελέτη εναλλακτικών πολιτικών σύνδεσης χρηστών στο uplink, όπως η μέθοδος Decoupled Uplink and Downlink Access (DUDe), η οποία αναλύεται σε επόμενες ενότητες.

2.2 Διαφορές Uplink και Downlink στα Δίκτυα 5G

Στα κυψελωτά δίκτυα κινητών επικοινωνιών, το uplink και το downlink παρουσιάζουν θεμελιώδεις διαφορές ως προς τα χαρακτηριστικά και τους περιορισμούς τους, οι οποίες επηρεάζουν άμεσα την απόδοση και τη διαδικασία σύνδεσης χρηστών. Στο downlink, οι σταθμοί βάσης διαθέτουν υψηλή ισχύ εκπομπής και προηγμένες τεχνικές επεξεργασίας σήματος, γεγονός που επιτρέπει την επίτευξη ευρείας κάλυψης και υψηλών ρυθμών μετάδοσης. Αντίθετα, στο uplink η ισχύς εκπομπής περιορίζεται από τις δυνατότητες των τερματικών χρηστών, καθιστώντας το uplink πιο ευάλωτο σε παρεμβολές και απώλειες σήματος [2].

Η ασυμμετρία αυτή αποκτά ιδιαίτερη σημασία στα δίκτυα 5G, όπου παρατηρείται αυξημένη ζήτηση για uplink-intensive εφαρμογές, όπως υπηρεσίες Internet of Things, εφαρμογές χαμηλής καθυστέρησης και σενάρια μαζικής σύνδεσης συσκευών. Σε τέτοια περιβάλλοντα, η απόδοση του uplink συχνά καθορίζει τη συνολική εμπειρία χρήστη και τη λειτουργική

αποδοτικότητα του δικτύου, ακόμη και όταν το downlink παρουσιάζει ικανοποιητικές επιδόσεις.

Στις παραδοσιακές πολιτικές σύνδεσης χρηστών, η επιλογή του σταθμού βάσης βασίζεται κυρίως σε μετρήσεις του downlink, όπως η ισχύς του λαμβανόμενου σήματος. Η προσέγγιση αυτή οδηγεί συχνά τους χρήστες να συνδέονται με macro σταθμούς βάσης, οι οποίοι προσφέρουν ισχυρότερο downlink σήμα, ακόμη και όταν υπάρχουν πλησιέστεροι small cells που θα μπορούσαν να εξυπηρετήσουν αποτελεσματικότερα το uplink. Το αποτέλεσμα είναι η δημιουργία συμφόρησης στο uplink του macro επιπέδου και η μη αποδοτική αξιοποίηση των διαθέσιμων πόρων του δικτύου.

Η κατανόηση των διαφορών μεταξύ uplink και downlink αποτελεί βασικό στοιχείο για τον σχεδιασμό αποδοτικών πολιτικών σύνδεσης χρηστών στα δίκτυα 5G. Η ανάγκη για μηχανισμούς που λαμβάνουν υπόψη τα χαρακτηριστικά του uplink οδηγεί στην εξέταση εναλλακτικών προσεγγίσεων, όπως η αποσυνδεδεμένη σύνδεση uplink και downlink, οι οποίες επιδιώκουν την εξισορρόπηση του φορτίου και τη βελτίωση της συνολικής απόδοσης του δικτύου.

2.3 Μέτρα Απόδοσης: SINR και Χωρητικότητα

Η αξιολόγηση της απόδοσης των κυψελωτών δικτύων βασίζεται σε συγκεκριμένα μεγέθη, τα οποία αποτυπώνουν την ποιότητα της επικοινωνίας και την αποδοτική αξιοποίηση των διαθέσιμων πόρων. Στο πλαίσιο της παρούσας εργασίας, βασικό ρόλο διαδραματίζουν ο λόγος σήματος προς παρεμβολή και θόρυβο (Signal-to-Interference-plus-Noise Ratio – SINR) και η χωρητικότητα του καναλιού, καθώς αποτελούν άμεσα συνδεδεμένα μεγέθη με την απόδοση του uplink.

Το SINR εκφράζει την ποιότητα του λαμβανόμενου σήματος και ορίζεται ως ο λόγος της ισχύος του επιθυμητού σήματος προς το άθροισμα της ισχύος των παρεμβολών και του θερμικού θορύβου. Στο uplink, το SINR επηρεάζεται έντονα από την περιορισμένη ισχύ εκπομπής των χρηστών, τη γεωμετρική απόσταση από τον σταθμό βάσης και το επίπεδο παρεμβολών που προκαλούν ταυτόχρονες μεταδόσεις άλλων χρηστών. Ως αποτέλεσμα, μικρές μεταβολές στη θέση ή στη σύνδεση ενός χρήστη μπορούν να οδηγήσουν σε σημαντικές διαφοροποιήσεις στο uplink SINR.

Η χωρητικότητα του καναλιού αποτελεί μέτρο της μέγιστης θεωρητικά εφικτής ταχύτητας μετάδοσης δεδομένων και συνδέεται άμεσα με το SINR μέσω του τύπου του Shannon [1]. Υψηλότερες τιμές SINR οδηγούν σε αυξημένη χωρητικότητα, ενώ χαμηλές τιμές περιορίζουν σημαντικά την απόδοση της σύνδεσης. Στα ετερογενή δίκτυα, η επιλογή του σταθμού βάσης στο uplink επηρεάζει καθοριστικά τη χωρητικότητα, καθώς η σύνδεση με έναν πλησιέστερο small cell μπορεί να προσφέρει σημαντικά καλύτερο SINR σε σύγκριση με τη σύνδεση σε έναν απομακρυσμένο macro σταθμό βάσης.

Τα μεγέθη SINR και χωρητικότητας αποτελούν, επομένως, κρίσιμους δείκτες για την αξιολόγηση πολιτικών σύνδεσης χρηστών στο uplink. Η ανάλυσή τους επιτρέπει την ποσοτική σύγκριση διαφορετικών προσεγγίσεων, όπως της συζευγμένης σύνδεσης και της μεθόδου DUDe, και παρέχει σαφή εικόνα των επιπτώσεων της επιλογής σταθμού βάσης στην απόδοση και την κατανομή φορτίου του δικτύου.

2.4 Κατανομή Φορτίου και Πρόβλημα Σύνδεσης Χρηστών

Η κατανομή φορτίου αποτελεί κρίσιμο ζήτημα στη λειτουργία των ετερογενών κυψελωτών δικτύων, καθώς επηρεάζει άμεσα τόσο την ποιότητα υπηρεσίας των χρηστών όσο και τη συνολική αποδοτικότητα του δικτύου. Ως φορτίο ορίζεται το επίπεδο χρήσης των διαθέσιμων πόρων ενός σταθμού βάσης, το οποίο εξαρτάται από τον αριθμό των εξυπηρετούμενων χρηστών και τις απαιτήσεις τους σε ρυθμό μετάδοσης. Σε δίκτυα με συνύπαρξη macro και small cells, η μη ισορροπημένη κατανομή φορτίου μπορεί να οδηγήσει σε φαινόμενα συμφόρησης, ακόμη και όταν υπάρχουν διαθέσιμοι πόροι σε γειτονικούς σταθμούς βάσης.

Στις συζευγμένες πολιτικές σύνδεσης, η επιλογή σταθμού βάσης πραγματοποιείται με βάση μετρήσεις του downlink, γεγονός που οδηγεί σε συγκέντρωση μεγάλου αριθμού χρηστών στο macro επίπεδο. Η πρακτική αυτή δημιουργεί συνθήκες αυξημένου φορτίου στο uplink των macro σταθμών βάσης και ταυτόχρονα περιορισμένη αξιοποίηση των small cells, ιδιαίτερα σε ετερογενή δίκτυα 5G. Ως αποτέλεσμα, η κατανομή των διαθέσιμων ραδιοπόρων καθίσταται μη ισορροπημένη και επηρεάζεται αρνητικά η συνολική απόδοση του συστήματος [3]. Για την αντιμετώπιση του προβλήματος απαιτούνται πολιτικές σύνδεσης που λαμβάνουν υπόψη τα χαρακτηριστικά του uplink και επιτρέπουν πιο αποτελεσματική κατανομή φορτίου μεταξύ macro και small cells, οδηγώντας σε εναλλακτικές προσεγγίσεις όπως η αποσυνδεδεμένη σύνδεση uplink και downlink που παρουσιάζεται στη συνέχεια.

2.5 Decoupled Uplink and Downlink Access (DUDe)

Η μέθοδος Decoupled Uplink and Downlink Access (DUDe) αποτελεί μια εναλλακτική προσέγγιση στη διαδικασία σύνδεσης χρηστών στα κυψελωτά δίκτυα, η οποία στοχεύει στην αντιμετώπιση των περιορισμών της παραδοσιακής συζευγμένης σύνδεσης uplink και downlink. Σε αντίθεση με την κλασική σύνδεση του χρήστη, όπου μοναδικό κριτήριο σύνδεσης (τόσο στο downlink όσο και στο uplink) αποτελεί η λαμβανόμενη ισχύς στην downlink κατεύθυνση, το DUDe επιτρέπει τη σύνδεση του χρήστη σε διαφορετικούς σταθμούς βάσης για το uplink και το downlink, λαμβάνοντας υπόψη τα ιδιαίτερα χαρακτηριστικά κάθε κατεύθυνσης μετάδοσης [7].

Στα ετερογενή δίκτυα 5G, η εφαρμογή του DUDe μπορεί να οδηγήσει σε ουσιαστική αποσυμφόρηση του macro επιπέδου στο uplink και σε καλύτερη αξιοποίηση των small cells. Η ανεξάρτητη επιλογή σταθμού βάσης για το uplink επιτρέπει την πιο ισορροπημένη κατανομή φορτίου και τη βελτίωση της συνολικής χωρητικότητας του δικτύου, χωρίς να απαιτούνται αυθαίρετες παρεμβάσεις ή τεχνητή ανακατανομή χρηστών.

Παρά τα πλεονεκτήματα που προσφέρει, η υλοποίηση του DUDe εισάγει νέες προκλήσεις, όπως η αυξημένη πολυπλοκότητα στη διαχείριση συνδέσεων και η ανάγκη για κατάλληλους μηχανισμούς σηματοδότησης μεταξύ των σταθμών βάσης. Για τον λόγο αυτό, η μελέτη και η αξιολόγηση του DUDe σε ρεαλιστικά σενάρια αποτελεί αντικείμενο ενεργούς έρευνας, ενώ εξετάζονται και εναλλακτικές προσεγγίσεις που αξιοποιούν τεχνικές μηχανικής μάθησης για την υποστήριξη της διαδικασίας σύνδεσης χρηστών στο uplink.

3

Μοντέλο Συστήματος και Μεθοδολογία

Στο παρόν κεφάλαιο περιγράφεται η μεθοδολογία που ακολουθείται για τη μελέτη και αξιολόγηση των πολιτικών σύνδεσης χρηστών στο uplink σε ετερογενή κυψελωτά δίκτυα. Με βάση το θεωρητικό υπόβαθρο που παρουσιάστηκε στο προηγούμενο κεφάλαιο, ορίζεται το μοντέλο του δικτύου, οι παραδοχές που υιοθετούνται, καθώς και οι βασικές παράμετροι προσομοίωσης που χρησιμοποιούνται για την ανάλυση της απόδοσης.

Αρχικά, παρουσιάζεται η αρχιτεκτονική του εξεταζόμενου ετερογενούς κυψελωτού δικτύου, το οποίο περιλαμβάνει macro και pico σταθμούς βάσης, καθώς και ο τρόπος χωρικής κατανομής των χρηστών. Στη συνέχεια, περιγράφονται τα μοντέλα καναλιού, παρεμβολών και υπολογισμού του λόγου σήματος προς παρεμβολή και θόρυβο (SINR), τα οποία χρησιμοποιούνται για την εκτίμηση της uplink χωρητικότητας.

Επιπλέον, ορίζονται οι πολιτικές σύνδεσης χρηστών που εξετάζονται, συμπεριλαμβανομένης της παραδοσιακής συζευγμένης πολιτικής, της μεθόδου Decoupled Uplink and Downlink Access (DUDe), καθώς και της προτεινόμενης μη εποπτευόμενης προσέγγισης μηχανικής μάθησης (ML-DUDe). Τέλος, παρουσιάζεται η διαδικασία συλλογής και επεξεργασίας των δεδομένων που προκύπτουν από τις προσομοιώσεις και καθορίζονται οι μετρικές απόδοσης που χρησιμοποιούνται για τη συγκριτική αξιολόγηση των εξεταζόμενων πολιτικών.

3.1 Περιγραφή Σεναρίου και Τοπολογία Δικτύου

Η τοπολογία που χρησιμοποιείται στο προτεινόμενο μοντέλο είναι ένα ασύμμετρο ετερογενές κυψελωτό δίκτυο (HetNet), αποτελούμενο από έναν μακροκυψελοειδή σταθμό βάσης (macro BS) και εννέα μικροκυψελοειδείς σταθμούς βάσης (pico BS). Οι pico σταθμοί βάσης τοποθετούνται σε διάταξη πλέγματος 3×3 εντός τετραγωνικής περιοχής πλευράς 500 m, ώστε να εξασφαλίζεται ομοιόμορφη κάλυψη. Ο macro σταθμός βάσης βρίσκεται στο κέντρο της περιοχής.

Στο εξεταζόμενο σενάριο τοποθετούνται 200 χρήστες (UEs), οι οποίοι κατανέμονται τυχαία στον χώρο προσομοίωσης. Το ύψος όλων των σταθμών βάσης θεωρείται ίσο με 25 m. Οι ίδιες παραδοχές τοπολογίας και παραμέτρων χρησιμοποιούνται σε όλα τα σενάρια σύνδεσης, ώστε να διασφαλίζεται η συγκρισιμότητα των αποτελεσμάτων.

3.1.1 Ορισμός Απόστασης, Pathloss, Λαμβανόμενης Ισχύος και Uplink SINR

Η απόσταση μεταξύ κάθε χρήστη i και του macro ή των pico σταθμών βάσης υπολογίζεται με βάση την ευκλείδεια απόσταση στον τρισδιάστατο χώρο. Συγκεκριμένα, η απόσταση από τον macro σταθμό βάσης δίνεται από:

$$d_{m,i} = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2 + h_{bs}^2} \quad (3.1) [2]$$

Η εξίσωση (3.1) υπολογίζει την ευκλείδεια απόσταση στον τρισδιάστατο χώρο μεταξύ του χρήστη i και του macro σταθμού βάσης, λαμβάνοντας υπόψη και το ύψος του σταθμού βάσης. ενώ η απόσταση από έναν pico σταθμό βάσης υπολογίζεται ως:

$$d_{p,i} = \sqrt{(x_i - x_p)^2 + (y_i - y_p)^2 + h_{bs}^2} \quad (3.2) [2]$$

όπου (x_i, y_i) είναι οι συντεταγμένες του χρήστη, (x_m, y_m) και (x_p, y_p) οι συντεταγμένες του macro και pico σταθμού βάσης αντίστοιχα και h_{bs} το ύψος των σταθμών βάσης.

Στο σενάριο DUDe, η σύνδεση ενός χρήστη με pico σταθμό βάσης επιτρέπεται μόνο εφόσον η απόσταση από τον πλησιέστερο pico σταθμό βάσης είναι μικρότερη ή ίση με 180 m, ώστε να διασφαλίζεται ευνοϊκή uplink σύνδεση.

3.1.2 Μοντέλο Διάδοσης και Παρεμβολών

Η απώλεια διάδοσης στο uplink μοντελοποιείται με λογαριθμικό μοντέλο εξασθένησης:

$$PL = PL_0 + 10n \log_{10}(d) \quad (3.3) \quad [1]$$

όπου PL_0 η απώλεια αναφοράς σε απόσταση d η απόσταση μεταξύ χρήστη και σταθμού βάσης και n ο εκθέτης εξασθένησης του καναλιού.

Η εξίσωση (3.3) περιγράφει την απώλεια διάδοσης στο uplink σύμφωνα με το λογαριθμικό μοντέλο εξασθένησης, όπου η απώλεια εξαρτάται από την απόσταση και τον εκθέτη εξασθένησης του καναλιού.

Η λαμβανόμενη ισχύς στο uplink από έναν χρήστη προς τον macro και τον pico σταθμό βάσης αντίστοιχα δίνεται από:

$$PR_{macro} = P_{UE} + G_{UE} + G_{macro} - PL_{macro} \quad (3.4) \quad [2],[10]$$

όπου P_{UE} η ισχύς εκπομπής του χρήστη, G_{UE} το κέρδος κεραίας του χρήστη, G_{macro} το κέρδος κεραίας του macro σταθμού βάσης και PL_{macro} η απώλεια διάδοσης μεταξύ χρήστη και macro σταθμού βάσης. Η εξίσωση (3.4) εκφράζει το uplink link budget και υπολογίζει τη λαμβανόμενη ισχύ στο macro σταθμό βάσης, λαμβάνοντας υπόψη την ισχύ εκπομπής, τα κέρδη κεραίων και την απώλεια διάδοσης.

$$PR_{pico} = P_{UE} + G_{UE} + G_{pico} - PL_{pico} \quad (3.5) \quad [2],[10]$$

όπου PR_{pico} είναι η λαμβανόμενη ισχύς στο uplink στον pico σταθμό βάσης, G_{pico} το κέρδος κεραίας του pico σταθμού βάσης και PL_{pico} η αντίστοιχη απώλεια διάδοσης. Η εξίσωση (3.5) υπολογίζει τη λαμβανόμενη ισχύ στο uplink προς pico σταθμό βάσης, ακολουθώντας το ίδιο μοντέλο link budget με το macro επίπεδο.

3.1.3 Υπολογισμός Uplink Χωρητικότητας

Ο λόγος σήματος προς παρεμβολή και θόρυβο (SINR) στο uplink υπολογίζεται στην κλίμακα dB ως:

$$SINR_{macro} = PR_{macro} - NOISE_{dBm} - I_{macro} \quad (3.6) [11]$$

όπου $P_{R,macro}$ η λαμβανόμενη ισχύς στο uplink στον macro σταθμό βάσης, $NOISE_{dBm}$ ο θερμικός θόρυβος του συστήματος σε dBm και I_{macro} η συνολική παρεμβολή που δέχεται ο macro σταθμός βάσης στο uplink. Η εξίσωση (3.6) ορίζει τον λόγο σήματος προς παρεμβολή και θόρυβο (SINR) στο uplink για macro σταθμό βάσης σε κλίμακα dB.

Αντίστοιχα για pico, έχουμε:

$$SINR_{pico} = PR_{pico} - NOISE_{dBm} - I_{pico} \quad (3.7) [11]$$

όπου $P_{R,pico}$ η λαμβανόμενη ισχύς στο uplink στον pico σταθμό βάσης (σε dBm) και I_{pico} η ισοδύναμη παρεμβολή στο uplink που δέχεται ο pico σταθμός βάσης (σε dBm), ενώ ο $NOISE_{dBm}$ ορίζεται όπως προηγουμένως. Η εξίσωση (3.7) ορίζει τον λόγο σήματος προς παρεμβολή και θόρυβο (SINR) στο uplink για τον pico σταθμό βάσης σε κλίμακα dB.

Η αντίστοιχη uplink χωρητικότητα υπολογίζεται μέσω του τύπου Shannon:

$$C = B \log_2 \left(1 + 10^{\frac{SINR}{10}} \right) \quad (3.8) [12]$$

Η εξίσωση (3.8) υπολογίζει τη θεωρητική χωρητικότητα του uplink καναλιού, με βάση τον τύπο του Shannon, χρησιμοποιώντας ως είσοδο το $SINR$ (σε dB) που προκύπτει από τις εξισώσεις (3.6)–(3.7).

3.2 Πολιτικές Σύνδεσης Χρηστών

Στο πλαίσιο της παρούσας εργασίας εξετάζονται δύο βασικά σενάρια σύνδεσης χρηστών, που είναι τα εξής:

3.2.1 Coupled Baseline (DL-based Association με CRE bias)

Στο βασικό σενάριο, οι χρήστες συνδέονται με τον σταθμό βάσης που προσφέρει τη μεγαλύτερη ισχύ λήψης στο downlink. Στους pico σταθμούς βάσης εφαρμόζεται bias 12 dB για Cell Range Expansion (CRE), με στόχο την ενίσχυση της σύνδεσης με μικροκυψέλες. Η uplink σύνδεση ακολουθεί υποχρεωτικά την επιλογή του downlink (coupled σύνδεση).

3.2.2 DUDe (Decoupled Uplink and Downlink Access)

Στο σενάριο DUDe, η επιλογή σταθμού βάσης για το uplink πραγματοποιείται ανεξάρτητα από το downlink. Ένας χρήστης συνδέεται με pico σταθμό βάσης όταν το uplink SINR προς τον pico υπερβαίνει το αντίστοιχο προς τον macro σταθμό βάσης και η απόσταση από τον πλησιέστερο pico σταθμό βάσης δεν υπερβαίνει τα 180 m. Με τον τρόπο αυτό διασφαλίζεται σύνδεση με κοντινό σταθμό βάσης, βελτιώνοντας την ποιότητα του uplink σήματος και μειώνοντας τη συμφόρηση των macro κυψελών.

Η απόφαση uplink συσχέτισης μπορεί να εκφραστεί συνοπτικά ως:

$$UL_i = \begin{cases} \text{pico}, & \text{αν } \max(\text{SINR}_{\text{pico},i}) > \text{SINR}_{\text{macro},i} \\ \text{macro}, & \text{διαφορετικά} \end{cases} \quad (3.9) [8]$$

Η εξίσωση (3.9) περιγράφει τον κανόνα απόφασης uplink συσχέτισης της τεχνολογίας DUDe, όπου ο χρήστης συνδέεται με τον σταθμό βάσης που προσφέρει καλύτερο uplink SINR.

3.3 Δημιουργία Συνόλου Δεδομένων και Επιλογή Χαρακτηριστικών

Για την αξιολόγηση των πολιτικών σύνδεσης χρηστών δημιουργείται ένα σύνολο δεδομένων που προκύπτει από προσομοιώσεις του εξεταζόμενου ετερογενούς δικτύου. Κάθε εγγραφή αντιστοιχεί σε έναν χρήστη και περιλαμβάνει πληροφορίες σχετικές με τη γεωμετρική του θέση, την ποιότητα της uplink σύνδεσης και την εκτιμώμενη χωρητικότητα.

Το σύνολο δεδομένων περιλαμβάνει χαρακτηριστικά που αποτυπώνουν τις αποστάσεις των χρηστών από τον macro και τους pico σταθμούς βάσης, καθώς και μετρικές άμεσα συνδεδεμένες με την uplink απόδοση, όπως το uplink SINR και η αντίστοιχη χωρητικότητα. Η

επιλογή των χαρακτηριστικών στοχεύει στη διατήρηση ενός ισορροπημένου συνόλου δεδομένων, το οποίο συνδυάζει γεωμετρικές και ραδιοζεύξεις παραμέτρους.

Το παραγόμενο σύνολο δεδομένων χρησιμοποιείται τόσο για τη συγκριτική αξιολόγηση των κλασικών πολιτικών σύνδεσης χρηστών όσο και ως είσοδος στις τεχνικές μη εποπτευόμενης μηχανικής μάθησης που εφαρμόζονται στη συνέχεια. Με τον τρόπο αυτό καθίσταται δυνατή η διερεύνηση του βαθμού στον οποίο μια δεδομενοκεντρική προσέγγιση μπορεί να προσεγγίσει τη συμπεριφορά της πολιτικής DUDe. Η ανάλυση των αποτελεσμάτων παρουσιάζεται στο επόμενο κεφάλαιο.

