

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης  
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Τηλεπικοινωνιών



Ασκητόπουλος Αθανάσιος, ΑΕΜ 8780

## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Full and Partial Offloading in a Mobile Edge Computing(MEC)  
scenario with Distributed Uplink Non Orthogonal Multiple  
Access(DU-NOMA)

Επιβλέπων:

Θεσσαλονίκη, Απρίλιος 2024

© Αλαμπάνος Αλέξανδρος

© Α.Π.Θ.

Full and Partial Offloading in a Mobile Edge Computing(MEC) scenario with Distributed Uplink Non Orthogonal Multiple Access(DU-NOMA)

«Η έγκριση της παρούσης Διδακτορικής Διατριβής από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέως»

(Ν. 5343/1932, άρθρο 202, παρ. 2)

# Ευχαριστίες

Η παρούσα διπλωματική εργασία σηματοδοτεί την ολοκλήρωση των προπτυχιακών σπουδών μου στο τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ. Κατ'άρχάς, θα ήθελα να εκφράσω την αμέριστη ευγνωμοσύνη στην οικογένεια μου, για την απόλυτη στήριξη και υπομονή που μου παρέχουν απλόχερα, όλα αυτά τα χρόνια. Επιπλέον, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κύριο Γιώργο Καραγιαννίδη, αλλά και τον μεταδιδακτορικό ερευνητή Παναγιώτη Διαμαντουλάκη, για την εμπιστοσύνη που μου έδειξαν, αναθέτοντάς μου ένα ενδιαφέρον και απαιτητικό θέμα, αλλά και τη πολύτιμη καθοδήγηση και συμβουλές τους, κατά την διάρκεια της εκπόνησης της εργασίας. Τέλος, ένα μεγάλο ευχαριστώ σε όλους τους φίλους μου για την συμπαράσταση, την κατανόηση, τα γέλια και τις στιγμές που μοιραστήσαμε.

Ασκητόπουλος Αθανάσιος



# Περίληψη

Τα συστήματά ασύρματης επικοινωνίας αποτελούν αναπόσπαστο κομμάτι των τεχνολογικών δομών που συγκροτούν την μοντέρνα ζωή. Οι ρυθμοί της τεχνολογικής ανάπτυξης αυτών των δομών προβάλλουν μια σειρά νέων απαιτήσεων και προκλήσεων στο σχεδιασμό τους ,όπως η μαζική αύξηση των αριθμού και της ποικιλίας των συσκευών που εξυπηρετούνται από το δίκτυο τηλεπικοινωνιών καθώς και οι νέες εφαρμογές που προτάσσουν πολύ πιο αυστηρές απαιτήσεις ρυθμού μετάδοσης ,καθυστερήσης , φασματικής αποδοτικότητας και ποιότητας εξυπηρέτησης. Η ανάγκη για αξιόπιστη , ενεργειακά αποδοτική επικοινωνία σε κλίμακα πολύ μεγαλύτερη από τις προηγούμενες δεκαετίες καλύπτεται από την εμφάνιση καινοτομιών στο επίπεδο της αρχιτεκτονικής του δικτύου και της θέσπισης νέων πρωτοκόλλων στα πλαίσια της 5ης και 6ης γενιάς ασύρματων δικτύων.

Μία από τις υποσχόμενες τεχνολογίες για την επίτευξη αξιόπιστης επικοινωνίας μεταξύ χρηστών αποτελεί η μη ορθογώνια πολλαπλή πρόσβαση(Non Orthogonal Multiple Access), η οποία επιτρέπει την από κοινού πρόσβαση σε ένα ορθογώνιο μπλοκ πόρων, από ένα σύνολο χρηστών, με αποτέλεσμα την αύξηση της συνδεσιμότητας και αποδοτικότητας διαχείρισης των πόρων ενός συστήματος. Η τεχνική αυτή, σε συνδυασμό με την αναδυόμενη τεχνολογία του Mobile Edge Computing, η οποία αποτελεί εξέλιξη της Cloud Computing αρχιτεκτονικής και παρέχει τη δυνατότητα εκφόρτωσης υπολογιστικού φόρτου τοπικών συσκευών σε διακομιστές άκρης (edge servers) κοντά στους χρήστες, μπορεί να αποτελέσει μια σημαντική προσέγγιση για την κάλυψη των αναγκών καθυστέρησης και ενεργειακής αποδοτικότητας επερχόμενων εφαρμογών, όπως το Internet of Things(IoT).