3.4 Πίνακας Μεταβλητών Υλοποίησης

Οι μεταβλητές που χρησιμοποιήθηκαν στην υλοποίηση και τα πειράματα της εργασίας, δίνονται στον Πίνακα 1 που ακολουθεί.

Πίνακας 1: Μεταβλητές πειραμάτων

Μεταβλητή	Περιγραφή
N_USERS	Αριθμός χρηστών (UEs) στο δίκτυο
R	Ακτίνα περιοχής προσομοίωσης (m)
h_bs	Ύψος σταθμών βάσης (m)
B	Bandwidth uplink (Hz)
NOISE_DBM	Θερμικός θόρυβος για το bandwidth B (dBm)
P_UE_DBM	Ισχύς εκπομπής χρήστη στο uplink (dBm)
G_UE	Κέρδος κεραίας χρήστη (dB)
G_MACRO	Κέρδος κεραίας macro σταθμού βάσης (dB)
G_PICO	Κέρδος κεραίας pico σταθμού βάσης (dB)
N_MACRO	Εκθέτης εξασθένησης καναλιού για macro
N_PICO	Εκθέτης εξασθένησης καναλιού για pico
x, y	Συντεταγμένες χρηστών στην περιοχή
xm, ym	Συντεταγμένες macro σταθμού βάσης
xp, yp	Συντεταγμένες pico σταθμού βάσης
d_macro	Απόσταση χρήστη-macro BS
d_pico	Απόσταση χρήστη-pico BS
PL_macro	Path loss uplink προς macro BS (dB)
PL_pico	Path loss uplink προς pico BS (dB)
PR_macro	Λαμβανόμενη uplink ισχύς στο macro BS (dBm)
PR_pico	Λαμβανόμενη uplink ισχύς στο pico BS (dBm)
I_macro	Παρεμβολή uplink στο macro BS
I_pico	Παρεμβολή uplink στο pico BS

Μεταβλητή	Περιγραφή
SINR_macro	Uplink SINR προς macro BS (dB)
SINR_pico	Uplink SINR προς pico BS (dB)
C_macro	Uplink ρυθμός μετάδοσης προς macro BS (bps)
C_pico	Uplink ρυθμός μετάδοσης προς pico BS (bps)
df	Κύριος πίνακας δεδομένων χρηστών
FEATURE_POOL	Σύνολο υποψήφιων χαρακτηριστικών για clustering
used	Τελικό σύνολο χαρακτηριστικών που χρησιμοποιούνται
X_raw	Αρχικός πίνακας χαρακτηριστικών
X_imp	Πίνακας χαρακτηριστικών μετά από imputation
X_scaled	Κανονικοποιημένα χαρακτηριστικά
PCA1, PCA2	Πρώτη και δεύτερη κύρια συνιστώσα PCA
X_pca	Προβολή δεδομένων στον χώρο PCA
kmeans_cluster	Ετικέτα cluster από KMeans
mbkmeans_cluster	Ετικέτα cluster από MiniBatch KMeans
spectral_cluster	Ετικέτα cluster από Spectral Clustering
agg_cluster	Ετικέτα cluster από Agglomerative (Ward)
gmm_cluster	Ετικέτα cluster από Gaussian Mixture Model
birch_cluster	Ετικέτα cluster από BIRCH
dbscan_cluster	Ετικέτα cluster από DBSCAN
optics_cluster	Ετικέτα cluster από OPTICS
anomaly	Δείκτης ανωμαλίας χρήστη (Isolation Forest)
K	Αριθμός clusters
LINKAGE_METHOD	Μέθοδος σύνδεσης στο hierarchical clustering
TRUNCATE_LEVEL	Επίπεδο αποκοπής dendrogram
N_CLUSTERS_TO_CUT	Αριθμός clusters μετά την αποκοπή
UL_baseline	Baseline uplink συσχέτιση
UL_DUDe	DUDe uplink συσχέτιση
UL_assoc_ML	ML-assisted uplink συσχέτιση
CLUSTER_USED	Clustering που χρησιμοποιείται για ML-DUDe
cluster_policy	Πολιτική uplink ανά cluster
CAP_GAIN_TH	Κατώφλι απαιτούμενου uplink κέρδους
MAX_PICO_FRAC	Μέγιστο επιτρεπτό ποσοστό χρηστών σε pico BS
current_pico_frac	Τρέχον ποσοστό χρηστών σε pico BS
rate	Uplink ρυθμός ανά χρήστη
eff_rate	Αποτελεσματικός uplink ρυθμός (με load)
gain	Σχετικό uplink κέρδος χρήστη
p5_macro, p5_pico	5ο εκατοστημόριο uplink απόδοσης
fair_baseline	Jain fairness για baseline
fair_dude	Jain fairness για DUDe
stat, p	Στατιστικά Wilcoxon ελέγχου
ci_low, ci_high	Κάτω και άνω όριο 95% διαστήματος εμπιστοσύνης
rng	Γεννήτρια τυχαίων αριθμών

Μεταβλητή	Περιγραφή
boot_means	Bootstrap μέσοι όροι
flip_rates	Ποσοστά λανθασμένων αποφάσεων

4

Μη Εποπτευόμενη Μάθηση & Clustering

Στο παρόν κεφάλαιο εξετάζεται η αξιοποίηση τεχνικών μη εποπτευόμενης μηχανικής μάθησης για την υποστήριξη της διαδικασίας σύνδεσης χρηστών στο uplink σε ετερογενή κυψελωτά δίκτυα. Σε αντίθεση με τις παραδοσιακές πολιτικές σύνδεσης, οι οποίες βασίζονται σε ρητούς κανόνες ή πλήρη γνώση του καναλιού, οι προσεγγίσεις μηχανικής μάθησης στοχεύουν στην εξαγωγή δομών και προτύπων απευθείας από τα διαθέσιμα δεδομένα [9].

Με βάση τα δεδομένα που προκύπτουν από το μοντέλο και τις προσομοιώσεις που περιγράφηκαν στο προηγούμενο κεφάλαιο, διαμορφώνεται ένα σύνολο χαρακτηριστικών που περιγράφει τη συμπεριφορά των χρηστών στο uplink, όπως η απόσταση από τους σταθμούς βάσης, ο λόγος σήματος προς παρεμβολή και θόρυβο (SINR) και η επιτυγχανόμενη χωρητικότητα. Τα χαρακτηριστικά αυτά χρησιμοποιούνται ως είσοδος για την εφαρμογή αλγορίθμων clustering, με σκοπό την ομαδοποίηση χρηστών με παρόμοια uplink χαρακτηριστικά.

Στη συνέχεια, παρουσιάζονται και συγκρίνονται διαφορετικοί αλγόριθμοι μη εποπτευόμενης μάθησης ως προς την ικανότητά τους να παράγουν σταθερές και ερμηνεύσιμες ομαδοποιήσεις. Τέλος, τεκμηριώνεται η επιλογή της καταλληλότερης μεθόδου clustering για την υλοποίηση της προτεινόμενης πολιτικής ML-DUDe, καθώς και ο τρόπος με τον οποίο τα αποτελέσματα της ομαδοποίησης αξιοποιούνται για τη λήψη αποφάσεων σύνδεσης χρηστών στο uplink

4.1 Μη Εποπτευόμενη Μάθηση – Θεωρητικό Υπόβαθρο

Η μη εποπτευόμενη μάθηση (unsupervised learning) αποτελεί έναν από τους θεμελιώδεις κλάδους της μηχανικής μάθησης και επικεντρώνεται στην εξαγωγή γνώσης από δεδομένα χωρίς τη διαθεσιμότητα ετικετών ή πληροφορίας σχετικά με την επιθυμητή έξοδο [13]. Σε αντίθεση με την εποπτευόμενη μάθηση, όπου ο αλγόριθμος εκπαιδεύεται στη συσχέτιση εισόδου-εξόδου βάσει γνωστών ετικετών, στη μη εποπτευόμενη μάθηση η ανάλυση είναι διερευνητική και στοχεύει στην αποκάλυψη της λανθάνουσας δομής των δεδομένων.

Κεντρικός στόχος της μη εποπτευόμενης μάθησης είναι η ανακάλυψη κρυφών προτύπων, συσχετίσεων και ομάδων που δεν είναι άμεσα εμφανείς. Στο πλαίσιο αυτό, δύο βασικές κατηγορίες μεθόδων κυριαρχούν: η ομαδοποίηση (clustering), η οποία επιδιώκει τη δημιουργία ομάδων με υψηλή εσωτερική ομοιογένεια και εξωτερική ετερογένεια, και η μείωση διαστασιμότητας (dimensionality reduction), η οποία στοχεύει στη συμπίκνωση της πληροφορίας σε μικρότερο αριθμό διαστάσεων. Η δεύτερη είναι ιδιαίτερα σημαντική σε πολυδιάστατα περιβάλλοντα, όπως τα τηλεπικοινωνιακά συστήματα, όπου κάθε χρήστης περιγράφεται από πλήθος μετρικών καναλιού και απόδοσης [14].

Ένα από τα βασικά ζητήματα στη μη εποπτευόμενη μάθηση είναι η αξιολόγηση των αποτελεσμάτων, καθώς απουσιάζει ένα αντικειμενικό «σωστό» αποτέλεσμα. Για τον λόγο αυτό χρησιμοποιούνται εσωτερικές μετρικές (intrinsic), οι οποίες βασίζονται αποκλειστικά στη γεωμετρία και τη συνοχή των δεδομένων, καθώς και εξωτερικές μετρικές (extrinsic), οι οποίες απαιτούν πρόσθετη πληροφορία αναφοράς.

Η θεωρητική θεμελίωση των μεθόδων clustering βασίζεται σε τρεις κύριους άξονες: (i) τον ορισμό της ομοιότητας και της απόστασης μεταξύ σημείων δεδομένων, (ii) τη συνάρτηση κόστους που ποσοτικοποιεί την ποιότητα μιας ομαδοποίησης και (iii) τις μεθόδους βελτιστοποίησης που χρησιμοποιούνται για την εύρεση αποδεκτών λύσεων. Η επιλογή των παραπάνω παραμέτρων επηρεάζει καθοριστικά το τελικό αποτέλεσμα.

Στον τομέα των τηλεπικοινωνιών, η μη εποπτευόμενη μάθηση βρίσκει εκτεταμένες εφαρμογές, καθώς επιτρέπει την ανάλυση μετρήσεων δικτύου σε πραγματικό χρόνο, την αναγνώριση ομάδων χρηστών με παρόμοια χαρακτηριστικά ποιότητας καναλιού και τον εντοπισμό ανωμαλιών που ενδέχεται να υποδηλώνουν προβλήματα λειτουργίας. Στο πλαίσιο της παρούσας εργασίας, οι τεχνικές clustering αξιοποιούνται για την υποστήριξη της

αποσυνδεδεμένης σύνδεσης uplink και downlink (DUDe), επιτρέποντας μια πιο αποδοτική και σταθερή ανάθεση χρηστών σε macro και pico σταθμούς βάσης.

Στα επόμενα υποκεφάλαια παρουσιάζονται αναλυτικά οι βασικότερες κατηγορίες αλγορίθμων ομαδοποίησης, τα πλεονεκτήματα και οι περιορισμοί τους, καθώς και η καταλληλότητά τους για εφαρμογές σε ετερογενή κυψελωτά δίκτυα.

Το σύνολο χαρακτηριστικών που χρησιμοποιήθηκαν κατά την ομαδοποίηση στην υλοποίηση της εργασίας, παρουσιάζονται στον Πίνακα 2 που ακολουθεί.

Πίνακας 2: Σύνολο χαρακτηριστικών που χρησιμοποιούνται για την ομαδοποίηση

Χαρακτηριστικό	Περιγραφή
Απόσταση από macro σταθμό βάσης	Ευκλείδεια απόσταση μεταξύ χρήστη και macro BS
Απόσταση από pico σταθμό βάσης	Απόσταση μεταξύ χρήστη και πλησιέστερου pico BS
Uplink SINR (macro)	Λόγος σήματος προς παρεμβολή και θόρυβο στο uplink προς macro BS
Uplink SINR (pico)	Λόγος σήματος προς παρεμβολή και θόρυβο στο uplink προς pico BS
Uplink χωρητικότητα (macro)	Χωρητικότητα uplink προς macro BS
Uplink χωρητικότητα (pico)	Χωρητικότητα uplink προς pico BS
PCA συνιστώσα 1	Πρώτη κύρια συνιστώσα μετά από PCA
PCA συνιστώσα 2	Δεύτερη κύρια συνιστώσα μετά από PCA

4.2 Μέθοδοι Ομαδοποίησης (Clustering)

Η ομαδοποίηση δεδομένων, γνωστή διεθνώς ως clustering, αποτελεί μία από τις σημαντικότερες τεχνικές μη εποπτευόμενης μάθησης. Ο βασικός στόχος της είναι η κατηγοριοποίηση ενός συνόλου παρατηρήσεων σε ομάδες με τέτοιο τρόπο ώστε τα στοιχεία της ίδιας ομάδας να παρουσιάζουν υψηλή εσωτερική ομοιότητα, ενώ ταυτόχρονα να διαφέρουν όσο το δυνατόν περισσότερο από τα στοιχεία των άλλων ομάδων [15]. Με αυτόν τον τρόπο επιτυγχάνεται μία ανακάλυψη της υποκείμενης δομής των δεδομένων, χωρίς να υπάρχει ανάγκη για εκ των προτέρων καθορισμένες ετικέτες.

Στο πεδίο των τηλεπικοινωνιών, η ομαδοποίηση αποκτά ιδιαίτερη σημασία καθώς μπορεί να αποκαλύψει ομάδες χρηστών με παρόμοια χαρακτηριστικά καναλιού, επίπεδα παρεμβολής ή πρότυπα κίνησης. Η πληροφορία αυτή είναι ιδιαίτερα χρήσιμη για την ανάθεση πόρων, τη βελτιστοποίηση της συνδεσιμότητας ή την ανίχνευση ανωμαλιών στο δίκτυο. Για παράδειγμα, η ομαδοποίηση μπορεί να χρησιμοποιηθεί για τον εντοπισμό ομάδων χρηστών που εξυπηρετούνται καλύτερα από μια pico κυψέλη αντί για μία macro, οδηγώντας σε πιο ισορροπημένη κατανομή φόρτου.

Οι αλγόριθμοι ομαδοποίησης ποικίλλουν σημαντικά ως προς τη φιλοσοφία και τη μεθοδολογία τους. Ένας από τους πιο κλασικούς και ευρέως χρησιμοποιούμενους είναι ο αλγόριθμος K-Means. Η βασική του λογική έγκειται στη διαίρεση των δεδομένων σε k ομάδες, όπου κάθε ομάδα αναπαρίσταται από το κέντρο βάρους της. Η διαδικασία είναι επαναληπτική και περιλαμβάνει δύο βήματα: την ανάθεση κάθε σημείου στο πλησιέστερο κέντρο και τον επαναυπολογισμό των κέντρων με βάση τα σημεία που ανήκουν σε αυτά. Ο αλγόριθμος συνεχίζει μέχρι να σταθεροποιηθούν οι θέσεις των κέντρων. Παρότι είναι ιδιαίτερα αποδοτικός, παρουσιάζει ορισμένους περιορισμούς, όπως η ανάγκη καθορισμού του αριθμού των ομάδων εκ των προτέρων και η ευαισθησία στην αρχική επιλογή κέντρων. Επιπλέον, αποδίδει καλύτερα όταν τα clusters έχουν σφαιρικό σχήμα και παρόμοιο μέγεθος, κάτι που δεν ισχύει πάντοτε σε πραγματικά δεδομένα.

Μια σημαντική γενίκευση του K-Means είναι τα Μοντέλα Μιγμάτων Κανονικών Κατανομών (Gaussian Mixture Models – GMM). Στην περίπτωση αυτή, θεωρείται ότι τα δεδομένα προέρχονται από έναν συνδυασμό πολλών κανονικών κατανομών, καθεμία από τις οποίες αντιστοιχεί σε ένα cluster. Αντί να αποδίδεται κάθε σημείο σε μία μοναδική ομάδα, υπολογίζεται η πιθανότητα συμμετοχής του σε καθεμία από αυτές. Ο υπολογισμός πραγματοποιείται μέσω του αλγορίθμου Expectation-Maximization (EM), ο οποίος εναλλάσσει τον υπολογισμό πιθανοτήτων ένταξης σημείων στα clusters με την εκ νέου εκτίμηση των παραμέτρων των κατανομών. Η μέθοδος αυτή είναι πιο ευέλικτη, καθώς μπορεί να περιγράψει clusters με ελλειπτικά σχήματα ή να χειριστεί καλύτερα την επικάλυψη μεταξύ διαφορετικών ομάδων. Ωστόσο, έχει και αυτή περιορισμούς, όπως η πολυπλοκότητα και η πιθανότητα παγίδευσης σε τοπικά ακρότατα.

Η ιεραρχική ομαδοποίηση αποτελεί μία ακόμη σημαντική κατηγορία μεθόδων. Σε αυτήν, τα δεδομένα οργανώνονται σε μία δενδροειδή δομή που ονομάζεται δενδρόγραμμα. Η διαδικασία μπορεί να είναι είτε συγκλητική, όπου ξεκινά με κάθε σημείο ως ξεχωριστό cluster και συγχωνεύει προοδευτικά τα πιο κοντινά, είτε διαχωριστική, όπου ξεκινά με όλα τα δεδομένα μαζί και τα διαχωρίζει σταδιακά. Η απόφαση για το ποια clusters θα συγχωνευτούν ή θα διαχωριστούν βασίζεται σε μετρικές συνάφειας, όπως η μέση απόσταση, η μέγιστη απόσταση ή το κριτήριο Ward [16]. Το βασικό πλεονέκτημα της ιεραρχικής προσέγγισης είναι ότι δεν απαιτεί εκ των προτέρων γνώση του αριθμού των clusters και προσφέρει μία πολυεπίπεδη αναπαράσταση της δομής των δεδομένων. Από την άλλη, είναι πιο υπολογιστικά δαπανηρή και μπορεί να δυσκολέψει την ερμηνεία όταν το dataset είναι πολύ μεγάλο.

Μία διαφορετική φιλοσοφία ακολουθούν οι πυκνοκεντρικές μέθοδοι, όπως ο DBSCAN και ο OPTICS. Αυτές στηρίζονται στην ιδέα ότι τα clusters αποτελούν περιοχές υψηλής πυκνότητας που διαχωρίζονται από περιοχές χαμηλής πυκνότητας. Ο DBSCAN σχηματίζει clusters εντοπίζοντας περιοχές όπου κάθε σημείο έχει αρκετούς γείτονες μέσα σε μία προκαθορισμένη ακτίνα. Με τον τρόπο αυτόν, μπορεί να ανακαλύψει clusters αυθαίρετου σχήματος και να αγνοήσει σημεία που θεωρούνται θόρυβος. Ο OPTICS επεκτείνει την ιδέα αυτή επιτρέποντας την καλύτερη ανίχνευση clusters διαφορετικής πυκνότητας, αποδίδοντας μια πιο λεπτομερή αναπαράσταση της δομής των δεδομένων. Οι μέθοδοι αυτές είναι ιδιαίτερα κατάλληλες για την ανάλυση δεδομένων τηλεπικοινωνιών, καθώς συχνά παρατηρούνται περιοχές με πολύ διαφορετικά επίπεδα δραστηριότητας χρηστών και παρεμβολής.

Τέλος, για μεγάλης κλίμακας δεδομένα έχουν αναπτυχθεί μέθοδοι όπως ο MiniBatch K-Means και ο BIRCH. Ο πρώτος μειώνει τον χρόνο εκτέλεσης εφαρμόζοντας τον K-Means σε τυχαία δείγματα μικρού μεγέθους, διατηρώντας ικανοποιητική ακρίβεια. Ο δεύτερος χρησιμοποιεί δομές δέντρου για να συμπυκνώνει τις πληροφορίες και να δημιουργεί clusters με αποδοτικό τρόπο, ακόμη και όταν ο όγκος των δεδομένων είναι τεράστιος. Οι τεχνικές αυτές είναι καίριες για τηλεπικοινωνιακά περιβάλλοντα όπου παράγονται δεδομένα σε πραγματικό χρόνο και απαιτείται ανάλυση μεγάλης κλίμακας με περιορισμένους υπολογιστικούς πόρους.

Συνοψίζοντας, οι μέθοδοι ομαδοποίησης παρουσιάζουν διαφορετικά πλεονεκτήματα και μειονεκτήματα, και η επιλογή της κατάλληλης τεχνικής εξαρτάται από τη φύση των δεδομένων και τον στόχο της ανάλυσης. Στο πλαίσιο του DUDe και της βελτιστοποίησης της ανάθεσης χρηστών σε ετερογενή δίκτυα, η χρήση διαφορετικών μεθόδων clustering επιτρέπει την πολυδιάστατη μελέτη των δεδομένων και οδηγεί σε πιο ολοκληρωμένα και αξιόπιστα συμπεράσματα.

4.2.1 K-Means και παραλλαγές

Ο αλγόριθμος K-Means αποτελεί έναν από τους πιο διαδεδομένους και θεμελιώδεις αλγορίθμους ομαδοποίησης. Η δημοτικότητά του οφείλεται στην απλότητα, την ταχύτητά του και στην ικανότητά του να παράγει εύκολα ερμηνεύσιμα αποτελέσματα. Στον πυρήνα του, ο K-Means επιδιώκει να χωρίσει ένα σύνολο δεδομένων σε k ομάδες (clusters), όπου κάθε δεδομένο ανήκει στο cluster με το πλησιέστερο κέντρο (centroid).

Η λειτουργία του αλγορίθμου βασίζεται στην επαναληπτική ελαχιστοποίηση της εξής συνάρτησης κόστους:

$$J = \sum_{i=1}^N \sum_{j=1}^k r_{ij} \|x_i - \mu_j\|^2 \quad (4.1) \quad [15]$$

όπου N είναι το πλήθος των δειγμάτων, x_i το i -οστό δείγμα, μ_j το κέντρο του j -οστού cluster και r_{ij} μια δυαδική μεταβλητή που ισούται με 1 αν το δείγμα x_i ανατεθεί στο cluster j , και 0 διαφορετικά. Ουσιαστικά, ο K-Means επιδιώκει να ελαχιστοποιήσει το άθροισμα των τετραγωνικών αποστάσεων κάθε σημείου από το κοντινότερο κέντρο του.

Η βασική διαδικασία του αλγορίθμου μπορεί να συνοψιστεί στα παρακάτω στάδια:

1. Επιλογή αρχικών κέντρων (συχνά με τυχαίο τρόπο ή με μεθόδους όπως το k-means++ για καλύτερη σταθερότητα).
2. Ανάθεση κάθε δείγματος στο πλησιέστερο κέντρο με βάση μια μετρική απόστασης (συνήθως την Ευκλείδεια).
3. Ανακαθορισμός των κέντρων ως ο μέσος όρος όλων των σημείων που ανήκουν στο cluster.
4. Επανάληψη των βημάτων 2–3 μέχρι σύγκλιση, δηλαδή μέχρι τα κέντρα να μην αλλάζουν σημαντικά ή να ικανοποιείται κάποιο κριτήριο τερματισμού.

Αν και ο K-Means είναι απλός και αποτελεσματικός, παρουσιάζει ορισμένους περιορισμούς. Απαιτεί εκ των προτέρων τον αριθμό των clusters k , είναι ευαίσθητος στην αρχικοποίηση των κέντρων, ενώ η χρήση της Ευκλείδειας απόστασης τον καθιστά κατάλληλο κυρίως για σφαιρικά ή ισομεγέθη clusters. Για την αντιμετώπιση αυτών των ζητημάτων έχουν αναπτυχθεί παραλλαγές, όπως ο MiniBatchKMeans: προσαρμογή του αλγορίθμου για μεγάλα datasets, όπου τα κέντρα ενημερώνονται χρησιμοποιώντας μικρά τυχαία υποσύνολα (mini-batches) των δεδομένων, μειώνοντας δραστικά τον υπολογιστικό φόρτο. Στο πεδίο των τηλεπικοινωνιών, ο K-Means χρησιμοποιείται ευρέως για την ανάλυση μετρήσεων ποιότητας σήματος και για την ομαδοποίηση χρηστών με παρόμοια χαρακτηριστικά. Για παράδειγμα, μετρικές όπως η απόσταση από τον σταθμό βάσης, το SNR ή το throughput μπορούν να χρησιμοποιηθούν για να διαχωριστούν οι χρήστες σε ομάδες με κοινή εμπειρία δικτύου. Αυτό έχει πρακτικές εφαρμογές στη βελτιστοποίηση της ανάθεσης χρηστών σε macro ή pico κυψέλες, στην εξισορρόπηση φόρτου και στη διαχείριση παρεμβολών. Η απλότητα του K-Means τον καθιστά ιδανικό ως αφετηρία για συγκρίσεις με πιο περίπλοκους αλγορίθμους, ενώ οι παραλλαγές του

(MiniBatchKMeans, K-Medoids) επιτρέπουν την εφαρμογή του σε μεγάλης κλίμακας δεδομένα τηλεπικοινωνιακών δικτύων, όπως αυτά που προκύπτουν σε περιβάλλοντα 5G.

4.2.2 Ιεραρχική Ομαδοποίηση (Hierarchical Clustering)

Η ιεραρχική ομαδοποίηση αποτελεί μία από τις πιο διαδεδομένες τεχνικές στη μη εποπτευόμενη μάθηση και ειδικότερα στην ανάλυση συστάδων. Σε αντίθεση με τον αλγόριθμο K-Means, όπου απαιτείται εκ των προτέρων ο καθορισμός του αριθμού των clusters, η ιεραρχική ομαδοποίηση βασίζεται στη δημιουργία μιας ιεραρχικής δομής, η οποία μπορεί να αναπαρασταθεί με τη μορφή δενδρογράμματος (dendrogram). Μέσω του δενδρογράμματος, ο ερευνητής έχει τη δυνατότητα να παρατηρήσει τη δομή των δεδομένων σε διαφορετικά επίπεδα συσσωμάτωσης και να επιλέξει τον αριθμό των clusters που θεωρεί πιο αντιπροσωπευτικό για το πρόβλημα που μελετά.

Η βασική ιδέα της μεθόδου είναι ότι κάθε δεδομένο αρχικά θεωρείται ως μία ξεχωριστή ομάδα και στη συνέχεια τα δεδομένα ή οι ομάδες συγχωνεύονται επαναληπτικά με βάση κάποιο μέτρο ομοιότητας. Η διαδικασία αυτή ονομάζεται συγχωνευτική (agglomerative) ιεραρχική ομαδοποίηση. Υπάρχει και η αντίστροφη εκδοχή, η λεγόμενη διαιρετική (divisive), όπου όλα τα δεδομένα θεωρούνται ότι ανήκουν σε μία ενιαία ομάδα και σταδιακά διαχωρίζονται σε μικρότερα clusters. Στην πράξη, η συγχωνευτική προσέγγιση είναι η πιο συνηθισμένη, καθώς είναι απλούστερη υπολογιστικά και παρέχει πιο ερμηνεύσιμα αποτελέσματα.

Η διαδικασία της ιεραρχικής ομαδοποίησης στηρίζεται στον ορισμό ενός μέτρου απόστασης μεταξύ σημείων και ενός κριτηρίου συνένωσης μεταξύ ομάδων. Συνήθη μέτρα απόστασης είναι η ευκλείδεια απόσταση, η Manhattan απόσταση ή η συνημιτονοειδής ομοιότητα, ανάλογα με τη φύση των δεδομένων. Το κριτήριο συνένωσης καθορίζει πώς υπολογίζεται η απόσταση μεταξύ δύο clusters και διακρίνεται σε διάφορες μεθόδους, όπως:

- Single linkage: η απόσταση μεταξύ των δύο πλησιέστερων σημείων.
- Complete linkage: η απόσταση μεταξύ των πιο απομακρυσμένων σημείων.
- Average linkage: ο μέσος όρος των αποστάσεων όλων των ζευγών σημείων.
- Ward's method: επιλέγει τη συγχώνευση που ελαχιστοποιεί την αύξηση της συνολικής ενδο-ομαδικής διακύμανσης.

Μαθηματικά, η συνάρτηση κόστους που ελαχιστοποιεί η μέθοδος του Ward μπορεί να γραφεί ως:

$$\Delta E = \frac{n_A n_B}{n_A + n_B} \cdot \|\mu_A - \mu_B\|^2 \quad (4.2) \quad [16]$$

Όπου $n_A n_B$ είναι τα μεγέθη των clusters A και B και μ_A, μ_B τα αντίστοιχα κεντροειδή. Η συγχώνευση των clusters που δίνει τη μικρότερη αύξηση στο ΔE θεωρείται η βέλτιστη σε κάθε βήμα.

Το πλεονέκτημα της ιεραρχικής ομαδοποίησης είναι ότι δεν απαιτείται ο προκαθορισμός του αριθμού των clusters. Ωστόσο, η μέθοδος είναι πιο απαιτητική υπολογιστικά, καθώς η πολυπλοκότητα αυξάνεται περίπου με $O(n^3)$, γεγονός που την καθιστά δύσχρηστη σε πολύ μεγάλα σύνολα δεδομένων. Παράλληλα, η ευαισθησία της στο θόρυβο και στα outliers μπορεί να οδηγήσει σε μη ικανοποιητικά αποτελέσματα, εάν δεν υπάρξει κατάλληλη προεπεξεργασία των δεδομένων.

Στο πλαίσιο των τηλεπικοινωνιών και ειδικότερα της τεχνολογίας DUDe, η ιεραρχική ομαδοποίηση μπορεί να χρησιμοποιηθεί για την κατανόηση των προτύπων απόδοσης των κυψελών. Μέσω του δενδρογράμματος, είναι δυνατή η παρατήρηση της φυσικής συσχέτισης μεταξύ χρηστών που βιώνουν παρόμοιο επίπεδο θορύβου, παρεμβολής ή καθυστέρησης. Με αυτόν τον τρόπο, η μέθοδος προσφέρει ένα εργαλείο ανάλυσης που συμπληρώνει πιο αυστηρούς αλγορίθμους, όπως το K-Means, καθώς παρέχει μια πολυεπίπεδη εικόνα της δομής των δεδομένων.

4.2.3 Gaussian Mixture Models (GMM)

Τα Gaussian Mixture Models (GMM) αποτελούν μια από τις πιο ισχυρές και ευέλικτες μεθόδους ομαδοποίησης. Η βασική ιδέα τους είναι ότι τα δεδομένα δεν κατανέμονται απαραίτητα γύρω από ένα μοναδικό κέντρο (όπως στο K-Means), αλλά μπορούν να περιγραφούν ως γραμμικός συνδυασμός από πολλές Γκαουσιανές κατανομές. Κάθε μία από αυτές τις κατανομές αντιστοιχεί σε ένα cluster, το οποίο ορίζεται όχι μόνο από το μέσο (mean) του αλλά και από την διασπορά του (covariance). [13]

Ας θεωρήσουμε ότι έχουμε K clusters. Το GMM υποθέτει ότι η πιθανότητα να παρατηρηθεί ένα σημείο x_i δίνεται από το μείγμα:

$$p(x_i) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \quad (4.3) \quad [13]$$

όπου:

- π_k είναι το βάρος (mixture weight) του k -οστού cluster, με την προϋπόθεση ότι $\sum_{k=1}^K \pi_k = 1$
- μ_k είναι το διάνυσμα μέσης τιμής (mean vector) του k -οστού cluster.
- Σ_k είναι ο πίνακας συνδιακύμανσης (covariance matrix) που περιγράφει το σχήμα του cluster.
- $\mathcal{N}(x_i | \mu_k, \Sigma_k)$ είναι η Γκαουσιανή κατανομή με παράμετρο μ_k και Σ_k .

Η εξίσωση μιας πολυδιάστατης Γκαουσιανής κατανομής δίνεται από:

$$N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right) \quad (4.4) \quad [13]$$

Όπου d είναι η διάσταση του διανύσματος x

Η εκτίμηση των παραμέτρων (π_k, μ_k, Σ_k) γίνεται με τη μέθοδο Expectation-Maximization (EM), η οποία εναλλάσσει δύο στάδια:

- E-step (Expectation): Υπολογίζει την πιθανότητα κάθε σημείου να ανήκει σε κάθε cluster.
- M-step (Maximization): Ενημερώνει τις παραμέτρους π_k, μ_k, Σ_k ώστε να μεγιστοποιείται η πιθανοφάνεια.

Σε αντίθεση με το K-Means, που αποδίδει κάθε σημείο αποκλειστικά σε ένα cluster, τα GMM παρέχουν πιθανολογική ανάθεση. Δηλαδή, ένα σημείο μπορεί να έχει πιθανότητες συμμετοχής σε πολλαπλά clusters, γεγονός που τα καθιστά πιο ευέλικτα για δεδομένα με επικαλυπτόμενες ομάδες.

Η ισχύς του GMM είναι ιδιαίτερα εμφανής σε τηλεπικοινωνιακά δεδομένα, όπου οι μετρήσεις SNR, throughput και παρεμβολής συχνά παρουσιάζουν επικαλύψεις. Με αυτόν τον τρόπο, το GMM μπορεί να αποτυπώσει καλύτερα την αβεβαιότητα και τη συνεχή μετάβαση των χαρακτηριστικών μεταξύ διαφορετικών περιοχών κάλυψης.

4.2.4 DBSCAN

Ο αλγόριθμος DBSCAN (Density-Based Spatial Clustering of Applications with Noise) αποτελεί μία από τις πιο γνωστές μεθόδους μη εποπτευόμενης μάθησης που στηρίζονται στην έννοια της πυκνότητας [15]. Σε αντίθεση με το K-Means και τα Gaussian Mixture Models, όπου απαιτείται εκ των προτέρων να δοθεί ο αριθμός των clusters, το DBSCAN προσδιορίζει τις ομάδες με βάση την κατανομή των δεδομένων στον χώρο. Αυτό το χαρακτηριστικό τον καθιστά ιδιαίτερα χρήσιμο σε εφαρμογές όπου δεν είναι γνωστός ο αριθμός των συστάδων ή όταν τα δεδομένα δεν έχουν σφαιρική δομή.

Η βασική ιδέα του DBSCAN είναι ότι ένα cluster μπορεί να περιγραφεί ως μία περιοχή υψηλής πυκνότητας σημείων, η οποία διαχωρίζεται από άλλες περιοχές μέσω περιοχών χαμηλής πυκνότητας. Για τον καθορισμό των clusters, χρησιμοποιούνται δύο βασικές παράμετροι:

- ϵ (epsilon): η ακτίνα γειτονιάς γύρω από κάθε σημείο.
- minPts (minimum points): το ελάχιστο πλήθος σημείων που πρέπει να περιέχει η γειτονιά για να χαρακτηριστεί ως «πυκνή».

Με βάση αυτές τις δύο παραμέτρους, κάθε σημείο μπορεί να κατηγοριοποιηθεί σε μία από τις εξής κατηγορίες:

1. Core point (κεντρικό σημείο): έχει τουλάχιστον $minPts$ σημεία στη γειτονιά του.
2. Border point (οριακό σημείο): βρίσκεται στη γειτονιά ενός core point, αλλά δεν έχει αρκετά σημεία για να χαρακτηριστεί και το ίδιο core point.
3. Noise point (θόρυβος): δεν ανήκει σε κανένα cluster και βρίσκεται σε αραιή περιοχή.

Ο αλγόριθμος DBSCAN ξεκινάει από ένα τυχαίο σημείο και, εφόσον αυτό είναι core point, επεκτείνει το cluster περιλαμβάνοντας όλα τα σημεία της γειτονιάς του. Η διαδικασία επαναλαμβάνεται μέχρι να μην υπάρχουν άλλα σημεία που μπορούν να προστεθούν στο ίδιο cluster. Αντίθετα, τα σημεία που δεν πληρούν τα κριτήρια παραμένουν ως θόρυβος.

Μαθηματικά, ο DBSCAN ορίζεται ως εξής:

$$N_{\epsilon}(x) = \{y \in D \mid |x - y| \leq \epsilon\} \quad (4.5) [17]$$

Όπου $N_\varepsilon(x)$ είναι η ε -γειτονιά του σημείου x .

Η εξίσωση ορίζει την ε -γειτονιά ενός σημείου στον αλγόριθμο DBSCAN, η οποία χρησιμοποιείται για τον εντοπισμό περιοχών υψηλής πυκνότητας.

Ένα από τα μεγαλύτερα πλεονεκτήματα του DBSCAN είναι ότι μπορεί να εντοπίσει συστάδες αυθαίρετου σχήματος, κάτι που δεν είναι εφικτό με μεθόδους όπως το K-Means, οι οποίες «επιβάλλουν» σφαιρικές συστάδες. Επιπλέον, έχει τη δυνατότητα να εντοπίζει ανωμαλίες (outliers) με φυσικό τρόπο, καθώς τα δεδομένα χαμηλής πυκνότητας χαρακτηρίζονται ως θόρυβος.

Ωστόσο, ο DBSCAN παρουσιάζει και ορισμένα μειονεκτήματα. Αρχικά, η επιλογή των παραμέτρων ε και $minPts$ επηρεάζει δραστικά την ποιότητα των αποτελεσμάτων και συνήθως απαιτείται εμπειρικός καθορισμός. Επιπλέον, η μέθοδος δυσκολεύεται όταν οι συστάδες έχουν μεταβαλλόμενη πυκνότητα, καθώς τότε είναι δύσκολο να οριστεί μία ενιαία τιμή ε που να λειτουργεί ικανοποιητικά σε όλα τα δεδομένα.

Στο πεδίο των τηλεπικοινωνιών, ο DBSCAN βρίσκει εφαρμογές σε σενάρια όπως η ανίχνευση χρηστών με παρόμοια χαρακτηριστικά καναλιού, η αναγνώριση περιοχών υψηλής παρεμβολής και η διάκριση ακραίων χρηστών που προκαλούν δυσανάλογη επιβάρυνση στο uplink. Αυτή η δυνατότητα τον καθιστά ιδιαίτερα χρήσιμο για την τεχνολογία DUDe, καθώς μπορεί να συμβάλει στην αναγνώριση ομάδων χρηστών που επωφελούνται περισσότερο από τον αποσυζευγμένο συσχετισμό uplink και downlink.

4.2.5 Spectral Clustering

Ο αλγόριθμος Spectral Clustering βασίζεται στη θεωρία γραφημάτων και χρησιμοποιεί τις ιδιότητες του φάσματος (eigenvalues και eigenvectors) του πίνακα Laplacian ενός γράφου που αναπαριστά τα δεδομένα [17]. Σε αντίθεση με μεθόδους όπως το K-Means, που λειτουργούν απευθείας στον χώρο των χαρακτηριστικών, το Spectral Clustering μετασχηματίζει τα δεδομένα σε έναν νέο χώρο, όπου η δομή των clusters γίνεται πιο διακριτή.

Η διαδικασία ξεκινά με τη δημιουργία ενός γράφου γειτνίασης, όπου κάθε σημείο δεδομένων αντιστοιχεί σε έναν κόμβο, και οι ακμές μεταξύ κόμβων έχουν βάρη που εκφράζουν την ομοιότητα μεταξύ τους. Η ομοιότητα αυτή συνήθως υπολογίζεται μέσω ενός kernel, όπως ο Gaussian kernel:

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (4.6) [18]$$

όπου ο εκθετικός όρος $\exp(\cdot)$ μετατρέπει την απόσταση μεταξύ των σημείων σε μέτρο ομοιότητας, ο όρος $\|x_i - x_j\|^2$ εκφράζει το τετράγωνο της Ευκλείδειας απόστασης μεταξύ των διανυσμάτων χαρακτηριστικών x_i και x_j , ενώ ο παρονομαστής $2\sigma^2$ καθορίζει την κλίμακα εξασθένησης της ομοιότητας μέσω της παραμέτρου σ .

Από τον πίνακα βαρών W κατασκευάζεται ο πίνακας βαθμών D , με στοιχεία

$$d_{ii} = \sum_j w_{ij} \quad (4.7) \quad [18]$$

όπου το άθροισμα $\sum_j w_{ij}$ περιλαμβάνει όλα τα βάρη ομοιότητας μεταξύ του κόμβου i και των υπόλοιπων κόμβων του γράφου, εκφράζοντας τη συνολική συνδεσιμότητα του κόμβου i , και στη συνέχεια ο πίνακας Laplacian:

$$L = D - W \quad (4.8) \quad [18]$$

όπου εκφράζει την διαφορά του διαγώνιου πίνακα βαθμών D και του πίνακα βαρών W , με αποτέλεσμα ο πίνακας Laplacian L να ενσωματώνει τόσο τη συνδεσιμότητα κάθε κόμβου όσο και τις μεταξύ τους ομοιότητες.

Η ιδιότητα του πίνακα Laplacian είναι ότι τα ιδιοδιανύσματά του αποκαλύπτουν πληροφορίες για τη δομή του γράφου, ειδικά για τη συνδεσιμότητα μεταξύ των κόμβων.

Το Spectral Clustering επιλέγει τα πρώτα K ιδιοδιανύσματα του L (ή κάποια κανονικοποιημένη παραλλαγή του) και χρησιμοποιεί αυτά τα ιδιοδιανύσματα ως νέα χαρακτηριστικά για τα δεδομένα. Στη συνέχεια εφαρμόζει τον K-Means πάνω σε αυτόν τον νέο χώρο για να προσδιορίσει τα clusters.

Ένα από τα σημαντικά πλεονεκτήματα του Spectral Clustering είναι η ικανότητά του να εντοπίζει συστάδες με περίπλοκα σχήματα και μη γραμμικά διαχωρίσιμες δομές, κάτι που είναι δύσκολο για πιο απλούς αλγόριθμους. Ειδικότερα, σε εφαρμογές τηλεπικοινωνιών, η μέθοδος αυτή μπορεί να χρησιμοποιηθεί για την αναγνώριση ομάδων χρηστών με παρόμοια συμπεριφορά, ακόμα κι αν τα όρια μεταξύ τους δεν είναι ευδιάκριτα με παραδοσιακές μετρικές απόστασης.

Ωστόσο, η μέθοδος απαιτεί σημαντικούς υπολογιστικούς πόρους, ειδικά για μεγάλα σύνολα δεδομένων, καθώς προϋποθέτει την αποθήκευση και επεξεργασία του πίνακα ομοιοτήτων.

Αυτό καθιστά το Spectral Clustering πιο κατάλληλο για μεσαίου μεγέθους προβλήματα, όπου η ποιότητα των clusters υπερτερεί της ανάγκης για κλιμακωσιμότητα.

4.2.6 Birch (Balanced Iterative Reducing and Clustering using Hierarchies)

Ο αλγόριθμος Birch (Balanced Iterative Reducing and Clustering using Hierarchies) ανήκει στις ιεραρχικές μεθόδους ομαδοποίησης και έχει σχεδιαστεί ειδικά για μεγάλα σύνολα δεδομένων [19]. Βασίζεται στην κατασκευή μιας συμπαγούς δομής δέντρου, γνωστής ως Clustering Feature Tree (CF-tree), η οποία συμπυκνώνει προοδευτικά τα δεδομένα διατηρώντας τα βασικά στατιστικά χαρακτηριστικά τους.

Κάθε κόμβος του CF-tree περιγράφεται από ένα clustering feature (CF), το οποίο συνοψίζει μία ομάδα σημείων μέσω τριπλέτας:

- πλήθος σημείων του cluster
- άθροισμα διανυσμάτων χαρακτηριστικών
- άθροισμα τετραγώνων αποστάσεων από το κέντρο

Με αυτόν τον τρόπο αποφεύγεται η ανάγκη αποθήκευσης όλων των δεδομένων στη μνήμη.

Η διαδικασία λειτουργίας Birch περιλαμβάνει δύο βασικά στάδια: Την κατασκευή CF-tree, όπου τα δεδομένα εισάγονται διαδοχικά και το δέντρο προσαρμόζεται δυναμικά, καθώς και το προαιρετικό refinement, όπου γίνεται εφαρμογή δεύτερου αλγορίθμου (συνήθως K-Means) πάνω στα φύλλα του δέντρου. Τα κύρια πλεονεκτήματα του Birch είναι:

- εξαιρετική κλιμακωσιμότητα για πολύ μεγάλα dataset
- χαμηλές απαιτήσεις μνήμης
- δυνατότητα επεξεργασίας δεδομένων σε streaming μορφή

Ωστόσο παρουσιάζει και μειονεκτήματα, όπως ευαισθησία στη σειρά εισαγωγής δεδομένων και δυσκολία χειρισμού clusters πολύ διαφορετικής πυκνότητας.

Στο πλαίσιο της παρούσας εργασίας, ο Birch είναι ιδιαίτερα χρήσιμος λόγω του μεγάλου πλήθους χρηστών σε κυψελωτά δίκτυα, επιτρέποντας αποδοτική ομαδοποίηση και ανάλυση συμπεριφοράς χρηστών σε όρους uplink απόδοσης.

4.2.7 OPTICS (Ordering Points To Identify the Clustering Structure)

Ο αλγόριθμος OPTICS αποτελεί επέκταση του DBSCAN και στοχεύει στην αντιμετώπιση του βασικού του περιορισμού, δηλαδή της δυσκολίας επιλογής μίας μοναδικής τιμής της παραμέτρου ϵ για δεδομένα με μεταβαλλόμενη πυκνότητα [21].

Αντί να παράγει άμεσα συγκεκριμένα clusters, ο OPTICS δημιουργεί μια σειριοθέτηση σημείων βάσει πυκνότητας, παράγοντας δύο βασικά μεγέθη:

- Core distance, δηλαδή την ελάχιστη απόσταση για να θεωρηθεί ένα σημείο πυκνό
- Reachability distance, δηλαδή την απόσταση προσπελασιμότητας ενός σημείου από άλλο

Η γραφική αναπαράσταση των reachability distances δημιουργεί το reachability plot, από το οποίο μπορούν να εξαχθούν clusters διαφορετικής πυκνότητας με οπτική ή αλγοριθμική ανάλυση. Τα βασικότερα πλεονεκτήματα του OPTICS, είναι ότι εντοπίζει clusters διαφορετικών πυκνοτήτων, αναγνωρίζει φυσικά outliers καθώς και δεν απαιτεί προκαθορισμένο αριθμό clusters. Τα κύρια μειονεκτήματα είναι η αυξημένη υπολογιστική πολυπλοκότητα έναντι DBSCAN, και η μεγαλύτερη δυσκολία ερμηνείας αποτελεσμάτων.