Στην παρούσα διπλωματική διατριβή, μελετάται ένα σενάριο διαχείρισης πόρων ενός διανεμημένου συστήματος NOMA για τις μεταδόσεις άνω ζεύξης(Distributed uplink NOMA(ΔΥ-NOM)), στο οποίο διανεμημένα ραδιοστελέχη (Remote Radio Heads(RRH)), εξοπλισμένα με edge servers, υλοποιούν ένα Mobile Edge Computing περιβάλλον για την εκφόρτωση εργασιών των τοπικών χρηστών, μέσω της μη ορθογωνικής πρόσβασης. Περιγράφεται το μοντέλο και η λειτουργία του MEC με εφαρμογή DU-NOMA, διατυπώνονται προβλήματα ελαχιστοποίησης της καθυστέρησης και ενεργειακής κατανάλωσης στην περίπτωση πλήρους εκφόρτωσης(full offloading) και ελαχιστοποίησης της καθυστέρησης εκφόρτωσης και τοπικής επεξεργασίας, στην περίπτωση της μερικής εκφόρτωσης(partial offloading) και τέλος, επιλύονται μέσω μεθόδων βελτιστοποίησης σε περιβάλλον MatLab. Τα αποτελέσματα των προσομοιώσεων δίνονται για διαφορετικές αποστάσεις χρηστών από τα ραδιοστελέχη, δια-

θέσιμης ισχύος και χρόνου μέγιστης μετάδοσης, παρέχοντας χρήσιμες πληροφορίες για την απόδοση του συστήματος.

**Λέξεις κλειδιά:** τηλεπικοινωνίες, ασύρματα δίκτυα 5<sup>ης</sup> γενιάς και μελλοντικές γενιές, Centralized/Cloud - Radio Access Network (C-RAN), Distributed Uplink - Non Orthogonal Multiple Access (DU-NOMA), Mobile Edge Computing (MEC), task offloading, Full Offloading, Partial Offloading

# Abstract

Wireless communication systems constitute an integral part of the technological infrastructures shaping modern life. The rates of technological development in these structures present a series of new requirements and challenges in their design, such as the massive increase in the number and variety of devices served by the telecommunications network, as well as the new applications that impose much stricter requirements on transmission rates, latency, spectral efficiency, and quality of service. The need for reliable, energy-efficient communication on a much larger scale than in previous decades is addressed by the emergence of innovations in network architecture and the introduction of new protocols within the framework of the 5th and 6th generations of wireless networks.

One of the promising technologies for achieving reliable communication among users is Non-Orthogonal Multiple Access (NOMA), which allows joint access to an orthogonal block of resources by a set of users, resulting in increased connectivity and resource management efficiency of a system. This technique, combined with the emerging technology of Mobile Edge Computing (MEC), which is an evolution of Cloud Computing architecture and provides the ability to offload computational load from local devices to edge servers near users, can be a significant approach to address the latency and energy efficiency needs of upcoming applications, such as the Internet of Things (IoT).

In this thesis, a resource management scenario of a distributed NOMA system for uplink transmissions (Distributed uplink NOMA - DU-NOM) is studied. In this scenario, distributed Remote Radio Heads (RRHs), equipped with edge servers, implement a Mobile Edge Computing environment for offloading tasks of local users through non-orthogonal access. The model and operation of MEC with DU-NOMA are described, problems are formulated for minimizing delay and energy consumption in the case of full offloading, and minimizing offloading delay and local processing in the case of partial offloading. These problems are then solved using optimization methods in a Matlab environment. Simulation results are provided for different user distances from the RRHs, available power, and maximum transmission time, providing useful information on system performance.