Στα ετερογενή δίκτυα 5G/6G, όπου παρατηρούνται περιοχές με πυκνή και αραιή κίνηση χρηστών, ο OPTICS μπορεί να αναγνωρίσει ομάδες χρηστών cell-edge, να εντοπίσει hot-spots παρεμβολής και να υποστηρίξει βελτιωμένες στρατηγικές DUDe.

4.2.8 Isolation Forest (Ανίχνευση ανωμαλιών)

Ο Isolation Forest δεν αποτελεί κλασικό αλγόριθμο clustering, αλλά μέθοδο ανίχνευσης ανωμαλιών (anomaly detection). Η βασική ιδέα είναι ότι τα σημεία που θεωρούνται ανωμαλίες απομονώνονται ευκολότερα από τα κανονικά σημεία λόγω της διαφορετικής στατιστικής συμπεριφοράς τους [20].

Ο αλγόριθμος κατασκευάζει πολλά τυχαία δυαδικά δέντρα κατάτμησης: κάθε διαχωρισμός επιλέγει τυχαία χαρακτηριστικό και τιμή κατωφλίου, και τα σημεία που «απομονώνονται» με λίγους διαχωρισμούς θεωρούνται ανωμαλίες. Το αναμενόμενο μήκος διαδρομής ενός σημείου μέσα στο δέντρο χρησιμοποιείται ως δείκτης «κανονικότητας». Είναι πολύ αποδοτικός σε μεγάλα datasets, δεν κάνει υποθέσεις για την κατανομή των δεδομένων και εντοπίζει ακραίες

τιμές και outliers φυσικά. Στον αντίποδα, δεν επιστρέφει clusters αλλά ετικέτες normal/anomaly και απαιτεί επιλογή παραμέτρου contamination

Στο πλαίσιο DUDe και κυψελωτών δικτύων, ο Isolation Forest μπορεί να εντοπίσει χρήστες με εξαιρετικά χαμηλό uplink SINR, ασυνήθιστη παρεμβολή, εξοπλισμό με μη ομαλή συμπεριφορά, καθώς και cell-edge περιπτώσεις με ανώμαλα χαρακτηριστικά καναλιού, και έτσι να υποστηρίξει προσαρμοστικές πολιτικές σύνδεσης.

4.3 Μετρικές Αξιολόγησης Clustering

Η αξιολόγηση της ποιότητας ενός αλγορίθμου clustering αποτελεί θεμελιώδες στάδιο στη διαδικασία ανάλυσης δεδομένων μέσω μη εποπτευόμενης μηχανικής μάθησης. Σε αντίθεση με την εποπτευόμενη μάθηση, όπου η ύπαρξη προκαθορισμένων ετικετών επιτρέπει την άμεση ποσοτική εκτίμηση της απόδοσης ενός μοντέλου μέσω καθιερωμένων μετρικών, όπως η ακρίβεια και η ανάκληση, στην ομαδοποίηση δεν υπάρχουν εκ των προτέρων γνωστές «σωστές» κατηγορίες. Ως εκ τούτου, η αξιολόγηση της ποιότητας των παραγόμενων clusters βασίζεται σε εξειδικευμένες μετρικές που επιχειρούν να αποτυπώσουν τόσο τον βαθμό συνοχής των δεδομένων εντός κάθε ομάδας όσο και τον βαθμό διαχωρισιμότητας μεταξύ διαφορετικών ομάδων. Οι μετρικές αξιολόγησης των αλγορίθμων clustering διακρίνονται γενικά σε δύο βασικές κατηγορίες: τις εσωτερικές και τις εξωτερικές. Οι εσωτερικές μετρικές βασίζονται αποκλειστικά στη γεωμετρική δομή των δεδομένων και στις αναθέσεις τους σε clusters, χωρίς να απαιτούν εξωτερική πληροφορία αναφοράς. Αντίθετα, οι εξωτερικές μετρικές προϋποθέτουν την ύπαρξη ground truth labels και αξιολογούν την παραγόμενη ομαδοποίηση συγκρίνοντάς τη με τις πραγματικές κατηγορίες.

Στο πεδίο των τηλεπικοινωνιών, όπου σπάνια υπάρχουν διαθέσιμες πραγματικές ετικέτες για τους χρήστες, τις κυψέλες ή τις συνθήκες καναλιού, οι εσωτερικές μετρικές αποτελούν το βασικό εργαλείο αξιολόγησης της ποιότητας του clustering. Παράλληλα, σε προσομοιωτικά περιβάλλοντα ή πειραματικά σενάρια με γνωστές παραμέτρους, οι εξωτερικές μετρικές μπορούν να χρησιμοποιηθούν συμπληρωματικά για την επιβεβαίωση της αξιοπιστίας των αποτελεσμάτων. Η επιλογή της καταλληλότερης μετρικής εξαρτάται από τη φύση του προβλήματος και τα χαρακτηριστικά των δεδομένων. Στην πράξη, η συνδυαστική χρήση πολλαπλών μετρικών θεωρείται βέλτιστη πρακτική, καθώς επιτρέπει μια πιο σφαιρική και αξιόπιστη αποτίμηση της ποιότητας της ομαδοποίησης και μειώνει τον κίνδυνο εσφαλμένων

συμπερασμάτων που ενδέχεται να προκύψουν από την αποκλειστική χρήση μιας μόνο μετρικής.

4.3.1 Εσωτερικές Μετρικές

Οι εσωτερικές μετρικές (intrinsic metrics) εφαρμόζονται σε περιπτώσεις όπου δεν υπάρχουν διαθέσιμες ετικέτες αναφοράς και αξιοποιούν αποκλειστικά τις πληροφορίες που περιέχονται στα ίδια τα δεδομένα και στη δομή των σχηματισμένων clusters. Η βασική αρχή που τις διέπει είναι ότι μια αποδοτική ομαδοποίηση πρέπει να χαρακτηρίζεται από υψηλή συνοχή εντός των clusters, δηλαδή τα σημεία κάθε ομάδας να βρίσκονται όσο το δυνατόν πιο κοντά μεταξύ τους, και ταυτόχρονα από σαφή διαχωρισμό μεταξύ διαφορετικών ομάδων.

Η βασική λογική πίσω από τις εσωτερικές μετρικές είναι ότι μια «καλή» ομαδοποίηση πρέπει να εμφανίζει υψηλή ομοιογένεια εντός των clusters (τα σημεία του ίδιου cluster να είναι όσο το δυνατόν πιο κοντά μεταξύ τους) και ταυτόχρονα μεγάλη απόσταση ή διαφοροποίηση μεταξύ των διαφορετικών clusters. Για τον λόγο αυτό, οι περισσότερες μετρικές συνδυάζουν έναν δείκτη ενδο-ομαδικής συνοχής (cohesion) με έναν δείκτη μεταξύ-ομαδικής διαχωρισιμότητας (separation).

Ο πιο διαδεδομένος δείκτης αυτής της κατηγορίας είναι ο δείκτης Silhouette [22]. Για κάθε δείγμα i , ορίζεται αρχικά η μέση απόσταση $a(i)$ από όλα τα υπόλοιπα σημεία του ίδιου cluster και η μέση απόσταση $b(i)$ από τα σημεία του πλησιέστερου γειτονικού cluster. Ο δείκτης Silhouette για το σημείο i δίνεται από την σχέση

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4.9) \quad [23]$$

Η τιμή του $s(i)$ κυμαίνεται από -1 έως $+1$. Τιμές κοντά στο $+1$ υποδηλώνουν ότι το σημείο είναι καλά τοποθετημένο στο cluster του, τιμές κοντά στο μηδέν υποδεικνύουν αβεβαιότητα, ενώ αρνητικές τιμές σηματοδοτούν ότι το σημείο ίσως θα έπρεπε να ανήκει σε διαφορετικό cluster. Ο συνολικός δείκτης Silhouette ενός clustering προκύπτει ως ο μέσος όρος όλων των $s(i)$

Ένας άλλος σημαντικός δείκτης είναι ο δείκτης Davies–Bouldin (DBI) [24]. Ο δείκτης αυτός μετρά τον μέσο όρο της ομοιότητας κάθε cluster με το πιο «όμοιο» cluster του, όπου η ομοιότητα ορίζεται ως ο λόγος της ενδο-ομαδικής διασποράς προς την απόσταση μεταξύ των

κέντρων των clusters. Συνεπώς, όσο μικρότερη είναι η τιμή του DBI, τόσο καλύτερη θεωρείται η ομαδοποίηση, καθώς υποδηλώνει ότι τα clusters είναι καλά διαχωρισμένα και με μικρή εσωτερική διασπορά.

Συναφής μετρική είναι ο δείκτης Calinski–Harabasz (CH) [25], γνωστός και ως Variance Ratio Criterion. Ο δείκτης αυτός υπολογίζει το πηλίκο της διασποράς μεταξύ των clusters προς τη διασπορά εντός των clusters. Υψηλότερες τιμές του δείκτη σημαίνουν ότι τα clusters είναι πιο διαχωρισμένα και συνεκτικά, επομένως η ομαδοποίηση είναι καλύτερης ποιότητας.

Εκτός από τους παραπάνω, έχουν προταθεί και άλλες εσωτερικές μετρικές, όπως ο δείκτης Dunn, ο οποίος εστιάζει στον λόγο της ελάχιστης απόστασης μεταξύ clusters προς τη μέγιστη ενδο-ομαδική διάμετρο. Ωστόσο, στην πράξη οι πιο διαδεδομένες μετρικές που χρησιμοποιούνται είναι οι Silhouette, Davies–Bouldin και Calinski–Harabasz, τόσο λόγω της ευκολίας υπολογισμού όσο και λόγω της ευρείας αποδοχής τους στην ερευνητική κοινότητα.

Στο πλαίσιο των τηλεπικοινωνιακών συστημάτων, όπου η ύπαρξη ground truth labels είναι σπάνια, οι εσωτερικές μετρικές αποκτούν ιδιαίτερη σημασία. Για παράδειγμα, κατά την ομαδοποίηση χρηστών με βάση μετρήσεις όπως το SNR, το throughput ή την παρεμβολή, οι δείκτες αυτοί επιτρέπουν την αξιολόγηση της ποιότητας της ομαδοποίησης χωρίς να απαιτούνται εξωτερικές πληροφορίες. Έτσι, ο ερευνητής ή ο μηχανικός μπορεί να συγκρίνει διαφορετικούς αλγορίθμους clustering και να επιλέξει εκείνον που αποδίδει τα πιο συνεπή και διαχωρίσιμα clusters.

4.3.2 Εξωτερικές Μετρικές

Οι εξωτερικές μετρικές χρησιμοποιούνται όταν υπάρχει διαθέσιμη κάποια «αναφορά» ή ground truth, δηλαδή γνωστές ετικέτες που περιγράφουν τις πραγματικές κατηγορίες των δεδομένων. Σε αντίθεση με τις εσωτερικές μετρικές που στηρίζονται αποκλειστικά στη δομή των δεδομένων, οι εξωτερικές μετρικές επιτρέπουν την άμεση αξιολόγηση της ομαδοποίησης με βάση το πόσο κοντά βρίσκεται στο πραγματικό μοτίβο ταξινόμησης.

Μια από τις πιο διαδεδομένες μετρικές είναι ο Adjusted Rand Index (ARI), ο οποίος μετρά την ομοιότητα μεταξύ δύο διαφορετικών ομαδοποιήσεων [26]. Ο δείκτης διορθώνεται ως προς την τυχαία συμφωνία, ώστε η τιμή του να κυμαίνεται από -1 (πλήρης ασυμφωνία) έως 1 (απόλυτη ταύτιση), με το 0 να δηλώνει τυχαία ομοιότητα.

Οι εξωτερικές μετρικές έχουν το πλεονέκτημα της σαφούς ερμηνείας, ωστόσο η χρήση τους είναι περιορισμένη σε σενάρια όπου υπάρχουν διαθέσιμες πραγματικές ετικέτες. Στα

περισσότερα τηλεπικοινωνιακά δεδομένα πεδίου, τέτοιες πληροφορίες δεν υπάρχουν, γεγονός που καθιστά τις εσωτερικές μετρικές πιο συνηθισμένη επιλογή. Παρόλα αυτά, σε πειραματικά σενάρια, όπως προσομοιώσεις με γνωστή διάταξη σταθμών βάσης και χρηστών, οι εξωτερικές μετρικές αποτελούν εξαιρετικά χρήσιμο εργαλείο για την επιβεβαίωση της ορθότητας και της αξιοπιστίας των αποτελεσμάτων.

4.3.3 Ανάλυση Κύριων Συνιστωσών (PCA)

Στο πλαίσιο της ανάλυσης των δεδομένων uplink και της ερμηνείας των αποτελεσμάτων της μη εποπτευόμενης μηχανικής μάθησης, αξιοποιείται η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis – PCA) ως εργαλείο διερεύνησης της δομής των δεδομένων [21]. Η PCA αποτελεί μια γραμμική μέθοδο μετασχηματισμού των αρχικών χαρακτηριστικών σε ένα νέο σύστημα αξόνων, όπου οι κύριες συνιστώσες ορίζονται κατά τρόπο ώστε να αποτυπώνουν το μεγαλύτερο μέρος της συνολικής διακύμανσης των δεδομένων.

Μέσω της PCA, τα δεδομένα προβάλλονται σε έναν νέο χώρο αναφοράς, με τις πρώτες κύριες συνιστώσες να συγκεντρώνουν την κυρίαρχη πληροφορία που περιέχεται στα αρχικά χαρακτηριστικά. Στην παρούσα εργασία εξετάζονται οι δύο πρώτες κύριες συνιστώσες, γεγονός που επιτρέπει τη διδιάστατη απεικόνιση των χρηστών και των παραγόμενων clusters, διευκολύνοντας την ποιοτική ανάλυση και την κατανόηση της μεταξύ τους συσχέτισης.

Επισημαίνεται ότι η PCA δεν χρησιμοποιείται ως είσοδος ή κριτήριο στη διαδικασία λήψης αποφάσεων σύνδεσης χρηστών, ούτε επηρεάζει άμεσα την υλοποίηση της πολιτικής ML-DUDe. Ο ρόλος της περιορίζεται αποκλειστικά στην ανάλυση, την οπτικοποίηση και την ερμηνεία των αποτελεσμάτων της μη εποπτευόμενης μάθησης, επιτρέποντας την καλύτερη κατανόηση της συμπεριφοράς των χρηστών ως προς τα χαρακτηριστικά του uplink.

Με τον τρόπο αυτό, η χρήση της PCA συμβάλλει στην ενίσχυση της ερμηνευσιμότητας των αποτελεσμάτων της εργασίας, χωρίς να αλλοιώνει τη φυσική σημασία των χαρακτηριστικών που χρησιμοποιούνται για τη διαδικασία clustering και τη σύγκριση των εξεταζόμενων πολιτικών σύνδεσης.

4.4 Συγκριτική Ανάλυση Μετρικών Clustering

Η επιλογή της κατάλληλης μετρικής για την αξιολόγηση ενός αλγορίθμου clustering αποτελεί κρίσιμο βήμα στη διαδικασία ανάλυσης δεδομένων. Καθώς η μη εποπτευόμενη μάθηση δεν

διαθέτει «σωστές» ετικέτες αναφοράς, η εκτίμηση της ποιότητας των παραγόμενων ομάδων μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα, ανάλογα με το ποια μετρική χρησιμοποιείται. Για τον λόγο αυτόν, στη διεθνή βιβλιογραφία επισημαίνεται συχνά η ανάγκη για συνδυαστική χρήση περισσότερων από μίας μετρικής, ώστε να επιτυγχάνεται μια πιο σφαιρική αποτίμηση της ποιότητας του clustering.

Οι εσωτερικές μετρικές, όπως ο δείκτης Silhouette, ο δείκτης Davies-Bouldin και ο δείκτης Calinski-Harabasz, παρέχουν χρήσιμες ενδείξεις για τη συνοχή και τη διαχωρισιμότητα των ομάδων. Ο δείκτης Silhouette είναι ιδιαίτερα διαισθητικός, καθώς αξιολογεί ταυτόχρονα την εγγύτητα ενός σημείου στο cluster του και την απόστασή του από τα γειτονικά clusters. Ο δείκτης Davies-Bouldin, αντίθετα, επικεντρώνεται στη σχέση ενδο-ομαδικής διασποράς και δια-ομαδικής απόστασης, με χαμηλότερες τιμές να σηματοδοτούν καλύτερη ομαδοποίηση. Ο δείκτης Calinski-Harabasz αξιοποιεί λόγους διασπορών, προτείνοντας υψηλότερες τιμές όταν τα clusters είναι σαφώς διακριτά.

Οι εξωτερικές μετρικές, όπως ο Adjusted Rand Index (ARI) και η Normalized Mutual Information (NMI), έχουν το πλεονέκτημα ότι συγκρίνουν απευθείας την παραγόμενη ομαδοποίηση με μια γνωστή κατανομή αναφοράς. Αυτό τις καθιστά εξαιρετικά αποτελεσματικές σε περιπτώσεις πειραματικών δεδομένων ή προσομοιώσεων, όπου οι πραγματικές ετικέτες είναι διαθέσιμες. Ωστόσο, σε πραγματικά τηλεπικοινωνιακά δεδομένα η απουσία ground truth περιορίζει την εφαρμογή τους.

Η συγκριτική παρουσίαση των μετρικών, με βάση τα πλεονεκτήματα και τα μειονεκτήματά τους, παρουσιάζεται στον ακόλουθο πίνακα. Ο πίνακας αυτός μπορεί να αποτελέσει χρήσιμο εργαλείο για την επιλογή των κατάλληλων μετρικών ανάλογα με το είδος της ανάλυσης και τα διαθέσιμα δεδομένα.

Η παρουσίαση αυτή καταδεικνύει ότι καμία μετρική από μόνη της δεν επαρκεί για την πλήρη αξιολόγηση ενός clustering. Η επιλογή τους θα πρέπει να λαμβάνει υπόψη το είδος των δεδομένων, τη διαθεσιμότητα ετικετών αναφοράς, καθώς και τους στόχους της εκάστοτε ανάλυσης. Στην περίπτωση της παρούσας διπλωματικής, όπου τα δεδομένα προέρχονται από προσομοίωση και η ground truth μπορεί να είναι μερικώς γνωστή, η συνδυαστική χρήση τόσο εσωτερικών όσο και εξωτερικών μετρικών κρίνεται αναγκαία για την αξιόπιστη αποτίμηση των αποτελεσμάτων.

4.4.1 Σύνολο δεδομένων και διαδικασία ομαδοποίησης

Το σύνολο δεδομένων που χρησιμοποιείται στην παρούσα εργασία προκύπτει από προσομοίωση ενός ετερογενούς κυψελωτού δικτύου και περιλαμβάνει πληροφορίες για κάθε μεμονωμένο χρήστη (User Equipment – UE). Συγκεκριμένα, στο εξεταζόμενο σενάριο τοποθετούνται 200 χρήστες στον χώρο προσομοίωσης, οι οποίοι κατανέμονται τυχαία εντός της περιοχής κάλυψης του δικτύου. Για κάθε χρήστη υπολογίζονται γεωμετρικά και ραδιοζεύξης χαρακτηριστικά που σχετίζονται άμεσα με την απόδοση της επικοινωνίας στο uplink.

Κάθε δείγμα του συνόλου δεδομένων αντιστοιχεί σε έναν χρήστη και περιγράφεται από ένα σύνολο χαρακτηριστικών που αποτυπώνουν τόσο τη χωρική σχέση με τους σταθμούς βάσης όσο και την ποιότητα της ασύρματης σύνδεσης. Τα χαρακτηριστικά αυτά περιλαμβάνουν την απόσταση από τον macro σταθμό βάσης και από τον πλησιέστερο pico σταθμό βάσης, τις τιμές του λόγου σήματος προς παρεμβολή και θόρυβο (SINR) προς κάθε τύπο σταθμού βάσης, καθώς και τη χωρητικότητα uplink που αντιστοιχεί σε κάθε σύνδεση. Με τον τρόπο αυτό, το σύνολο δεδομένων ενσωματώνει τόσο πληροφορία θέσης όσο και πληροφορία απόδοσης του καναλιού.

Πριν από την εφαρμογή των αλγορίθμων clustering, τα δεδομένα υφίστανται βασική προεπεξεργασία, η οποία περιλαμβάνει τη διαχείριση πιθανών ελλিপών τιμών μέσω αντικατάστασης με τη διάμεσο κάθε χαρακτηριστικού και την κανονικοποίηση των χαρακτηριστικών με χρήση τυπικής κλιμάκωσης (standardization). Η κανονικοποίηση είναι απαραίτητη ώστε όλα τα χαρακτηριστικά να συνεισφέρουν ισότιμα στη διαδικασία ομαδοποίησης και να μην κυριαρχούν όσα έχουν μεγαλύτερη αριθμητική κλίμακα.

Στη συνέχεια, οι αλγόριθμοι μη εποπτευόμενης μάθησης εφαρμόζονται στο κανονικοποιημένο σύνολο χαρακτηριστικών. Εξετάζονται διάφορες τεχνικές clustering, μεταξύ των οποίων οι KMeans, MiniBatch KMeans, Agglomerative Clustering, Spectral Clustering, Gaussian Mixture Models (GMM), BIRCH, DBSCAN και OPTICS. Κάθε αλγόριθμος παράγει μία διαφορετική ομαδοποίηση των χρηστών, με βάση την ομοιότητα των χαρακτηριστικών τους στο uplink.

Σκοπός της ομαδοποίησης είναι η ανάδειξη ομάδων χρηστών με παρόμοια συμπεριφορά ως προς την ποιότητα σύνδεσης και τη γεωμετρική τους θέση στο δίκτυο. Τα παραγόμενα clusters

χρησιμοποιούνται στη συνέχεια τόσο για τη συγκριτική αξιολόγηση των αλγορίθμων clustering όσο και για την ανάλυση της αποτελεσματικότητας της πολιτικής σύνδεσης ML-DUDe.

Η ποιότητα των παραγόμενων ομαδοποιήσεων αξιολογείται μέσω εσωτερικών και εξωτερικών μετρικών clustering, με έμφαση στις εσωτερικές μετρικές λόγω της απουσίας πραγματικών ετικετών αναφοράς σε ρεαλιστικά τηλεπικοινωνιακά σενάρια. Μεταξύ αυτών περιλαμβάνονται οι δείκτες Silhouette, Davies–Bouldin και Calinski–Harabasz, οι οποίοι επιτρέπουν την αποτίμηση της συνοχής των clusters και του βαθμού διαχωρισμού τους.

4.4.2 Συγκριτική αξιολόγηση αλγορίθμων clustering και τελική επιλογή

Στην παρούσα ενότητα πραγματοποιείται συγκριτική αξιολόγηση των αλγορίθμων μη εποπτευόμενης μάθησης που εξετάστηκαν στα προηγούμενα υποκεφάλαια, με στόχο την επιλογή της καταλληλότερης μεθόδου clustering για την υλοποίηση της προσέγγισης ML-assisted DUDe. Η αξιολόγηση βασίζεται σε ποσοτικούς δείκτες ποιότητας ομαδοποίησης, καθώς και σε ποιοτικά κριτήρια που σχετίζονται με τη σταθερότητα των clusters και τη δυνατότητα εξαγωγής πρακτικής πολιτικής συσχέτισης uplink.