**Keywords:** telecommunications, 5G and wireless networks beyond 5G, Centralized / Cloud - Radio Access Network (C-RAN), Distributed Uplink - Non Orthogonal Multiple Access (DU-NOMA), Mobile Edge Computing (MEC), task offloading, Full Offloading, Partial Offloading





# Περιεχόμενα

Ευχαριστίες	i
Περίληψη	ii
Περιεχόμενα	iii
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Ασύρματα δίκτυα 5 <sup>ης</sup> γενιάς και μελλοντικές γενιές . . . . .	1
1.1.1 Non-Orthogonal Multiple Access (NOMA) . . . . .	2
1.1.2 Cloud - Radio Access Network (C-RAN) . . . . .	3
1.2 Mobile Edge Computing (MEC) . . . . .	6
1.3 Τεχνολογίες αιχμής - σχετικές έρευνες . . . . .	8
1.4 Συνεισφορά . . . . .	10
1.5 Δομή εργασίας . . . . .	11
<b>2 Παρουσίαση της προτεινόμενης αρχιτεκτονικής - διατύπωση του προβλήματος</b>	<b>13</b>
2.1 Μοντέλο και διάταξη συστήματος . . . . .	13
2.2 Λειτουργία συστήματος - Rates χρηστών . . . . .	14
2.3 Περίπτωση Full Offloading . . . . .	16
2.4 Περίπτωση Partial Offloading . . . . .	16
2.5 Πρόβλημα βελτιστοποίησης: Full offloading . . . . .	17
2.5.1 Energy constrained delay minimization . . . . .	18
2.5.2 Delay constrained energy minimization . . . . .	18
2.6 Πρόβλημα βελτιστοποίησης: Partial offloading . . . . .	19
<b>3 Επίλυση προβλημάτων</b>	<b>23</b>
3.1 Προτεινόμενες Λύσεις Full Offloading . . . . .	23
3.1.1 Επίλυση Energy constrained minimization . . . . .	23
3.1.2 Επίλυση Delay constrained minimization . . . . .	29
3.2 Προτεινόμενη Λύση Partial Offloading . . . . .	30

<b>4</b>	<b>Αποτελέσματα Προσομοιώσεων</b>	<b>35</b>
4.1	Περιγραφή benchmark . . . . .	35
4.2	Προσομοιώσεις delay minimization . . . . .	36
4.3	Προσομοιώσεις energy minimization . . . . .	44
<b>5</b>	<b>Συμπεράσματα - Μελλοντική Έρευνα</b>	<b>49</b>
5.1	Επίλογος - Συμπεράσματα . . . . .	49
5.2	Πρόταση για Μελλοντική Έρευνα . . . . .	49
	<b>Βιβλιογραφία</b>	<b>51</b>

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Ασύρματα δίκτυα 5<sup>ης</sup> γενιάς και μελλοντικές γενιές

Με την εμπορική ανάπτυξη και διάθεση του δικτύων 5<sup>ης</sup> γενιάς, εν εξελίξει σε παγκόσμιο επίπεδο, παρατηρείται έντονο ενδιαφέρον τόσο από την ακαδημαϊκή κοινότητα όσο και από την βιομηχανία για τον σχεδιασμό των δικτύων επερχόμενων γενεών (5G and beyond και 6G) καθώς και τον καθορισμό των προδιαγραφών, των πεδίων εφαρμογής και των κατάλληλων τεχνικών για την κάλυψη των αυξανόμενων αναγκών. Κάθε καινούργια γενιά μέχρι στιγμής, χαρακτηριζόταν από μια επαύξηση των δυνατοτήτων επικοινωνίας με την υποστήριξη ομιλίας (1G και 2G), γραπτών μηνυμάτων (3G) και σύνδεσης με το διαδίκτυο (4G+) καθώς και με την συνοδευόμενη αύξησή των απαιτήσεων σε αριθμό χρηστών, ικανότητα κάλυψης, ρυθμών επικοινωνίας και φασματικής αποδοτικότητας. Η μετάβαση στην 5<sup>η</sup> και 6<sup>η</sup> γενιά δεν απαιτεί μονάχα μεγέθυνσή των παραπάνω απαιτήσεων σε εκθετικό βαθμό αλλά και προσθήκη ριζικά διαφορετικών λειτουργιών στο ασύρματο δίκτυο. Μια αρχική εκτίμηση της μεγέθυνσης σε σχέση με το 4G, προβλέπει την ικανότητα υποστήριξης 100 φορές υψηλότερο όγκο δεδομένων κινητής τηλεφωνίας ανά περιοχή, 10 με 100 φορές ταχύτερους ρυθμούς δεδομένων, μείωση της καθυστέρησης επικοινωνίας 5 φορές και υποστήριξη έως και 100 φορές μεγαλύτερο αριθμό συσκευών χωρίς επιβάρυνση σε κόστος ή κατανάλωση ενέργειας [1], [2]. Επιπροσθέτως, σε παγκόσμιο επίπεδο η μέση μηνιαία χρήση δεδομένων προβλέπεται να ανέλθει από 21 GB το 2023 έως και 56 GB το 2029, με το συνολικό τηλεπικοινωνιακό φόρτο να τριπλασιάζεται στο ίδιο διάστημα και το μερίδιο του 5G να φτάνει το 76% .

Η διαχείριση του υψηλού αυτού όγκου τηλεπικοινωνιακής κίνησης καθώς και οι επερχόμενες ανάγκες σε μαζική και αξιόπιστη επικοινωνία ετερογενών συσκευών, οδήγησε στη θέσπιση τριών βασικών κατηγοριών υπηρεσιών του 5G : της enhanced Mobile Broadband (eMBB), των Ultra-Reliable Low-Latency Communications (URLLC) και των massive Machine Type Communications (mMTC). Αναφορικά, η enhanced Mobile Broadband (eMBB) υποστηρίζει σταθερές συνδέσεις με υψηλό ρυθμό δεδομένων για σενάρια επικοινωνίας μεγάλης

κάλυψης και κινητικότητας και μπορεί να ειπωθεί ως άμεση επέκταση των υπάρχοντων δικτύων, οι Ultra-Reliable Low-Latency Communications (URLLC) υποστηρίζουν υψηλής αξιοπιστίας και πολύ χαμηλής καθυστέρησης, συνδέσεις από ένα περιορισμένο σύνολο ενεργών τερματικών με σκοπό την ανάπτυξη αυτοματοποιήσεων σε κλάδους της βιομηχανίας και της υγείας και τέλος, οι mMTC υποστηρίζουν βραχείας διαρκείας, χαμηλής ισχύος συνδέσεις από μεγάλο αριθμό, περιστασιακά ενεργών συσκευών σε εφαρμογές δικτύων IoT. Οι παραπάνω κατηγορίες συνδεσιμότητας καθιστούν εμφανές πως το δίκτυο 5G σχεδιάζεται ως μια 'πλατφόρμα' τεχνολογιών που θα επιτρέπει την κάλυψη ετερογενών προδιαγραφών σε ένα φάσμα πιθανών περιπτώσεων χρήσης.[3],[4],[5].

Οι δυνατότητες που θα παρέχει το σύνολο αυτών των τεχνολογιών μπορούν να συνοψιστεί ως εξής: η μαζική συνδεσιμότητα, η αυξημένη φασματική αποδοτικότητα, η μειωμένη απο-άκρο-σε-άκρο καθυστέρηση, οι αυξημένοι ρυθμοί μετάδοσης και χαμηλή κατανάλωση ενέργειας.