Η συγκριτική παρουσίαση των μετρικών, με βάση τα πλεονεκτήματα και τα μειονεκτήματά τους, παρουσιάζεται στον Πίνακα 3 που ακολουθεί. Ο πίνακας αυτός μπορεί να αποτελέσει χρήσιμο εργαλείο για την επιλογή των κατάλληλων μετρικών ανάλογα με το είδος της ανάλυσης και τα διαθέσιμα δεδομένα

Πίνακας 3: Συγκριτική παρουσίαση αλγορίθμων ομαδοποίησης

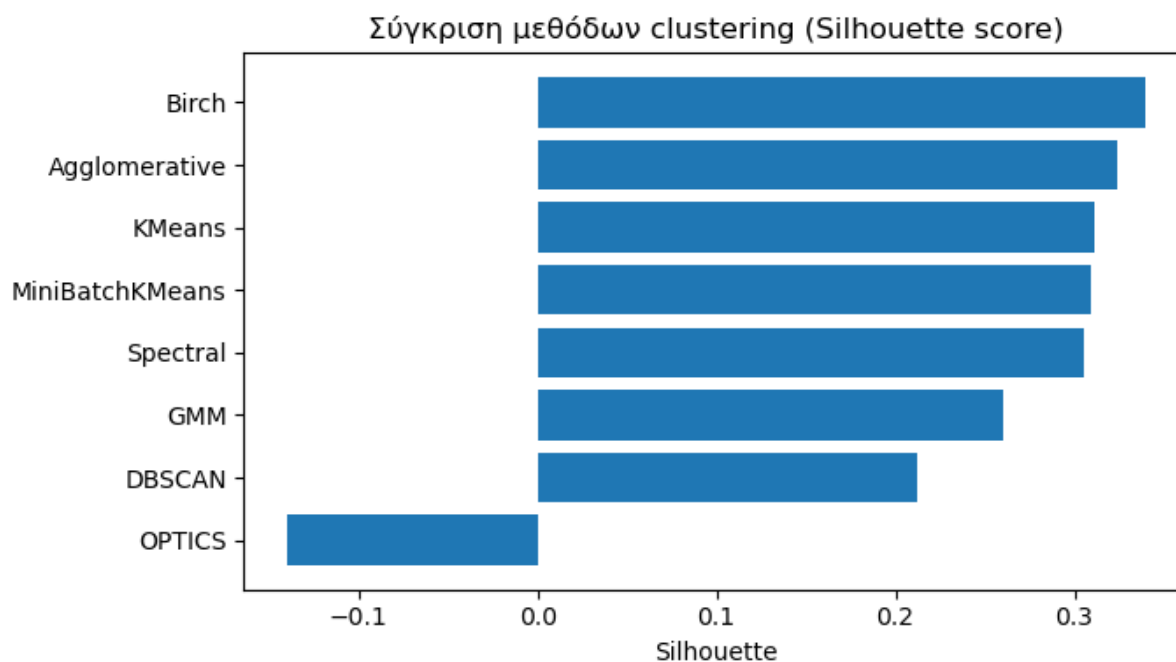
Αλγόριθμος	Αριθμός ομάδων (clusters)	Πλεονεκτήματα	Περιορισμοί
K-Means	k (προκαθορισμένος)	Γρήγορος, απλός, εύκολα ερμηνεύσιμος	Απαιτεί εκ των προτέρων τον αριθμό ομάδων, ευαίσθητος στην αρχικοποίηση
MiniBatch K-Means	k (προκαθορισμένος)	Κατάλληλος για μεγάλα σύνολα δεδομένων, χαμηλό υπολογιστικό κόστος	Προσεγγιστική λύση, πιθανή απώλεια ακρίβειας
Gaussian Mixture Model (GMM)	k (προκαθορισμένος)	Μαλακή ομαδοποίηση, ευέλικτα σχήματα clusters	Ευαίσθησία στην αρχικοποίηση, αυξημένη πολυπλοκότητα
Ιεραρχική ομαδοποίηση (Ward)	Μεταβλητός	Δεν απαιτεί προκαθορισμό k, καλή ερμηνευσιμότητα	Υψηλό υπολογιστικό κόστος για μεγάλα σύνολα δεδομένων
DBSCAN	Μεταβλητός	Εντοπισμός θορύβου, clusters αυθαίρετου σχήματος	Ευαίσθησία στις παραμέτρους πυκνότητας
OPTICS	Μεταβλητός	Διαχειρίζεται μεταβαλλόμενη πυκνότητα	Δυσκολότερη ερμηνεία αποτελεσμάτων

Αλγόριθμος	Αριθμός ομάδων (clusters)	Πλεονεκτήματα	Περιορισμοί
BIRCH	Μεταβλητός	Αποδοτικός για μεγάλης κλίμακας δεδομένα	Μειωμένη ακρίβεια σε πολύπλοκες δομές

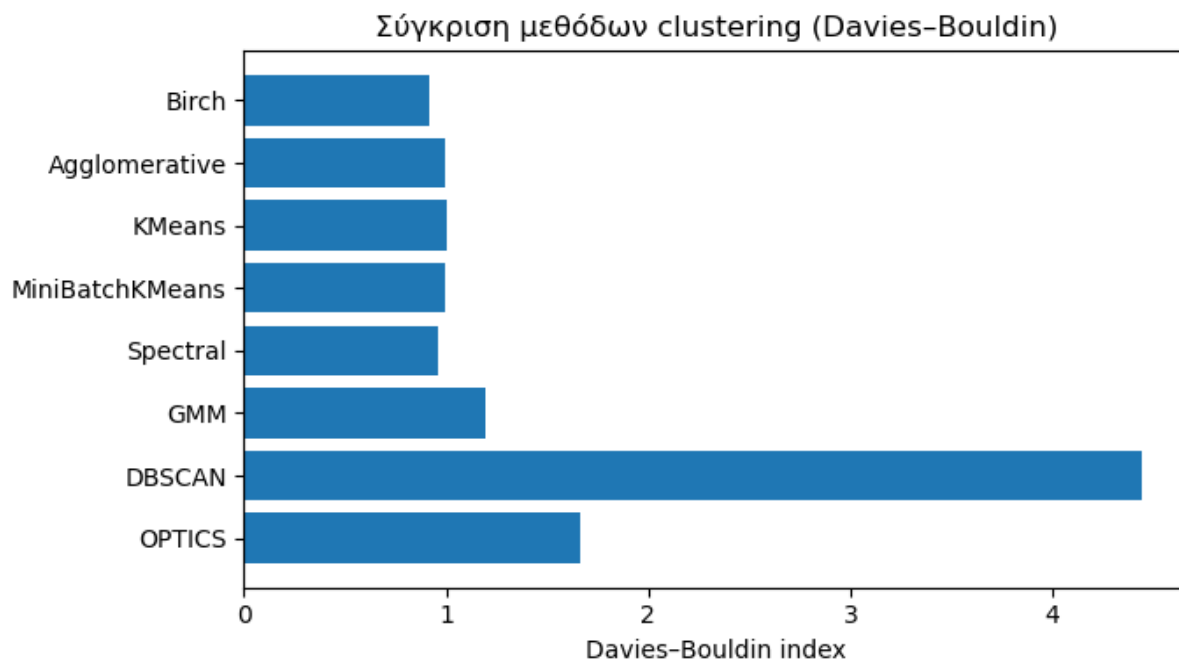
4.4.3 Ποσοτική αξιολόγηση ποιότητας clustering

Για την αντικειμενική σύγκριση των αλγορίθμων χρησιμοποιήθηκαν καθιερωμένοι δείκτες ποιότητας clustering. Συγκεκριμένα, ο δείκτης Silhouette χρησιμοποιήθηκε για την αξιολόγηση της συνοχής και του διαχωρισμού των clusters, ενώ ο δείκτης Davies–Bouldin χρησιμοποιήθηκε συμπληρωματικά για την εκτίμηση της μεταξύ τους διακριτότητας. Οι δείκτες αυτοί επιτρέπουν τη σύγκριση αλγορίθμων με διαφορετική εσωτερική λογική, χωρίς την ανάγκη ύπαρξης ετικετών αναφοράς.

Στις Εικόνες 1 και 2 που ακολουθούν, βλέπουμε τα αποτελέσματα της σύγκρισης Silhouette και Davies-Bouldin αντίστοιχα, για τους αλγορίθμους που αναλύθηκαν παραπάνω.



Εικόνα 1: Αποτελέσματα σύγκρισης Silhouette



Εικόνα 2: Αποτελέσματα σύγκρισης Davies-Bouldin

Από τα αποτελέσματα, παρατηρείται ότι οι αλγόριθμοι KMeans, BIRCH, Agglomerative και Spectral Clustering παρουσιάζουν συγκρίσιμη και υψηλή ποιότητα ομαδοποίησης, με τον αλγόριθμο BIRCH να εμφανίζει την υψηλότερη τιμή Silhouette. Ο αλγόριθμος KMeans ακολουθεί με ελαφρώς χαμηλότερη αλλά πρακτικά ισοδύναμη απόδοση, παρουσιάζοντας παράλληλα χαμηλή τιμή Davies–Bouldin, γεγονός που υποδηλώνει καλά διαχωρισμένα clusters.

Αντίθετα, οι αλγόριθμοι DBSCAN και OPTICS εμφανίζουν σημαντικά υποδεέστερη απόδοση, με χαμηλές ή αρνητικές τιμές Silhouette και ιδιαίτερα υψηλές τιμές Davies–Bouldin, γεγονός που καταδεικνύει ακατάλληλη ομαδοποίηση στο εξεταζόμενο σύνολο δεδομένων. Ως εκ τούτου, οι συγκεκριμένες μέθοδοι αποκλείονται από τη συνέχεια της ανάλυσης.

4.4.4 Τελική επιλογή αλγορίθμων clustering

Λαμβάνοντας υπόψη τα αποτελέσματα της ποσοτικής αξιολόγησης, ο αλγόριθμος KMeans επιλέγεται ως η τελική μέθοδος clustering για τη συνέχεια της εργασίας. Παρότι ο αλγόριθμος BIRCH παρουσιάζει οριακά υψηλότερη τιμή Silhouette, ο KMeans εμφανίζει συγκρίσιμη ποιότητα ομαδοποίησης, ενώ παράλληλα διακρίνεται για την απλότητα, τη σταθερότητα και την ευκολία ερμηνείας των παραγόμενων clusters. Τα χαρακτηριστικά αυτά καθιστούν τον KMeans ιδιαίτερα κατάλληλο για την εξαγωγή σαφών και επαναλήψιμων ομάδων χρηστών, οι

οποίες μπορούν να αξιοποιηθούν άμεσα για την εφαρμογή πολιτικών συσχέτισης uplink στα επόμενα κεφάλαια της μελέτης.

Η επιλογή του K-Means στην παρούσα εργασία βασίζεται στη σταθερότητα των αποφάσεων σύνδεσης και στην ερμηνευσιμότητα των clusters, και όχι στη βέλτιστη θεωρητική αναπαράσταση των δεδομένων.

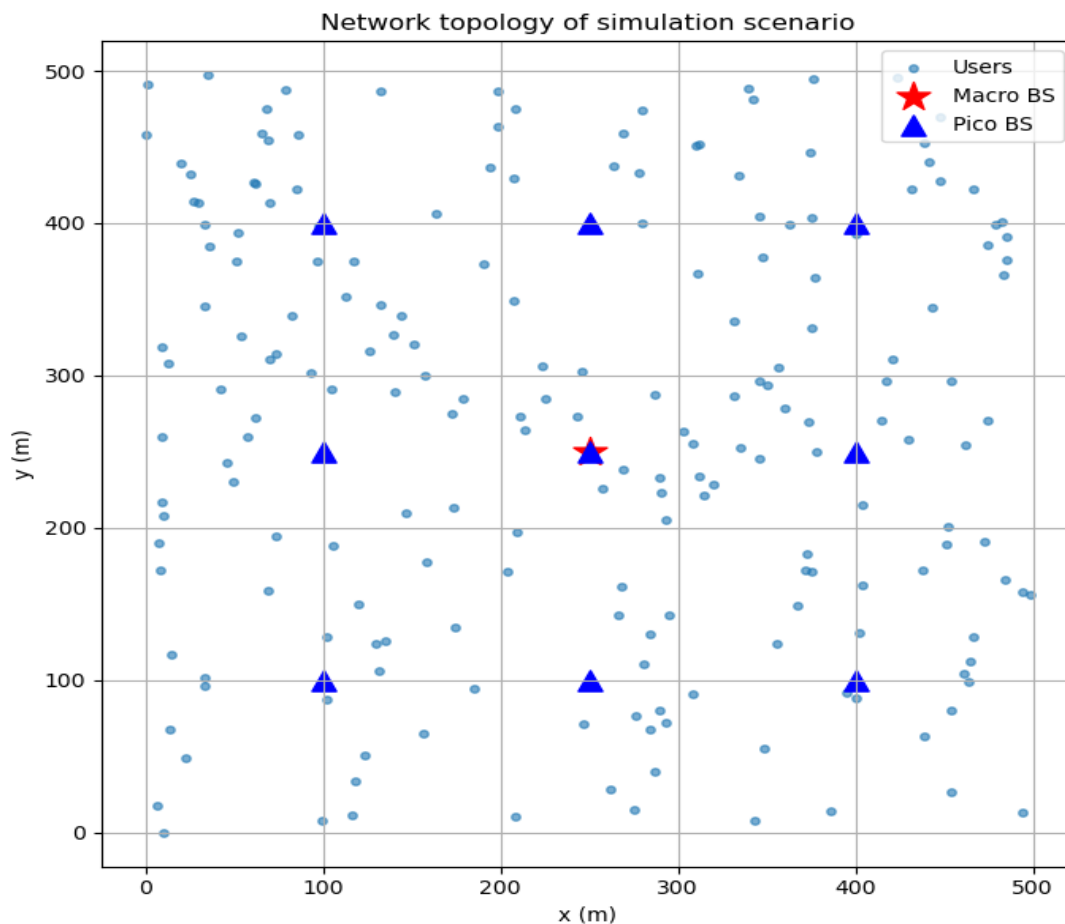
Στα επόμενα κεφάλαια, τα παραγόμενα clusters αξιοποιούνται για τη διαμόρφωση πολιτικών συσχέτισης uplink, με στόχο την προσέγγιση της συμπεριφοράς του DUDe μέσω μη επιβλεπόμενης μάθησης.

5

Πειραματικά Αποτελέσματα

Στο παρόν κεφάλαιο παρουσιάζονται και αναλύονται τα αποτελέσματα της προσομοίωσης που πραγματοποιήθηκε με στόχο την αξιολόγηση της απόδοσης της τεχνολογίας Decoupled Uplink and Downlink Access (DUDe), καθώς και της προτεινόμενης μη εποπτευόμενης προσέγγισης ML-assisted DUDe. Αρχικά εξετάζεται η συμπεριφορά του συστήματος στο βασικό σενάριο συσχέτισης χρηστών–σταθμών βάσης, όπου η uplink σύνδεση ακολουθεί το downlink. Στη συνέχεια αξιολογείται η βελτίωση που προσφέρει το DUDe στο uplink, τόσο ως προς τη χωρητικότητα όσο και ως προς την κατανομή φορτίου. Τέλος, παρουσιάζονται τα αποτελέσματα της μη εποπτευόμενης μηχανικής μάθησης, με στόχο την ποσοτική σύγκριση της ML-DUDe προσέγγισης με ένα θεωρητικά βέλτιστο σενάριο σύνδεσης uplink, στο οποίο κάθε χρήστης συνδέεται στον σταθμό βάσης που μεγιστοποιεί τη χωρητικότητα.

Η χωρική διάταξη των macro και pico σταθμών βάσης, καθώς και η τυχαία κατανομή των χρηστών στην περιοχή προσομοίωσης, απεικονίζονται σχηματικά ώστε να καταδεικνύεται η γεωμετρική σχέση των χρηστών με το δίκτυο.



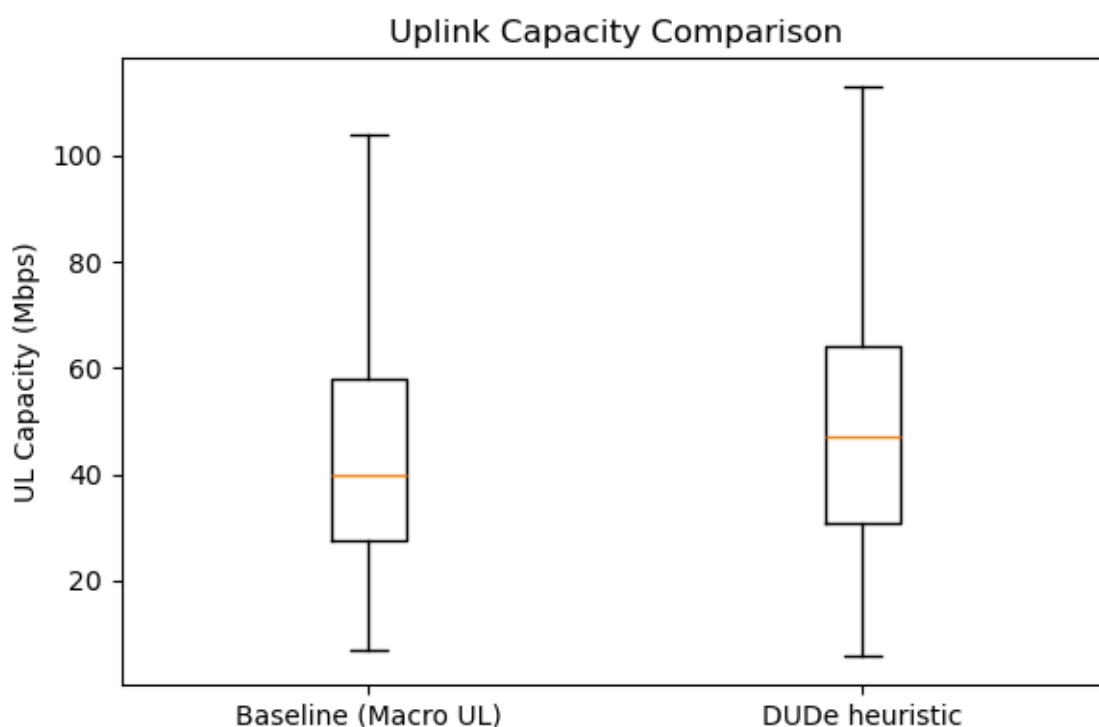
Εικόνα 3: Τοπολογία δικτύου υλοποίησης

Η Εικόνα 3 απεικονίζει την τοπολογία του δικτύου που χρησιμοποιήθηκε στο σενάριο προσομοίωσης. Ο macro σταθμός βάσης τοποθετείται στο κέντρο της περιοχής κάλυψης και αναπαρίσταται με κόκκινο σύμβολο, ενώ οι pico σταθμοί βάσης κατανέμονται ομοιόμορφα σε διάφορα σημεία του χώρου και σημειώνονται με μπλε τρίγωνα. Οι χρήστες απεικονίζονται με μπλε κυκλικά σημεία και είναι τυχαία κατανεμημένοι εντός της περιοχής, προσομοιώνοντας ρεαλιστικές συνθήκες χωρικής διασποράς.

Η συγκεκριμένη διάταξη αντικατοπτρίζει μια ετερογενή αρχιτεκτονική δικτύου (HetNet), όπου ο macro σταθμός παρέχει ευρεία κάλυψη, ενώ οι pico κυψέλες ενισχύουν την χωρητικότητα σε περιοχές με υψηλότερη πυκνότητα χρηστών. Η τοπολογία αυτή επιτρέπει τη μελέτη μηχανισμών αποσύζευξης uplink και downlink (DUDe), καθώς και την αξιολόγηση της κατανομής χρηστών μεταξύ macro και pico σταθμών βάσης υπό διαφορετικές πολιτικές σύνδεσης.

5.1 Απόδοση uplink στο βασικό σενάριο (Baseline)

Στο βασικό σενάριο, η συσχέτιση των χρηστών με τον σταθμό βάσης στο uplink ακολουθεί αποκλειστικά την επιλογή που πραγματοποιείται στο downlink. Η προσέγγιση αυτή είναι η πλέον διαδεδομένη στα σύγχρονα κυψελωτά δίκτυα, ωστόσο δεν λαμβάνει υπόψη τις ιδιαιτερότητες του uplink καναλιού, όπως τη διαφορά ισχύος εκπομπής των χρηστών και την αυξημένη ευαισθησία σε παρεμβολές.



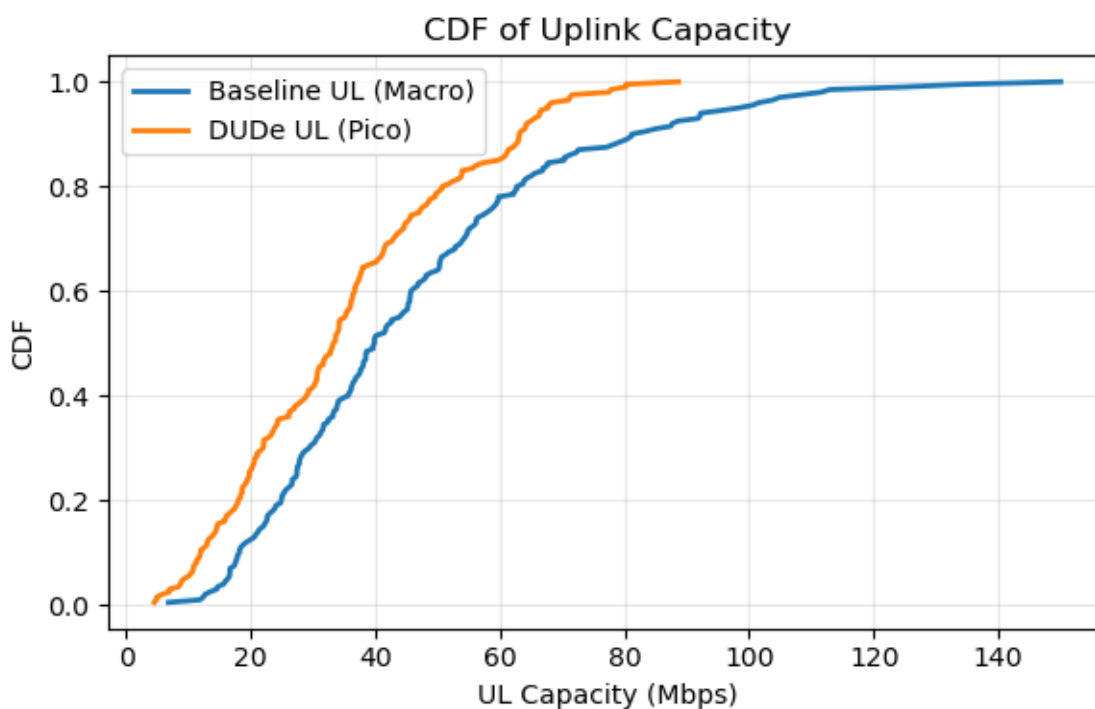
Εικόνα 4: Σύγκριση χωρητικότητας uplink

Όπως φαίνεται στην Εικόνα 4 παραπάνω, η κατανομή της uplink χωρητικότητας στο βασικό σενάριο σύνδεσης μέσω macro σταθμού βάσης παρουσιάζει μεγάλη διασπορά και σχετικά χαμηλές τιμές για σημαντικό ποσοστό των χρηστών. Το φαινόμενο αυτό υποδηλώνει ότι αρκετοί χρήστες εξυπηρετούνται από απομακρυσμένες κυψέλες macro, γεγονός που οδηγεί σε χαμηλές τιμές SINR και, κατά συνέπεια, σε μειωμένη απόδοση στο uplink. Αντίθετα, η εφαρμογή της πολιτικής DUDe οδηγεί σε εμφανή αύξηση της μέσης χωρητικότητας uplink, καθώς και σε μικρότερη διασπορά των τιμών, υποδεικνύοντας πιο ομοιόμορφη και αποδοτική κατανομή των χρηστών στους διαθέσιμους σταθμούς βάσης. Η βελτίωση αυτή καταδεικνύει

ότι η αποσυνδεδεμένη σύνδεση uplink επιτρέπει την αξιοποίηση πλησιέστερων μικροκυψελών, μειώνοντας το bottleneck του uplink και ενισχύοντας τη συνολική απόδοση του συστήματος.