Παρότι η μετάβαση στην 5η γενιά δεν έχει ολοκληρωθεί, ο οραματισμός γύρω από τα χαρακτηριστικά και τις πιθανές εφαρμογές του δικτύου 6G συντελείται ταχύτατα στους κόλπους της ακαδημαϊκής κοινότητας και της βιομηχανίας. Με βάση την ήδη υπάρχουσα δυνατότητες του 5G, το 6G θα συνεχίσει την προσπάθεια πραγμάτωσης του internet of Everything (IoE), επεκτείνοντας τις human-to-machine (ως eMBB) και machine-to-machine type communications (ως mMTC και URLLC), και ενσωματώνοντας σε βάθος την Τεχνητή Νοημοσύνη. Από πλευράς οράματος, το 6G προβλέπεται να έχει την μορφή ενός κατανεμημένου, πανταχού παρόντος, 'έξυπνου' δικτύου που θα παρέχει συνδέσμους επικοινωνίας μεταξύ του φυσικού και του ψηφιακού κόσμου, μεταξύ ανθρώπων, πραγμάτων και μηχανών[6]. Η οπτική αυτή στηρίζεται στην χρησιμοποίηση ανεμετάλλευτων περιοχών του φάσματος όπως τις sub-6GHz, THz και οπτικές μπάντες στην συγχώνευση δυνατοτήτων υπολογισμού, επικοινωνίας, ανίχνευσης, ελέγχου και 'έξυπνης' λήψης αποφάσεων και στην παροχή εμπειριστικών, αισθητηριακών εμπειριών στους χρήστες μέσω ολογραφικής (holographic) επικοινωνίας και εφαρμογών επαυξημένης πραγματικότητας (XR)[7],[8].

### 1.1.1 Non-Orthogonal Multiple Access (NOMA)

Η προβλεπόμενη εκθετική αύξηση του αριθμού των συνδέσεων και συσκευών έως το 2030[7] καθώς και οι αυστηρές προδιαγραφές για την επίτευξη των διαφόρων σεναρίων χρήσης των δικτύων επόμενης γενιάς[9], καθιστούν επιτακτική την χρήση προηγμένων τεχνικών πολλαπλής προσπέλασης (multiple access techniques). Η μη-ορθογώνια πολλαπλή πρόσβαση αποτελεί σημαντική εξέλιξη στο χώρο του 5G και προβλέπεται να αποτελέσει παραδειγματική μέθοδος για τον σχεδιασμό τεχνικών προσπέλασης ραδιοφάσματος για τις μελλοντικές γενιές δικτύων (Next generation Multiple access)[10],[11],[12].

Σε αντίθεση με τις παραδοσιακές τεχνικές προσπέλασης όπως το TDMA, FDMA, CDMA και OFDM οι οποίες χρησιμοποιούν ορθογωνικά, μη επικαλυπτόμενα μπλοκ πόρων (χρόνου,

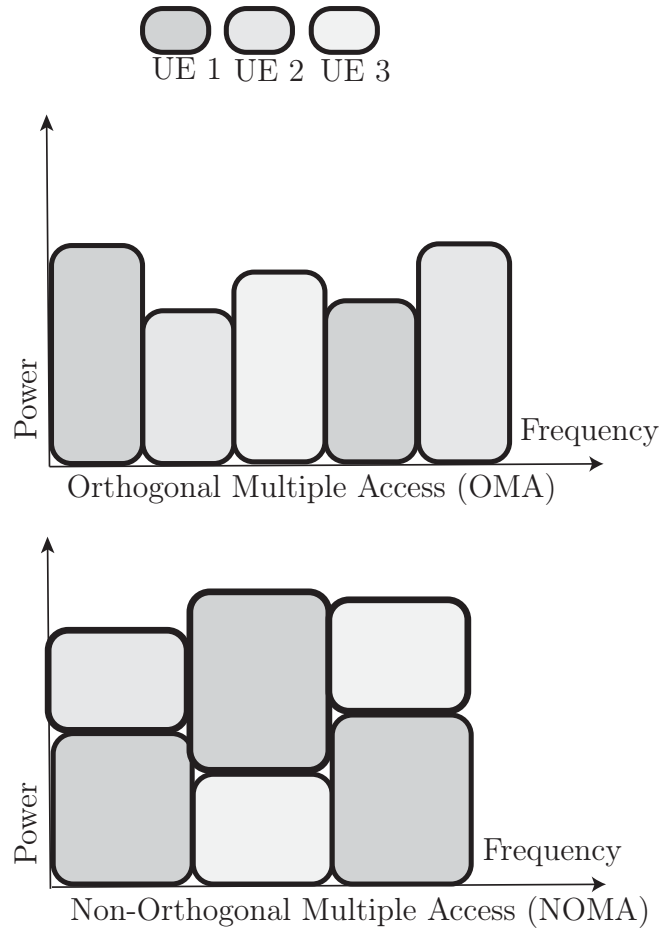
συχνότητας, κώδικα), περιορίζοντας τον αριθμό των εξυπηρετούμενων χρηστών στον αριθμό των εκάστοτε πόρων, στα συστήματα NOMA οι χρήστες μπορούν ταυτοχρόνως να χρησιμοποιούν το ίδιο μπλοκ για την πρόσβαση τους στο δίκτυο. Η υπερφόρτωση (Overloading) των υπάρχοντων πόρων επιτυγχάνεται με κάποια μορφή πολυπλεξίας στον πομπό και αποπολυπλεξίας σε μερικούς ή όλους τους δέκτες, με τελικό αποτέλεσμα την καλύτερη αξιοποίηση των τηλεπικοινωνιακών πόρων και την αύξηση της αποδοτικότητας.

Στο πρωταρχική του εκδοχή[12], κατά την downlink επικοινωνία μεταξύ ενός σταθμού βάσης(BS) και ενός συνόλου χρηστών, ο σταθμός βάσης χρησιμοποιεί superposition coding (SC) για την δημιουργία ενός γραμμικού συνδυασμού των μηνυμάτων των χρηστών, σταθμισμένο με ένα ποσοστό της εκπεμπόμενης ισχύος  $P$  για κάθε χρήστη και το εκπέμπει στο κανάλι. Το ποσοστό της ισχύος καθορίζεται από το κέρδος που δέχεται κάθε χρήστης στο κανάλι του,ορίζοντας ισχυρούς και αδύναμους χρήστες αντίστοιχα. Στην πλευρά των δεκτών, ο κάθε χρήστης αποκωδικοποιεί το μήνυμα του χρησιμοποιώντας successive interference cancelling (SIC). Τα σήματα των πιο αδύναμων χρηστών σε σχέση με το δέκτη, αποκωδικοποιούνται προοδευτικά και αφαιρούνται από το λαμβανόμενο σήμα ώστε να τελεστεί η αποκωδικοποίηση του μηνύματος, χρησιμοποιώντας τα σήματα των ισχυρότερων χρηστών ως παρεμβολή[9]. Είναι σημαντικό να σημειωθεί πως ο σταθμός βάσης αναθέτει μεγαλύτερη ισχύ στους αδύναμους, με βάση το συντελεστή καναλιού, χρήστες και λιγότερη στους ισχυρούς ώστε να μπορεί να συντελεστεί επιτυχημένα το SIC, με την αφαίρεση του μεγαλύτερου σήματος πρώτα, στους δέκτες.

Σε ένα σύστημα που χρησιμοποιεί OMA, κάθε χρήστης θα καταλάμβανε έναν πόρο, ακόμα και εκείνοι που διαθέτουν πολύ κακές συνθήκες καναλιού(σηαννελ στατε ινφορματιον), όπου στη περίπτωση ανάθεσης φάσματος, θα οδηγούσε σε μειωμένη φασματική αποδοτικότητα και ρυθμαπόδοση του συστήματος. Με τη χρήση NOMA,ισχυροί και αδύναμοι χρήστες μπορούν με το ίδιο φάσμα να εξυπηρετηθούν επαρκώς επιτυγχάνοντας καλύτερη δικαιοσύνη μεταξύ χρηστών (user fairness) και βελτιώσεις στη ρυθμαπόδοση.