5.2 Βελτίωση uplink μέσω DUDe

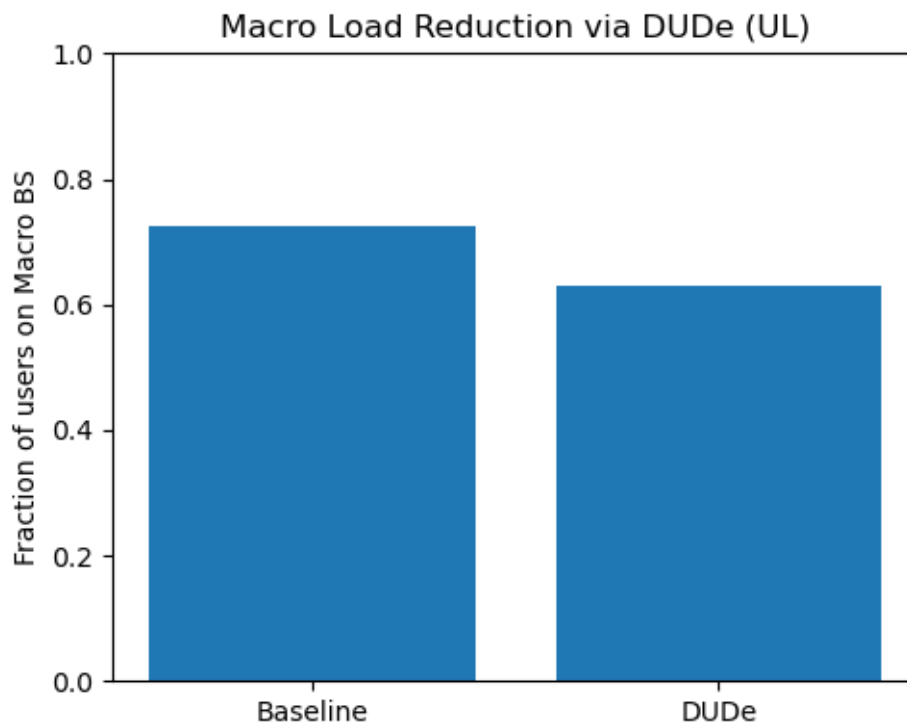
Στην ενότητα αυτή αξιολογείται η επίδραση της τεχνολογίας DUDe στην uplink απόδοση. Η βασική ιδέα του DUDe είναι η ανεξάρτητη επιλογή σταθμού βάσης για το uplink, επιτρέποντας στους χρήστες να συνδέονται με τον σταθμό που προσφέρει καλύτερες συνθήκες uplink, ανεξάρτητα από την επιλογή στο downlink.



Εικόνα 5: CDF της uplink χωρητικότητας για Baseline (Macro UL) και DUDe (Pico UL)

Η σωρευτική κατανομή της uplink χωρητικότητας στην Εικόνα 5, επιβεβαιώνει ότι η βελτίωση που προσφέρει η πολιτική DUDe δεν περιορίζεται μόνο στον μέσο όρο, αλλά επεκτείνεται σε ολόκληρη την κατανομή των χρηστών. Παρατηρείται σαφής μετατόπιση της καμπύλης DUDe προς υψηλότερες τιμές χωρητικότητας σε σύγκριση με το βασικό σενάριο σύνδεσης μέσω macro σταθμού βάσης. Ενδεικτικά, για τιμές uplink χωρητικότητας της τάξης των 40 Mbps, το ποσοστό των χρηστών που τις επιτυγχάνουν στο σενάριο DUDe είναι σημαντικά υψηλότερο

σε σχέση με το baseline σενάριο. Αυτό υποδεικνύει ότι μεγαλύτερο μέρος του πληθυσμού χρηστών επωφελείται από τη σύνδεση σε πλησιέστερους pico σταθμούς βάσης, βελτιώνοντας όχι μόνο τις κορυφαίες αλλά και τις χαμηλότερες αποδόσεις. Η συμπεριφορά αυτή καταδεικνύει ότι το DUDe συμβάλλει στη συνολική αναβάθμιση της εμπειρίας χρήστη, ιδιαίτερα για εκείνους που στο βασικό σενάριο παρουσίαζαν χαμηλή uplink απόδοση.



Εικόνα 6: Μείωση φορτίου macro σταθμού βάσης στο uplink μέσω DUDe

Η Εικόνα 6 παρουσιάζει το ποσοστό των χρηστών που εξυπηρετούνται από macro σταθμούς βάσης στο uplink για το βασικό σενάριο σύζευξης (Baseline) και για το σενάριο εφαρμογής της τεχνολογίας DUDe. Παρατηρείται ότι στο baseline περίπου το 72,5% των χρηστών συνδέεται στο uplink με macro σταθμούς βάσης, ενώ το υπόλοιπο 27,5% εξυπηρετείται από pico κυψέλες, γεγονός που οφείλεται στη γεωγραφική κατανομή των χρηστών και στη σχετική εγγύτητά τους προς τους μικροκυψελωτούς σταθμούς.

Με την εφαρμογή της τεχνολογίας DUDe, το ποσοστό των χρηστών που συνδέονται με macro σταθμούς βάσης μειώνεται στο 63%, ενώ αυξάνεται αντίστοιχα το ποσοστό των χρηστών που εξυπηρετούνται από pico κυψέλες στο uplink. Η μεταβολή αυτή καταδεικνύει ότι το DUDe επιτυγχάνει πιο αποδοτική εκμετάλλευση της ετερογενούς αρχιτεκτονικής του δικτύου,

κατευθύνοντας μέρος των χρηστών προς τους πλησιέστερους pico σταθμούς βάσης, όπου οι συνθήκες uplink είναι ευνοϊκότερες.

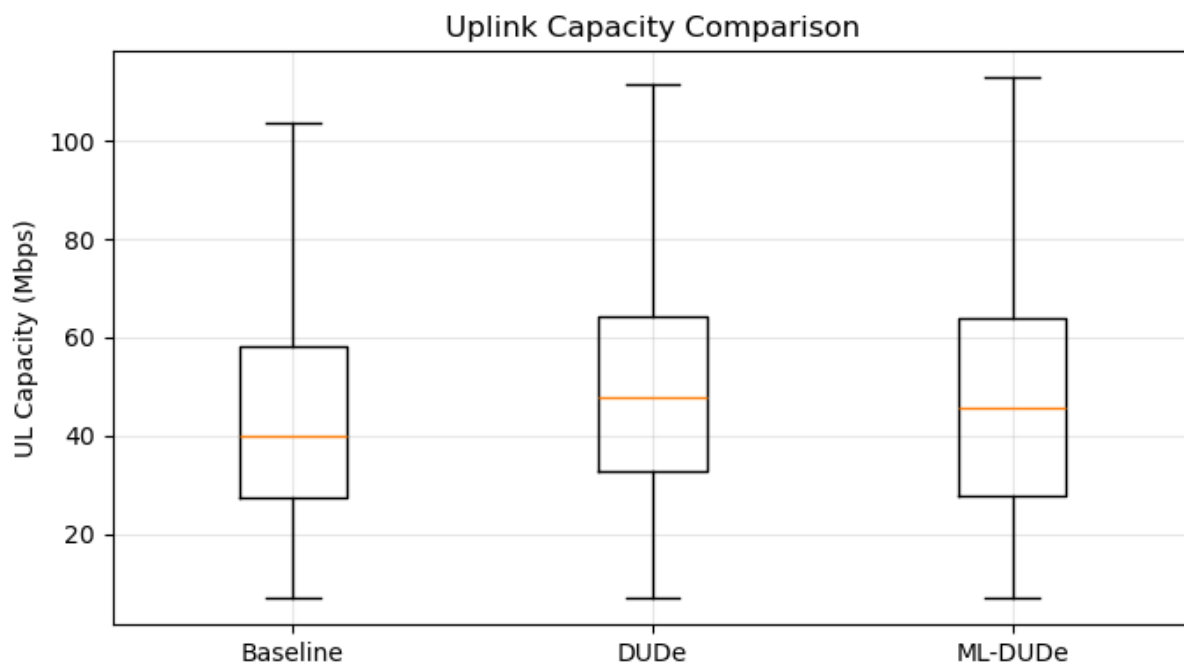
Η αποσυμφόρηση των macro κυψελών στο uplink οδηγεί σε μείωση των παρεμβολών και σε καλύτερη κατανομή των ραδιοπόρων, γεγονός που συμβάλλει άμεσα στη συνολική αύξηση της uplink χωρητικότητας που παρατηρείται στα προηγούμενα αποτελέσματα. Επιπλέον, η μετατόπιση χρηστών προς τα pico cells επιβεβαιώνει τον βασικό στόχο της τεχνολογίας DUDe, δηλαδή τον διαχωρισμό των συνδέσεων uplink και downlink με γνώμονα τη βέλτιστη ποιότητα καναλιού και όχι αποκλειστικά τη μέγιστη ισχύ εκπομπής του macro σταθμού βάσης.

Συνεπώς, το γράφημα αναδεικνύει με σαφήνεια τον ρόλο του DUDe στη βελτίωση της εξισορρόπησης φορτίου του δικτύου (load balancing), στοιχείο κρίσιμο για την αποδοτική λειτουργία ετερογενών τηλεπικοινωνιακών συστημάτων υψηλής πυκνότητας χρηστών.

Στη συνέχεια εξετάζεται η κατανομή του φορτίου στο uplink, μέσω του ποσοστού των χρηστών που εξυπηρετούνται από τον macro σταθμό βάσης για τα σενάρια Baseline, DUDe και ML-DUDe.

5.3 Απόδοση μη εποπτευόμενης προσέγγισης ML-DUDe

Παρότι το DUDe προσφέρει σημαντικά οφέλη, η πρακτική υλοποίησή του μπορεί να απαιτεί πρόσθετη σηματοδосία και πληροφορία καναλιού. Για τον λόγο αυτό, διερευνάται η δυνατότητα προσέγγισης της συμπεριφοράς του DUDe μέσω μη εποπτευόμενης μηχανικής μάθησης.



Εικόνα 7: Σύγκριση της uplink χωρητικότητας για τα σενάρια Baseline, DUDe και ML-DUDe

Στην Εικόνα 7 παρουσιάζεται η σύγκριση της uplink χωρητικότητας για τα σενάρια Baseline, DUDe και ML-DUDe. Παρατηρείται ότι τόσο το DUDe όσο και η προτεινόμενη προσέγγιση ML-DUDe επιτυγχάνουν σαφώς υψηλότερες τιμές uplink χωρητικότητας σε σχέση με το βασικό σενάριο σύζευξης, με μετατόπιση της κατανομής προς μεγαλύτερες τιμές και αύξηση της διάμεσης χωρητικότητας.

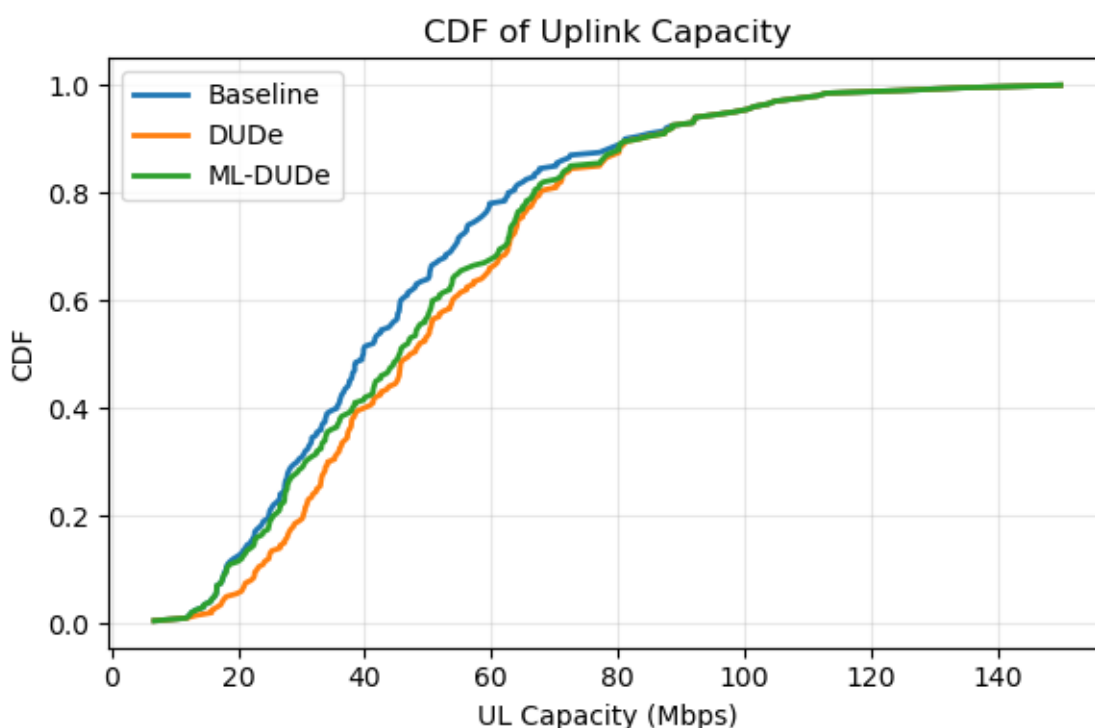
Η βελτίωση αυτή συνδέεται άμεσα με τον τρόπο με τον οποίο το ML-DUDe πραγματοποιεί την ομαδοποίηση των χρηστών. Συγκεκριμένα, οι χρήστες ομαδοποιούνται βάσει χαρακτηριστικών όπως η απόσταση από τον macro και τον πλησιέστερο pico σταθμό βάσης, οι αντίστοιχες τιμές SINR και η εκτιμώμενη uplink χωρητικότητα προς κάθε τύπο σταθμού βάσης. Στη συνέχεια, για κάθε ομάδα χρηστών λαμβάνεται απόφαση σύνδεσης στο uplink προς τον σταθμό βάσης που προσφέρει την υψηλότερη αναμενόμενη απόδοση.

Όπως προκύπτει και από την ανάλυση φορτίου, στο βασικό σενάριο περίπου το 72,5% των χρηστών εξυπηρετείται από macro σταθμούς βάσης στο uplink, ενώ με την εφαρμογή του DUDe το ποσοστό αυτό μειώνεται στο 63%, με αντίστοιχη αύξηση των συνδέσεων προς pico κυψέλες. Η ίδια τάση ακολουθείται και από το ML-DUDe, το οποίο ανακατανέμει αποτελεσματικά τους χρήστες προς τους πλησιέστερους και καταλληλότερους pico σταθμούς. Η αποσυμφόρηση των macro κυψελών και η βελτιωμένη εκμετάλλευση των μικροκυψελών οδηγούν σε μείωση παρεμβολών και σε αύξηση της διαθέσιμης uplink χωρητικότητας για

σημαντικό μέρος των χρηστών, γεγονός που αποτυπώνεται στη συνολική μετατόπιση των boxplots προς υψηλότερες τιμές.

Συνεπώς, το ML-DUDe καταφέρνει να προσεγγίσει την απόδοση της ιδανικής πολιτικής DUDe με καθαρά δεδομενοκεντρικό τρόπο, χωρίς ρητή χρήση κανόνων σύνδεσης, επιβεβαιώνοντας ότι οι τεχνικές μη εποπτευόμενης μάθησης μπορούν να συλλάβουν αποτελεσματικά τη δομή των uplink συνθηκών σε ετερογενή δίκτυα.

Έχοντας αναδείξει τη βελτίωση της uplink απόδοσης μέσω της πολιτικής DUDe, εξετάζεται στη συνέχεια κατά πόσο η προτεινόμενη προσέγγιση ML-DUDe μπορεί να αναπαράγει τη συμπεριφορά αυτή σε επίπεδο συνολικής κατανομής χωρητικότητας. Για τον σκοπό αυτό, παρουσιάζεται η αθροιστική συνάρτηση κατανομής (CDF) της uplink χωρητικότητας για τα σενάρια Baseline, DUDe και ML-DUDe, επιτρέποντας άμεση σύγκριση της απόδοσης των τριών πολιτικών σύνδεσης για το σύνολο των χρηστών.

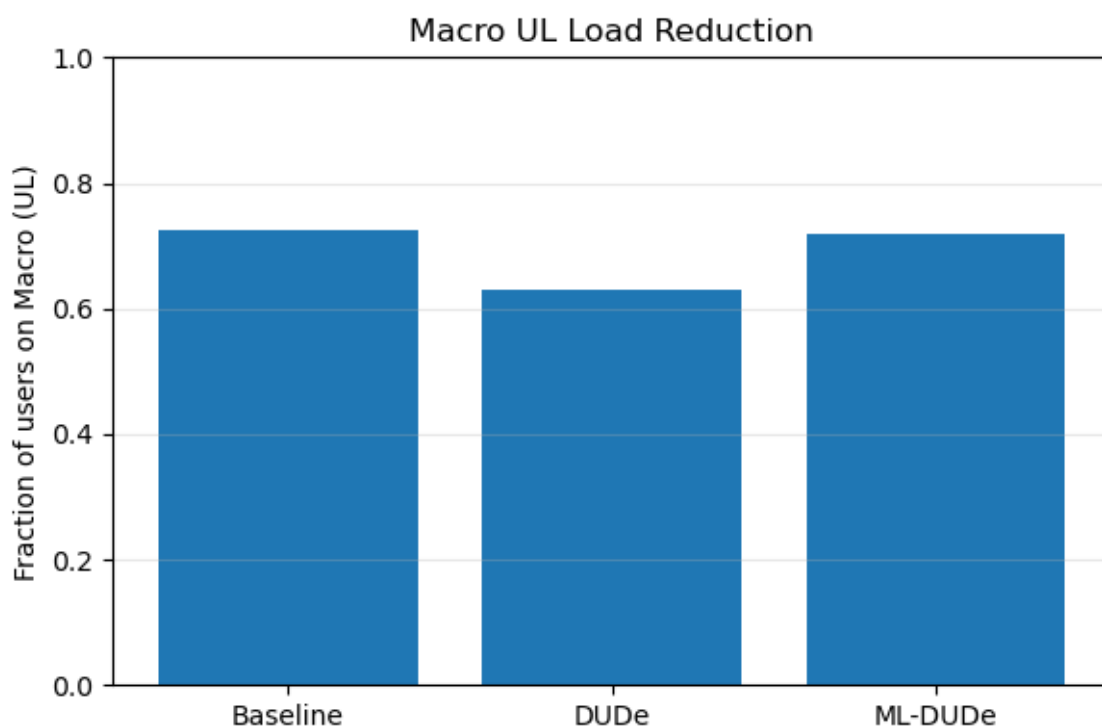


Εικόνα 8: CDF της uplink χωρητικότητας για τα σενάρια Baseline, DUDe και ML-DUDe

Όπως φαίνεται στην Εικόνα 8, οι καμπύλες DUDe και ML-DUDe παρουσιάζουν σαφή μετατόπιση προς υψηλότερες τιμές χωρητικότητας σε σύγκριση με το βασικό σενάριο, επιβεβαιώνοντας τη γενικευμένη βελτίωση της uplink απόδοσης. Η προσέγγιση ML-DUDe προσεγγίζει σε μεγάλο βαθμό τη συμπεριφορά του DUDe σε όλο το εύρος της κατανομής,

γεγονός που υποδηλώνει ότι η μη εποπτευόμενη μάθηση καταφέρνει να συλλάβει τη βασική δομή των uplink συνθηκών.

Στην προηγούμενη ενότητα αναδείχθηκε η μείωση του φορτίου στους macro σταθμούς βάσης στο uplink μέσω της εφαρμογής της πολιτικής DUDe. Στη συνέχεια εξετάζεται κατά πόσο η προτεινόμενη προσέγγιση ML-DUDe επιτυγχάνει αντίστοιχη ανακατανομή των χρηστών, συγκρίνοντας το ποσοστό των χρηστών που εξυπηρετούνται από macro σταθμούς βάσης για τα σενάρια Baseline, DUDe και ML-DUDe.



Εικόνα 9: Μείωση φορτίου macro σταθμού βάσης στο uplink για τα σενάρια Baseline, DUDe και ML-DUDe

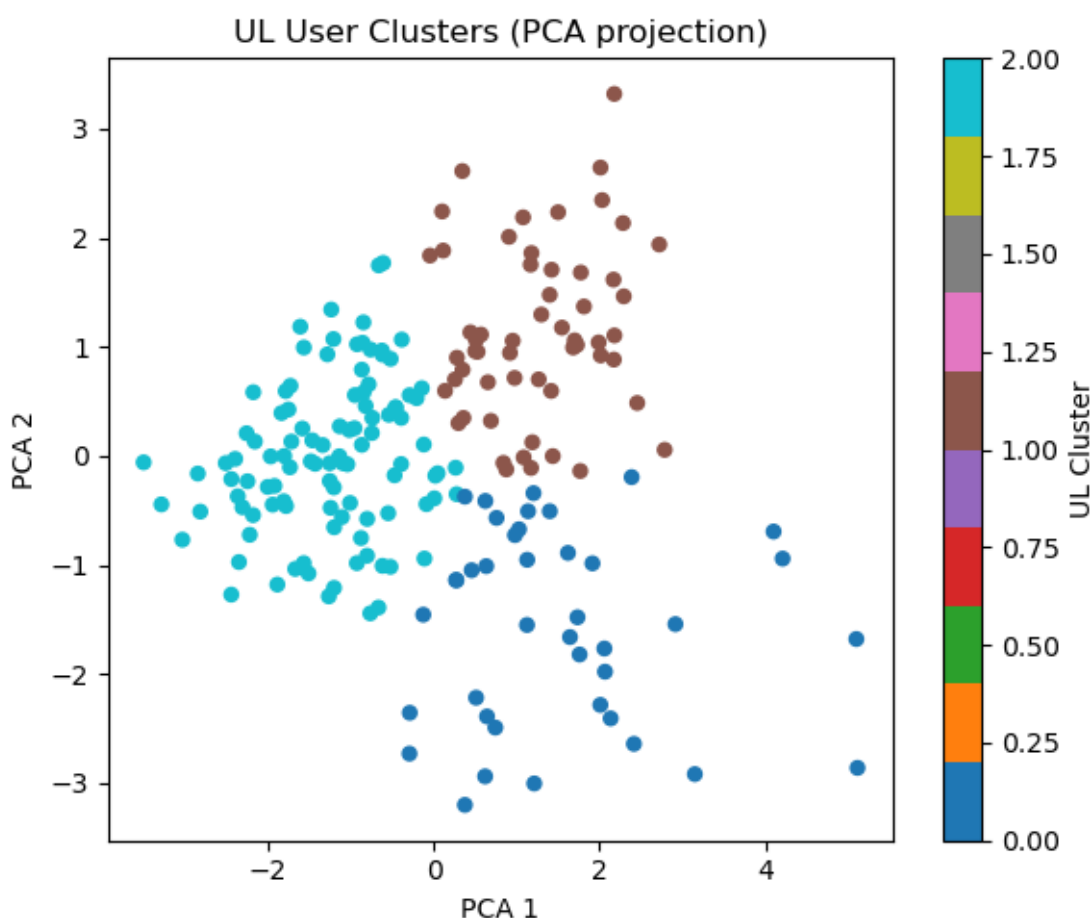
Όπως φαίνεται στην Εικόνα 9, στο βασικό σενάριο περίπου το 72.5% των χρηστών συνδέεται με τον macro σταθμό βάσης στο uplink, γεγονός που υποδηλώνει σημαντική συγκέντρωση φορτίου. Με την εφαρμογή της πολιτικής DUDe, το ποσοστό αυτό μειώνεται στο 63%, αποδεικνύοντας ότι η αποσύζευξη uplink–downlink οδηγεί σε αποτελεσματικότερη κατανομή των χρηστών και αποσυμφόρηση του macro σταθμού βάσης.