### 1.1.2 Cloud - Radio Access Network (C-RAN)

Πέρα από τις τεχνικές πολλαπλής πρόσβασης, οι αυστηρές απαιτήσεις των δικτύων επόμενης γενιάς σε ικανότητα κάλυψης μεγάλης πυκνότητα χρηστών και όγκου τηλεπικοινωνιακής κίνησης αναμένεται να καλυφθούν από την ριζική αναδόμηση του δικτύου ραδιοκάλυψης (radio access network,RAN). Η τάση για επανακαθορισμό της αρχιτεκτονικής του δικτύου ραδιοπρόσβασης απαντάται ήδη από την μετάβαση από του δικτύου από το 2G RAN, με την κοινή επεξεργασία σημάτων βασικής ζώνης και ραδιοεκπομών σε Σταθμούς Βάσης (Base Station,BS), στις ποικίλες αρχιτεκτονικές των δικτύων 3G, 4G βασισμένες στο D-RAN, όπου η επεξεργασία αυτή τελείται σε ξεχωριστούς κόμβους: τα Remote Radio Heads(RRH) και τα Baseband Units (BBU).Οι αναγκαίότητα κάλυψης ετερογενών απαιτήσεων σε ποιότητα υπηρεσιών(Quality-of-Service) μαζί με την ραγδαία αύξηση σε δεδομένα χρηστών οδήγησε τους

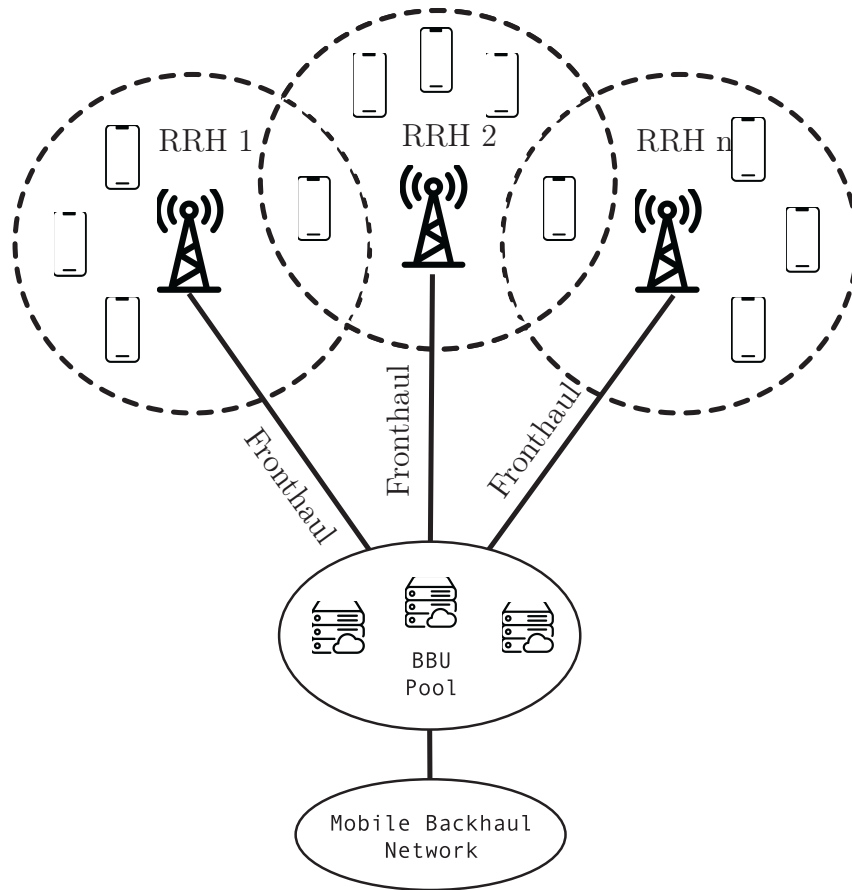


**Σχήμα 1.1:** Εικονική αναπαράσταση OMA και NOMA

παρόχους τηλεπικοινωνιακών υπηρεσιών (network operators στην προσπάθεια υιοθέτησης τεχνικών κεντροποίησης νεφοποίησης(cloudification) και εικονικοποίησης (virtualization των BBUs και των αντίστοιχων RRH.