Αντίθετα, η προσέγγιση ML-DUDe παρουσιάζει ποσοστό περίπου 72%, προσεγγίζοντας το baseline σενάριο, γεγονός που δείχνει ότι η μη επιβλεπόμενη μέθοδος δεν επιτυγχάνει στον ίδιο βαθμό τη βέλτιστη μεταφορά χρηστών προς τους pico σταθμούς στο uplink. Παρόλα αυτά, η

διαφοροποίηση από το baseline υποδηλώνει ότι η ML-DUDe επηρεάζει τη συμπεριφορά συσχέτισης, αν και με περιορισμένο όφελος ως προς τη μείωση φορτίου.

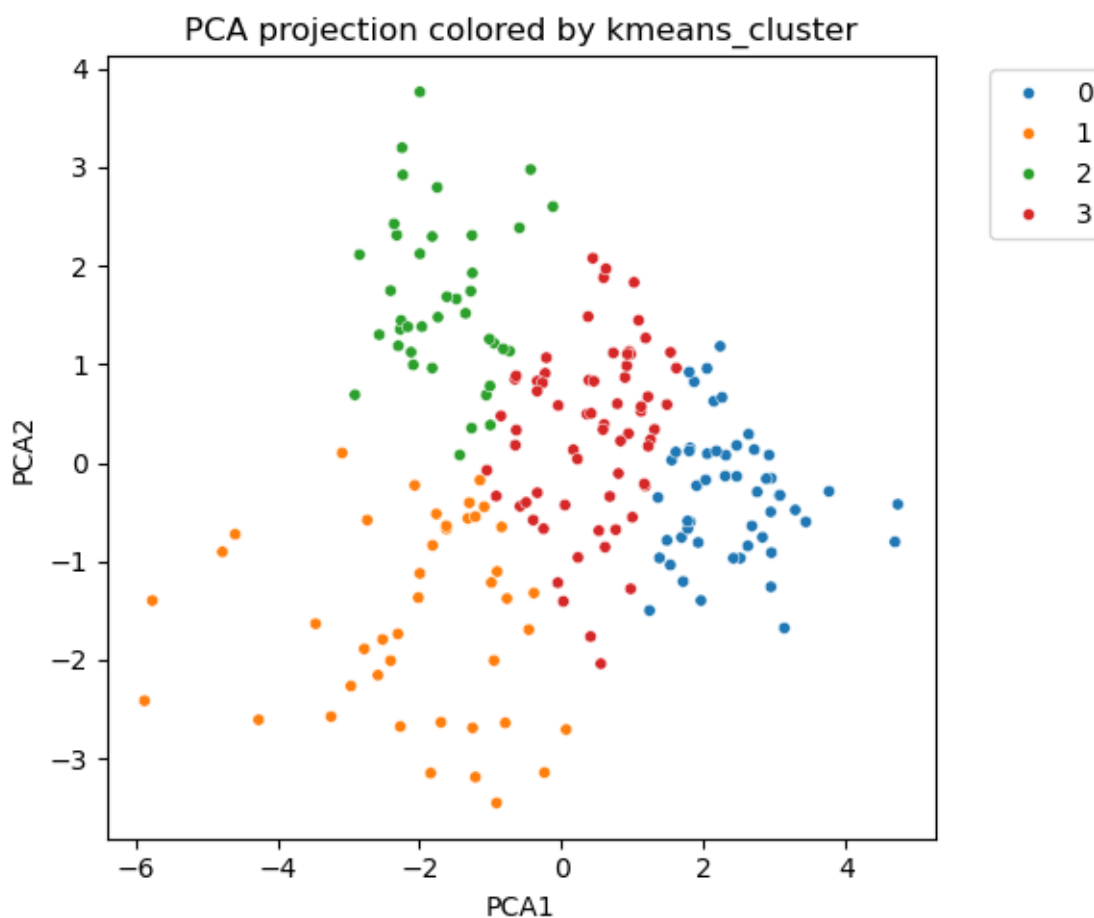
5.4 Δομή δεδομένων και ερμηνεία των αποτελεσμάτων

Για την κατανόηση της λειτουργίας της ML-DUDe προσέγγισης, εξετάζεται η δομή των δεδομένων στο χώρο χαρακτηριστικών.



Εικόνα 10: Ομαδοποίηση χρηστών στο uplink σε προβολή PCA

Η προβολή των χρηστών στον μειωμένο χώρο χαρακτηριστικών που φαίνεται στην Εικόνα 10, καταδεικνύει την ύπαρξη διακριτών ομάδων χρηστών στο uplink. Η παρατήρηση αυτή επιβεβαιώνει ότι τα δεδομένα δεν κατανέμονται τυχαία, αλλά παρουσιάζουν εγγενή δομή, η οποία μπορεί να αξιοποιηθεί για την εξαγωγή πολιτικών συσχέτισης.

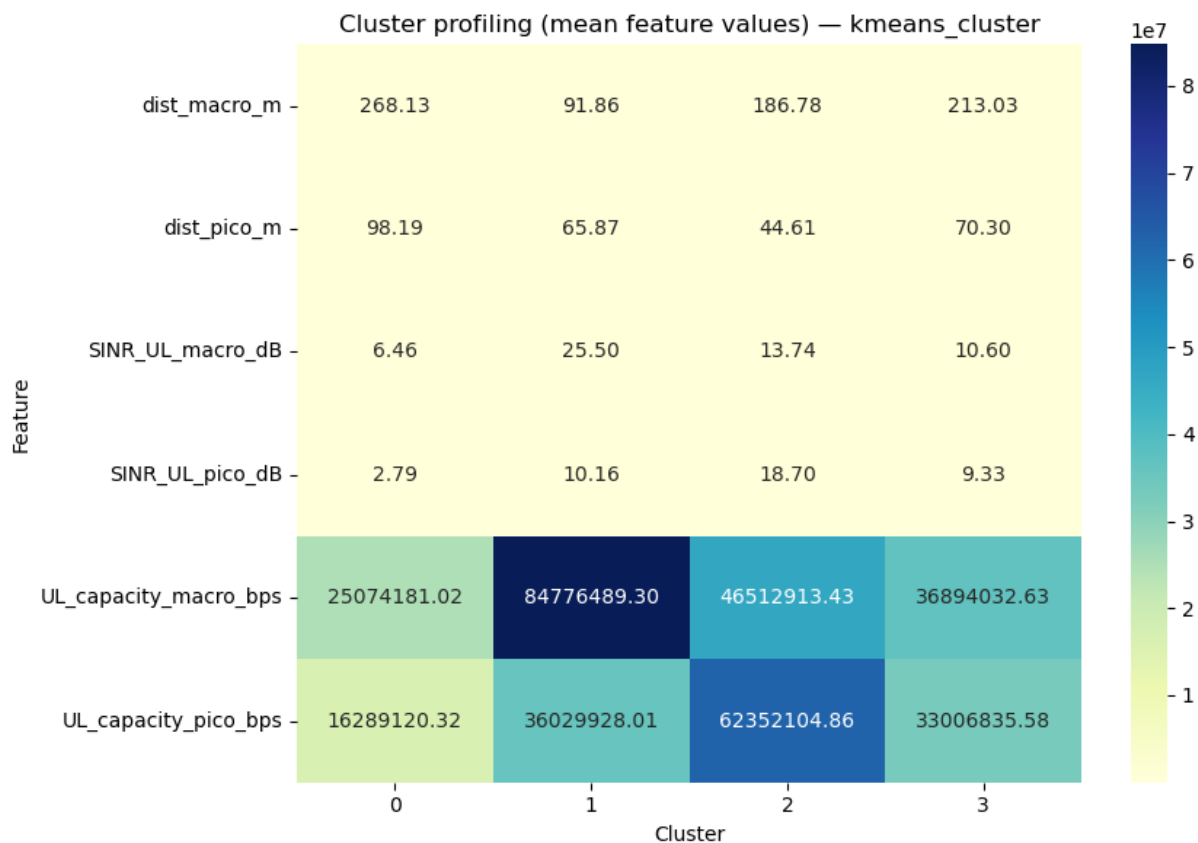


Εικόνα 11: Προβολή PCA των χρηστών με χρωματική ένδειξη των ομάδων K-Means

Όπως φαίνεται στην Εικόνα 11, οι χρήστες σχηματίζουν διακριτές ομάδες στον μειωμένο χώρο χαρακτηριστικών, γεγονός που υποδηλώνει την ύπαρξη εγγενούς δομής στα δεδομένα του uplink. Η παρατήρηση αυτή επιβεβαιώνει ότι η ομαδοποίηση μέσω μη εποπτευόμενης μάθησης μπορεί να αξιοποιήσει ουσιαστικές πληροφορίες του καναλιού.

Για την καλύτερη κατανόηση της λειτουργίας της ML-DUDe προσέγγισης, εξετάζεται η δομή των δεδομένων στο χώρο χαρακτηριστικών. Για τον σκοπό αυτό, εφαρμόστηκε προβολή των χρηστών σε μειωμένο διδιάστατο χώρο μέσω Ανάλυσης Κύριων Συνιστωσών (PCA). Κατά συνέπεια, είναι εφικτή η εφαρμογή πολιτικής συσχέτισης σε επίπεδο cluster, ώστε χρήστες με παρόμοιες uplink συνθήκες να ακολουθούν κοινό κανόνα επιλογής σταθμού βάσης

Για την ερμηνεία των παραγόμενων clusters, πραγματοποιήθηκε ανάλυση των μέσων τιμών των βασικών χαρακτηριστικών ανά ομάδα χρηστών.

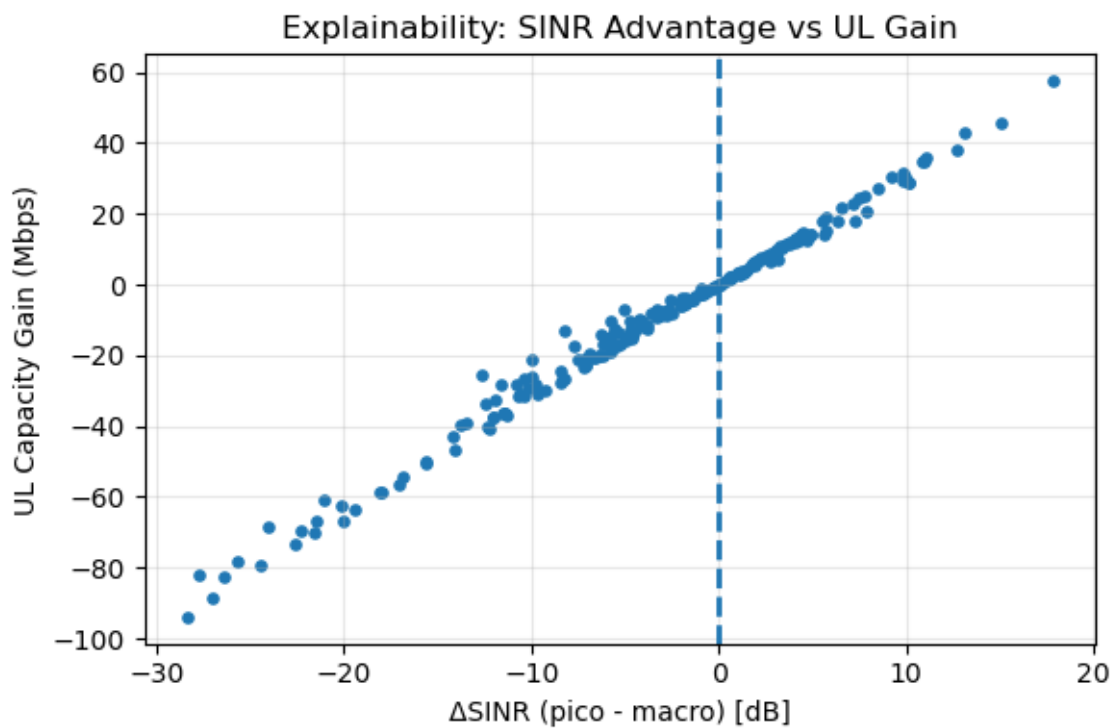


Εικόνα 12: Μέσες τιμές χαρακτηριστικών ανά ομάδα χρηστών (cluster profiling) για τον αλγόριθμο K-Means

Η Εικόνα 12 παρουσιάζει τις μέσες τιμές των χαρακτηριστικών ανά cluster, αποκαλύπτοντας σαφείς διαφοροποιήσεις μεταξύ των ομάδων χρηστών. Συγκεκριμένα, παρατηρούνται clusters με υψηλότερο uplink SINR και αυξημένη χωρητικότητα σε pico σταθμούς βάσης, καθώς και clusters που εξυπηρετούνται κυρίως από macro κυψέλες με χαμηλότερη uplink απόδοση. Η διαφοροποίηση αυτή δικαιολογεί τη χρήση πολιτικών συσχέτισης σε επίπεδο cluster και εξηγεί την ικανότητα της ML-DUDe προσέγγισης να προσεγγίζει αποτελεσματικά τη συμπεριφορά του DUDe.

Κατά συνέπεια, είναι εφικτή η εφαρμογή πολιτικής συσχέτισης σε επίπεδο cluster, ώστε χρήστες με παρόμοιες uplink συνθήκες να ακολουθούν κοινό κανόνα επιλογής σταθμού βάσης

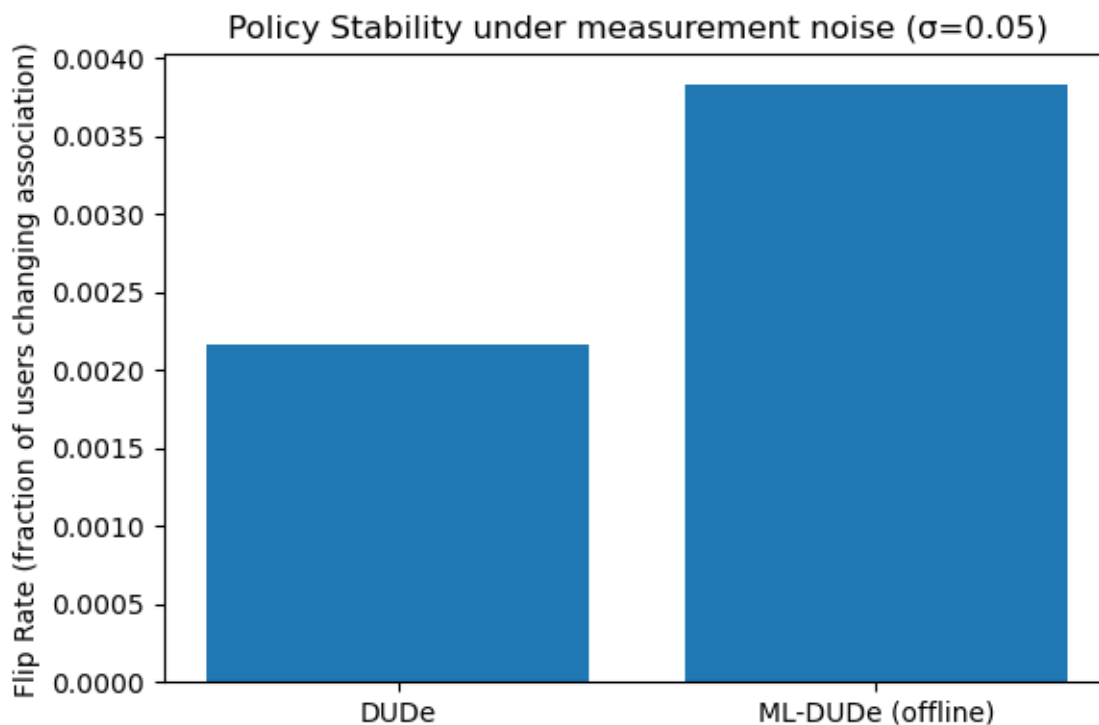
5.5 Ερμηνευσιμότητα των αποτελεσμάτων



Εικόνα 13: Σχέση πλεονεκτήματος SINR και κέρδους uplink χωρητικότητας

Η Εικόνα 13 απεικονίζει τη σχέση μεταξύ του πλεονεκτήματος SINR υπέρ των pico σταθμών βάσης και του αντίστοιχου κέρδους σε uplink χωρητικότητα. Παρατηρείται σαφής συσχέτιση μεταξύ των δύο μεγεθών, γεγονός που υποδεικνύει ότι η βελτίωση της απόδοσης δεν προκύπτει τυχαία, αλλά βασίζεται σε φυσικά χαρακτηριστικά του καναλιού. Η παρατήρηση αυτή ενισχύει την ερμηνευσιμότητα της ML-DUDe προσέγγισης και αποδεικνύει ότι το μοντέλο εκμεταλλεύεται ουσιαστικές πληροφορίες του uplink.

Για την αξιολόγηση της πρακτικής εφαρμοσιμότητας των πολιτικών συσχέτισης χρηστών, εξετάζεται η σταθερότητα των αποφάσεων uplink υπό μικρές μεταβολές των μετρήσεων. Συγκεκριμένα, μετράται το ποσοστό χρηστών που αλλάζουν συσχέτιση (flip rate) όταν προστίθεται τυχαίος θόρυβος στα uplink χαρακτηριστικά. Η σύγκριση πραγματοποιείται μεταξύ της DUDe πολιτικής και της ML-DUDe, με την τελευταία να εκπαιδεύεται offline και να εφαρμόζεται online.



Εικόνα 14: Σταθερότητα πολιτικής συσχέτισης (flip rate) για DUDe και ML-DUDe

Όπως παρατηρείται στην Εικόνα 14, παρουσιάζεται ο ρυθμός αλλαγής σύνδεσης χρηστών (flip rate) υπό την επίδραση μετρητικού θορύβου ($\sigma = 0.05$) για τις πολιτικές DUDe και ML-DUDe. Παρατηρείται ότι για την DUDe το μέσο flip rate ανέρχεται σε περίπου 0.002, ενώ για την ML-DUDe σε περίπου 0.004. Και στις δύο περιπτώσεις τα ποσοστά είναι εξαιρετικά χαμηλά, γεγονός που καταδεικνύει υψηλή σταθερότητα των πολιτικών σύνδεσης. Η ελαφρώς αυξημένη τιμή της ML-DUDe αποδίδεται στη χρήση ομαδοποιημένης πληροφορίας και δεδομενοκεντρικών αποφάσεων, χωρίς ωστόσο να επηρεάζει ουσιαστικά τη συνολική αξιοπιστία της πολιτικής.

5.6 Συνοπτική συζήτηση

Στο παρόν κεφάλαιο παρουσιάστηκαν και αναλύθηκαν τα αποτελέσματα της προσομοίωσης που αφορούν την απόδοση του uplink σε ετερογενή κυψελωτά δίκτυα, με έμφαση στη σύγκριση του βασικού σεναρίου συσχέτισης, της τεχνολογίας Decoupled Uplink and Downlink Access (DUDe) και της προτεινόμενης μη εποπτευόμενης προσέγγισης ML-DUDe. Τα αποτελέσματα καταδεικνύουν ότι η συμβατική συσχέτιση χρηστών βάσει του downlink οδηγεί σε

σημαντικούς περιορισμούς στο uplink, ιδιαίτερα λόγω της εξυπηρέτησης χρηστών από απομακρυσμένους macro σταθμούς βάσης και της επακόλουθης υποβάθμισης του λόγου σήματος προς παρεμβολή και θόρυβο.

Η εφαρμογή της τεχνολογίας DUDe επιβεβαιώνεται ως ιδιαίτερα αποτελεσματική λύση για την αντιμετώπιση των παραπάνω περιορισμών. Η ανεξάρτητη επιλογή σταθμού βάσης για το uplink οδηγεί σε σημαντική αύξηση της χωρητικότητας, βελτίωση των επιδόσεων των χρηστών χαμηλής απόδοσης και πιο ισορροπημένη κατανομή του φόρτου μεταξύ macro και pico σταθμών βάσης. Τα ευρήματα αυτά αναδεικνύουν το DUDe ως ένα άνω φράγμα απόδοσης για το uplink, το οποίο αξιοποιεί άμεσα τις ευνοϊκότερες συνθήκες καναλιού.

Παράλληλα, η προτεινόμενη ML-DUDe προσέγγιση, η οποία βασίζεται σε μη εποπτευόμενη μηχανική μάθηση και clustering χρηστών με παρόμοια uplink χαρακτηριστικά, καταφέρνει να προσεγγίσει σε μεγάλο βαθμό τη συμπεριφορά του DUDe. Τα αποτελέσματα δείχνουν ότι η ML-DUDe υπερέχει σαφώς του βασικού σεναρίου συσχέτισης, ενώ η απόκλιση από την απόδοση του DUDe παραμένει περιορισμένη. Το εύρημα αυτό είναι ιδιαίτερα σημαντικό, καθώς η ML-DUDe δεν αξιοποιεί ρητούς κανόνες συσχέτισης ούτε πλήρη πληροφορία καναλιού, αλλά βασίζεται αποκλειστικά στη δομή των διαθέσιμων δεδομένων uplink.

Η ανάλυση της δομής των δεδομένων μέσω προβολής σε μειωμένο χώρο χαρακτηριστικών ανέδειξε την ύπαρξη διακριτών ομάδων χρηστών με παρόμοιες uplink συνθήκες. Επιπλέον, η διαφοροποίηση των clusters ως προς βασικά μεγέθη, όπως το SINR και η uplink χωρητικότητα, προσδίδει ερμηνευσιμότητα στη μη εποπτευόμενη προσέγγιση και εξηγεί γιατί η ML-DUDe λαμβάνει συνεπείς και φυσικά αιτιολογημένες αποφάσεις συσχέτισης. Η παρατηρούμενη συσχέτιση μεταξύ πλεονεκτήματος SINR και κέρδους σε uplink χωρητικότητα επιβεβαιώνει ότι η βελτίωση της απόδοσης δεν είναι τυχαία, αλλά συνδέεται άμεσα με θεμελιώδη χαρακτηριστικά του καναλιού.

Συνολικά, τα αποτελέσματα της εργασίας δείχνουν ότι η μη εποπτευόμενη μηχανική μάθηση μπορεί να λειτουργήσει ως αποτελεσματικό εργαλείο υποστήριξης της τεχνολογίας DUDe, επιτρέποντας την προσέγγιση της συμπεριφοράς της με μειωμένες απαιτήσεις πληροφορίας και πολυπλοκότητας. Η ML-DUDe δεν στοχεύει στην υπέρβαση της θεωρητικής απόδοσης του DUDe, αλλά στη διευκόλυνση της πρακτικής υλοποίησής του μέσω δεδομενοκεντρικών και ερμηνεύσιμων πολιτικών συσχέτισης uplink. Τα συμπεράσματα αυτά συνοψίζονται και επεκτείνονται στο επόμενο κεφάλαιο, όπου παρουσιάζονται τα κύρια συμπεράσματα της εργασίας και προτείνονται κατευθύνσεις για μελλοντική έρευνα.

6

Συμπεράσματα & Μελλοντική Εργασία

6.1 Συμπεράσματα

Η παρούσα διπλωματική εργασία επικεντρώθηκε στη μελέτη της αποσυζευγμένης συσχέτισης uplink και downlink (Decoupled Uplink and Downlink Access – DUDe) σε ετερογενή κυψελωτά δίκτυα, καθώς και στη διερεύνηση της δυνατότητας προσέγγισης της συμπεριφοράς του DUDe μέσω τεχνικών μη εποπτευόμενης μηχανικής μάθησης. Το βασικό κίνητρο της εργασίας ήταν η αντιμετώπιση των περιορισμών του uplink, οι οποίοι προκύπτουν από τη συμβατική συσχέτιση χρηστών με βάση το downlink σήμα, ιδιαίτερα σε σενάρια υψηλής πυκνότητας χρηστών και ανομοιογενούς ανάπτυξης σταθμών βάσης, όπου η ανισοκατανομή φόρτου και οι αυξημένες παρεμβολές οδηγούν σε σημαντική υποβάθμιση της απόδοσης.

Στο πλαίσιο αυτό, υλοποιήθηκε ένα ρεαλιστικό σενάριο ετερογενούς κυψελωτού δικτύου που περιλαμβάνει macro και pico σταθμούς βάσης, με σκοπό τη μοντελοποίηση της συμπεριφοράς του uplink σε περιβάλλον πολλαπλών επιπέδων κάλυψης. Η προσομοίωση ενσωμάτωσε βασικά χαρακτηριστικά του ασύρματου καναλιού, όπως απώλειες διάδοσης, παρεμβολές και μεταβολές απόστασης, επιτρέποντας την αξιόπιστη αξιολόγηση των διαφορετικών πολιτικών σύνδεσης.

Τα αποτελέσματα επιβεβαίωσαν ότι η εφαρμογή της πολιτικής DUDe οδηγεί σε σημαντική βελτίωση της uplink απόδοσης, τόσο ως προς τη μέση χωρητικότητα όσο και ως προς τη συνολική κατανομή των επιδόσεων των χρηστών. Η ανεξάρτητη επιλογή σταθμού βάσης για το uplink επιτρέπει στους χρήστες να εκμεταλλεύονται ευνοϊκότερες συνθήκες καναλιού, μειώνοντας την εξάρτηση από απομακρυσμένες macro κυψέλες και περιορίζοντας τις επιπτώσεις της παρεμβολής. Παράλληλα, παρατηρήθηκε ουσιαστική ανακατανομή του φόρτου

προς τις pico κυψέλες, γεγονός που συμβάλλει στην αποσυμφόρηση των macro σταθμών βάσης και στη βελτιστοποίηση της χρήσης των διαθέσιμων ραδιοπόρων.

Ωστόσο, όπως αναδείχθηκε και στην ανάλυση, η πρακτική υλοποίηση της πολιτικής DUDe προϋποθέτει τη διαθεσιμότητα λεπτομερούς πληροφορίας uplink και τη λήψη αποφάσεων σε επίπεδο μεμονωμένου χρήστη σε πραγματικό χρόνο. Οι απαιτήσεις αυτές ενδέχεται να αυξήσουν την πολυπλοκότητα του συστήματος και το φορτίο σηματοδοσίας σε πραγματικά δίκτυα, καθιστώντας αναγκαία την αναζήτηση πιο απλών και αποδοτικών εναλλακτικών προσεγγίσεων.

Για τον λόγο αυτό, η εργασία διερεύνησε την αξιοποίηση μη εποπτευόμενης μηχανικής μάθησης ως δεδομενοκεντρικής προσέγγισης για την προσέγγιση της συμπεριφοράς του DUDe. Η προτεινόμενη πολιτική ML-DUDe βασίζεται στην ομαδοποίηση χρηστών με παρόμοια uplink χαρακτηριστικά, όπως η απόσταση από τους σταθμούς βάσης, οι τιμές SINR και η εκτιμώμενη χωρητικότητα προς κάθε τύπο κυψέλης. Μέσω της ανάλυσης αυτών των χαρακτηριστικών, το σύστημα είναι σε θέση να λαμβάνει αποφάσεις σύνδεσης που μεγιστοποιούν την αναμενόμενη απόδοση στο uplink.

Η αξιολόγηση των πειραματικών αποτελεσμάτων έδειξε ότι η μη εποπτευόμενη προσέγγιση μπορεί να αναπαράγει σε μεγάλο βαθμό τη συμπεριφορά του DUDe, επιτυγχάνοντας σαφώς ανώτερη απόδοση σε σχέση με τη συμβατική συζευγμένη πολιτική σύνδεσης και παρουσιάζοντας περιορισμένη απόκλιση από την ιδανική πολιτική DUDe. Η γενικευμένη βελτίωση της χωρητικότητας για το σύνολο των χρηστών, όπως αποτυπώθηκε τόσο στις κατανομές CDF όσο και στα boxplots, επιβεβαιώνει την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας.

Ιδιαίτερη σημασία έχει το γεγονός ότι η ανάλυση της δομής των δεδομένων και η ερμηνεία των παραγόμενων clusters ανέδειξαν ότι οι αποφάσεις της ML-DUDe βασίζονται σε ουσιαστικά φυσικά χαρακτηριστικά του uplink και όχι σε αυθαίρετες στατιστικές συσχετίσεις. Η ιδιότητα αυτή προσδίδει ερμηνευσιμότητα στην προτεινόμενη προσέγγιση και ενισχύει την αξιοπιστία της ως πρακτικό εργαλείο υποστήριξης αποφάσεων σε κυψελωτά δίκτυα.

Επιπλέον, η αξιολόγηση της σταθερότητας των αποφάσεων έδειξε ότι, όταν η μηχανική μάθηση εκπαιδεύεται offline και εφαρμόζεται online, η ML-DUDe παρουσιάζει συγκρίσιμη συμπεριφορά με το DUDe υπό μικρές μεταβολές των μετρήσεων καναλιού. Το εύρημα αυτό είναι ιδιαίτερα σημαντικό για την πρακτική εφαρμογή της μεθοδολογίας, καθώς υποδηλώνει

ότι η προσέγγιση μπορεί να λειτουργήσει αξιόπιστα σε δυναμικά περιβάλλοντα χωρίς συνεχή επανεκπαίδευση.

Συνολικά, η εργασία καταδεικνύει ότι η μη εποπτευόμενη μηχανική μάθηση μπορεί να αποτελέσει αποτελεσματική, ερμηνεύσιμη και πρακτικά υλοποιήσιμη λύση για τη βελτίωση της uplink απόδοσης σε ετερογενή κυψελωτά δίκτυα. Μέσω της δεδομενοκεντρικής προσέγγισης, καθίσταται δυνατή η γεφύρωση του χάσματος μεταξύ θεωρητικά βέλτιστων πολιτικών συσχέτισης και των ρεαλιστικών περιορισμών υλοποίησης των σύγχρονων ασύρματων δικτύων.

6.2 Μελλοντική Εργασία

Παρότι τα αποτελέσματα της παρούσας εργασίας είναι ιδιαίτερα ενθαρρυντικά, υπάρχουν αρκετές κατευθύνσεις για περαιτέρω έρευνα και επέκταση της προτεινόμενης προσέγγισης. Μια πρώτη και ουσιαστική κατεύθυνση αφορά τη μελέτη πιο δυναμικών σεναρίων δικτύου, όπου λαμβάνεται υπόψη η κινητικότητα των χρηστών και οι χρονικές μεταβολές των uplink συνθηκών. Σε τέτοια περιβάλλοντα, η στατική ομαδοποίηση χρηστών ενδέχεται να μην είναι επαρκής, καθιστώντας αναγκαία τη διερεύνηση μηχανισμών δυναμικής προσαρμογής των clusters, περιοδικής επανεκπαίδευσης του μοντέλου ή ακόμη και online μάθησης, ώστε το σύστημα να μπορεί να προσαρμόζεται σε πραγματικό χρόνο στις μεταβαλλόμενες συνθήκες του δικτύου.

Μια δεύτερη κατεύθυνση μελλοντικής έρευνας σχετίζεται με την ενσωμάτωση πρόσθετων παραμέτρων στη διαδικασία λήψης αποφάσεων, πέραν των βασικών uplink χαρακτηριστικών. Η εισαγωγή κριτηρίων ποιότητας υπηρεσίας (QoS), όπως απαιτήσεις καθυστέρησης, αξιοπιστίας, ρυθμού απώλειας πακέτων ή ενεργειακής κατανάλωσης των τερματικών χρηστών, θα μπορούσε να οδηγήσει σε μια πολυκριτηριακή πολιτική σύνδεσης. Μια τέτοια προσέγγιση θα επέτρεπε την προσαρμογή των αποφάσεων στις ανάγκες διαφορετικών εφαρμογών, όπως υπηρεσίες χαμηλής καθυστέρησης, μαζικές συνδέσεις IoT ή εφαρμογές υψηλής χωρητικότητας.

Επιπλέον, ενδιαφέρουσα κατεύθυνση αποτελεί η διερεύνηση υβριδικών προσεγγίσεων, όπου η μη εποπτευόμενη μάθηση συνδυάζεται με εποπτευόμενες ή ενισχυτικές τεχνικές μηχανικής μάθησης. Τέτοιες προσεγγίσεις θα μπορούσαν να αξιοποιούν ιστορικά δεδομένα, προβλέψεις

κίνησης χρηστών ή ανατροφοδότηση από το δίκτυο, βελτιώνοντας περαιτέρω την ακρίβεια και την προσαρμοστικότητα των αποφάσεων σύνδεσης. Για παράδειγμα, αλγόριθμοι ενισχυτικής μάθησης θα μπορούσαν να βελτιστοποιούν μακροπρόθεσμα την απόδοση του δικτύου λαμβάνοντας υπόψη τη χρονική εξέλιξη του φόρτου και των παρεμβολών.

Επιπρόσθετα, θα ήταν χρήσιμη η διερεύνηση της επίδρασης διαφορετικών μοντέλων διάδοσης, παρεμβολών και γεωμετρικών διατάξεων των σταθμών βάσης, ώστε να αξιολογηθεί η γενικευσιμότητα της προτεινόμενης ML-DUDe προσέγγισης σε ένα ευρύτερο φάσμα σεναρίων. Η ανάλυση σε περιβάλλοντα με διαφορετική πυκνότητα κυψελών, αστικές και αγροτικές περιοχές, καθώς και με ανομοιογενή κατανομή χρηστών θα προσέφερε πληρέστερη εικόνα της απόδοσης του συστήματος.

Τέλος, ιδιαίτερα σημαντική κατεύθυνση μελλοντικής έρευνας αποτελεί η αξιολόγηση της προτεινόμενης ML-DUDe προσέγγισης σε πιο ρεαλιστικά σενάρια δικτύων πέμπτης και έκτης γενιάς, καθώς και η μελέτη της ενσωμάτωσής της σε πραγματικές αρχιτεκτονικές δικτύων. Η ανάλυση ζητημάτων σηματοδότησης, υπολογιστικής πολυπλοκότητας, απαιτήσεων αποθήκευσης δεδομένων και κλιμάκωσης του συστήματος θα αποτελέσει κρίσιμο βήμα για τη μετάβαση από τη θεωρητική ανάλυση στην πρακτική εφαρμογή της προτεινόμενης μεθοδολογίας σε πραγματικές υποδομές κινητών επικοινωνιών.

Βιβλιογραφία

- [1]. A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.
- [2]. T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed., Prentice Hall, 2002.
- [3]. S. Singh, H. S. Dhillon, and J. G. Andrews, “Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights,” arXiv:1208.1977, 2012
- [4]. A. Damnjanovic et al., “A Survey on 3GPP Heterogeneous Networks,” *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, Jun. 2011.
- [5]. H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, “Modeling and Analysis of K-Tier Downlink Heterogeneous Cellular Networks,” *IEEE JSAC*, vol. 30, no. 3, pp. 550–560, Apr. 2012.
- [6]. H. S. Dhillon, R. K. Ganti, and J. G. Andrews, “Load-Aware Modeling and Analysis of Heterogeneous Cellular Networks,” *IEEE Trans. Wireless Communications*, 2013.
- [7]. H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, “Downlink and Uplink Decoupling: A Disruptive Architectural Design for 5G Networks,” in *Proc. IEEE GLOBECOM*, 2014.
- [8]. K. Smiljkovikj, H. Elshaer, P. Popovski, F. Boccardi, M. Dohler, L. Gavrilovska, and R. Irmer, “Capacity Analysis of Decoupled Downlink and Uplink Access in 5G Heterogeneous Systems,” arXiv:1410.7270, 2014.
- [9]. Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, “Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues,” *IEEE Communications Surveys & Tutorials*, 2019.
- [10]. 3GPP TS 36.213, “E-UTRA; Physical layer procedures.”
- [11]. D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge Univ. Press, 2005.
- [12]. C. E. Shannon, “A Mathematical Theory of Communication,” *Bell System Technical Journal*, 1948.
- [13]. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [14]. I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, 2002.
- [15]. J. MacQueen, “Some Methods for Classification and Analysis of Multivariate Observations,” in *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, 1967, pp. 281–297.
- [16]. J. H. Ward, Jr., “Hierarchical Grouping to Optimize an Objective Function,” *J. American Statistical Association*, vol. 58, no. 301, pp. 236–244
- [17]. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” in *Proc. KDD*, 1996, pp. 226–231.
- [18]. Chung, *Spectral Graph Theory*, AMS, 1997.
- [19]. A. Y. Ng, M. I. Jordan, and Y. Weiss, “On Spectral Clustering: Analysis and an Algorithm,” in *Advances in Neural Information Processing Systems*, 2002.
- [20]. T. Zhang, R. Ramakrishnan, and M. Livny, “BIRCH: An Efficient Data Clustering Method for Very Large Databases,” in *Proc. ACM SIGMOD*, 1996.
- [21]. M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “OPTICS: Ordering Points To Identify the Clustering Structure,” in *Proc. ACM SIGMOD*, 1999.
- [22]. F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *Proc. IEEE ICDM*, 2008, pp. 413–422.

- [23]. P. J. Rousseeuw, “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis,” *J. Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [24]. D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [25]. T. Calinski and J. Harabasz, “A Dendrite Method for Cluster Analysis”, 1974
- [26]. L. Pt and P. Arabie, “Comparing Partitions”, Adjusted Rand Index, 1985.

Παράρτημα: Κώδικας της Εργασίας

```
import os,warnings
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from scipy.cluster.hierarchy import linkage,dendrogram,cophenet,fcluster
from scipy.spatial.distance import pdist

from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import
KMeans,MiniBatchKMeans,SpectralClustering,AgglomerativeClustering,DBSCAN,OPTICS,Birch
from sklearn.mixture import GaussianMixture
from sklearn.ensemble import IsolationForest
from sklearn.metrics import silhouette_score,davies_bouldin_score,calinski_harabasz_score

warnings.filterwarnings("ignore")
np.random.seed(1)

# ===== PARAMETERS =====

N_USERS,R,h_bs,B=200,500.,25.,10e6
NOISE_DBM=-104.
P_UE_DBM,P_DL_MACRO_DBM,P_DL_PICO_DBM=23.,46.,30.
G_UE,G_MACRO,G_PICO=0.,14.,6.
N_EXP=3.7
PL0_MACRO,PL0_PICO=32.4,40.4
DL_CRE_BIAS_DB,UL_BIAS_DB=12.,0.
PICO_MAX_DIST_M=180.
N_PICO_BS=9

# ===== TOPOLOGY =====

x=np.random.uniform(0,R,N_USERS)
y=np.random.uniform(0,R,N_USERS)
xm,ym=R/2,R/2

grid=np.linspace(.2,.8,3)*R
xp,yp=np.meshgrid(grid,grid)
xp,yp=xp.flatten(),yp.flatten()

d_macro=np.sqrt((x-xm)**2+(y-ym)**2+h_bs**2)

d_pico_all=np.sqrt((x[:,None]-xp[None,:])**2+(y[:,None]-yp[None,:])**2+h_bs**2)
```

```

best=np.argmin(d_pico_all,1)
d_pico=d_pico_all[np.arange(N_USERS),best]

# ===== CHANNEL =====

PL_macro=PL0_MACRO+10*N_EXP*np.log10(d_macro)
PL_pico_all=PL0_PICO+10*N_EXP*np.log10(d_pico_all)
PL_pico=PL0_PICO+10*N_EXP*np.log10(d_pico)

PR_macro=P_UE_DBM+G_MACRO-PL_macro
PR_pico=P_UE_DBM+G_PICO-PL_pico
PR_pico_all=P_UE_DBM+G_PICO-PL_pico_all

I_macro=np.clip(np.random.normal(12,3,N_USERS),6,25)
I_pico=np.clip(np.random.normal(15,4,N_USERS),8,28)

SINR_macro=PR_macro-NOISE_DBM-I_macro
SINR_pico=PR_pico-NOISE_DBM-I_pico
SINR_pico_all=PR_pico_all-NOISE_DBM-I_pico[:,None]

C_macro=B*np.log2(1+10**(SINR_macro/10))
C_pico=B*np.log2(1+10**(SINR_pico/10))
C_pico_all=B*np.log2(1+10**(SINR_pico_all/10))

# ===== ASSOCIATION =====

DL_PR_macro=P_DL_MACRO_DBM+G_MACRO-PL_macro
DL_PR_pico_all=P_DL_PICO_DBM+G_PICO-PL_pico_all+DL_CRE_BIAS_DB
UL_baseline=(DL_PR_pico_all.max(1)>DL_PR_macro).astype(int)

UL_DUDe=((SINR_pico>SINR_macro)&(d_pico<=PICO_MAX_DIST_M)).astype(int)

# ===== DATAFRAME =====

df=pd.DataFrame({
"x":x,"y":y,"best_pico_id":best,
"dist_macro_m":d_macro,"dist_pico_m":d_pico,
"SINR_UL_macro_dB":SINR_macro,"SINR_UL_pico_dB":SINR_pico,
"UL_capacity_macro_bps":C_macro,"UL_capacity_pico_bps":C_pico,
"UL_baseline_macro0_pico1":UL_baseline,
"UL_DUDe_macro0_pico1":UL_DUDe
})

for p in range(N_PICO_BS):
df[f"d_pico_{p}_m"]=d_pico_all[:,p]
df[f"SINR_pico_{p}_dB"]=SINR_pico_all[:,p]
df[f"C_pico_{p}_bps"]=C_pico_all[:,p]

```

```
# ===== TOPOLOGY PLOT =====

plt.figure(figsize=(7,7))
plt.scatter(df.x,df.y,s=15,alpha=.6)
plt.scatter(xm,ym,c="r",s=250,marker="*")
plt.scatter(xp,yp,c="b",s=120,marker="^")
plt.axis("equal");plt.grid();plt.title("Topology");plt.show()

# ===== FEATURES =====

features=["dist_macro_m","dist_pico_m","SINR_UL_macro_dB","SINR_UL_pico_dB","UL_capacity_ma
cro_bps","UL_capacity_pico_bps"]

X=StandardScaler().fit_transform(SimpleImputer(strategy="median").fit_transform(df[features
]))

# ===== PCA =====

Xp=PCA(2,random_state=42).fit_transform(X)
df["PCA1"],df["PCA2"]=Xp[:,0],Xp[:,1]

plt.scatter(df.PCA1,df.PCA2,s=12)
plt.title("PCA");plt.show()

# ===== CLUSTERING =====

K=4

df["kmeans_cluster"]=KMeans(K,n_init=20,random_state=42).fit_predict(X)
df["mbkmeans_cluster"]=MiniBatchKMeans(K,n_init=20,random_state=42).fit_predict(X)
df["spectral_cluster"]=SpectralClustering(K,affinity="nearest_neighbors",assign_labels="kme
ans",random_state=42).fit_predict(X)
df["agg_cluster"]=AgglomerativeClustering(K).fit_predict(X)
df["gmm_cluster"]=GaussianMixture(K,random_state=42).fit(X).predict(X)
df["dbscan_cluster"]=DBSCAN(eps=1,min_samples=5).fit_predict(X)
df["optics_cluster"]=OPTICS(min_samples=10).fit_predict(X)
df["birch_cluster"]=Birch(n_clusters=K).fit_predict(X)

df["anomaly"] = pd.Series(
    IsolationForest(contamination=0.10, random_state=42).fit_predict(X)
).map({1:0, -1:1})

# ===== HEATMAP =====

profile=df.groupby("kmeans_cluster")[features].mean().T
sns.heatmap(profile,annot=True,fmt=".2f",cmap="YlGnBu")
plt.title("Cluster profiling");plt.show()

# ===== DENDROGRAM =====
```

```
Z=linkage(pdist(X),"ward")
dendrogram(Z,truncate_mode="level",p=5,no_labels=True)
plt.title("Hierarchical");plt.show()

# ===== METRICS =====

labels=df.kmeans_cluster
print("Silhouette:",silhouette_score(X,labels))
print("Davies:",davies_bouldin_score(X,labels))
print("Calinski:",calinski_harabasz_score(X,labels))

# ===== DUDe =====

df["UL_baseline"]="macro"
df["UL_DUDe"]=np.where(df.SINR_UL_pico_dB>df.SINR_UL_macro_dB,"pico","macro")

df["ul_capacity_macro"]=df.UL_capacity_macro_bps
df["ul_capacity_pico"]=df.UL_capacity_pico_bps

cap_base=df.ul_capacity_macro
cap_dude=np.where(df.UL_DUDe=="macro",df.ul_capacity_macro,df.ul_capacity_pico)

plt.boxplot([cap_base/1e6,cap_dude/1e6],labels=["Baseline","DUDe"],showfliers=False)
plt.ylabel("Mbps");plt.show()

# ===== CDF =====

def cdf(x,l):
    s=np.sort(x)
    plt.plot(s/1e6,np.arange(1,len(s)+1)/len(s),label=l)

cdf(cap_base,"Baseline")
cdf(cap_dude,"DUDe")
plt.legend();plt.show()

# ===== EXPLAIN =====

d=df.SINR_UL_pico_dB-df.SINR_UL_macro_dB
g=(df.ul_capacity_pico-df.ul_capacity_macro)/1e6

plt.scatter(d,g,s=14)
plt.axvline(0,ls="--")
plt.xlabel("ΔSINR");plt.ylabel("Gain Mbps")
plt.title("Explainability");plt.grid();plt.show()
```