Οι προσπάθειες αυτές οδήγησαν στην ανάδυση του Centralized Radio Access Network, μια καινοτόμα αρχιτεκτονική λύση η οποία προτάθηκε αρχικά από την IBM με την ονομασία Wireless Network Cloud[13] και στη συνέχεια αναλύθηκε πιο διεξοδικά σε ένα white paper του Chinese Mobile Research Institute[14]. Όπως παρουσιάζεται στο , στο βασικό σενάριο της εφαρμογής της αρχιτεκτονικής, τα BBUs διαχωρίζονται από τα RRHs τους και συγκεντρώνονται σε ένα κεντρικό ,εικονικοποιημένο pool βασισμένο στο υπολογιστικό νέφος. Κάθε ένα από τα RRH συνδέεται με το BBU pool μέσω Fronthaul ζεύξεων ενώ το pool, το οποίο μπορεί να στηρίζει πολλαπλά RRHs, συνδέεται με τον πυρήνα του δικτύου μέσω ζεύξης Backhaul. Τα RRH αποδεσμεύονται από ένα αποκλειστικό BBU και εξυπηρετούνται πλέον από virtualized BBUs τα οποία δημιουργούνται ανάλογα με τις ανάγκες εξυπηρέτησης. Με τη χρήση οπτικών ινών για τις ζεύξεις αυτές, εξασφαλίζεται μεγάλη χωρητικότητα και μειωμένη από-άκρο-σε-άκρο χωρητικότητα(latency) για την μετάδοση των δεδομένων μεταξύ του πυρήνα και των ραδιοσταθμών[15],[16].

Η αρχιτεκτονική C-RAN παρέχει ποικίλλα πλεονεκτήματα στα δίκτυα 5G όπως: την δυ-



**Σχήμα 1.2:** Εικονική αναπαράσταση εφαρμογής της αρχιτεκτονικής CRAN

νατότητα επέκτασης της κάλυψης του δικτύου ή της διάσπασης των υπάρχοντων κυψελών για την αναβάθμιση της χωρητικότητας, λόγω της ευκολίας προσθήκης νέων RRH συνδεδεμένων με το BBU pool, την υποστήριξη δια-κυψελιδικής(intra-cell) συνεργασίας για την διαχείριση των πόρων του δικτύου και την επεξεργασία σήματος, την βελτίωση της φασματικής αποδοτικότητας, της ευκολίας διαχείρισης της κινητικότητας και ισορρόπησης φόρου και της μείωσης της ενεργειακής κατανάλωσης του δικτύου[17],[18].

Η ενεργειακή αποδοτικότητα του C-RAN έχει σημειωθεί ως καίριο χαρακτηριστικό για την υποστήριξη των Πράσινων επικοινωνιών(Green Communications[19]) οι οποίες απαιτούν την προσεκτική επέκταση των δικτύων για τη στήριξη μεγαλύτερου αριθμού συσκευών, με γνώμονα την εξοικονόμηση ενέργειας χωρίς απώλειες στην ποιότητα επικοινωνίας. Το μεγαλύτερο μερίδιο της χρησιμοποιούμενης ενέργειας των δικτύων προηγούμενων γενεών καταναλωνόταν στους σταθμούς βάσης για την επίτευξη των διεργασιών ραδιοπρόσβασης (RAN processes) καθώς και για τη λειτουργία και ψύξη του εξοπλισμού. Οι τεχνικές υπολογιστικού νέφους σε συνδυασμό με την κεντροποιημένη επεξεργασία των ραδιοπομπών στην C-RAN αρχιτεκτονική, επιτρέπει τη μείωση του αριθμού BS χωρίς την συνοδευόμενη πτώση σε απόδοση και χωρητικότητα του δικτύου, καθώς τα virtualized BBUs μπορούν να καλύπτουν την τηλεπικοινωνιακή κίνηση δυναμικά και ανάλογα με τις συνθήκες ζήτησης των χρηστών[17].

Τέλος, η κεντρική διαχείριση των πόρων και η χρήση feedback link για την επικοινωνία των RRHs επιτρέπει την χρήση τεχνικών συνεργατικής επικοινωνίας (CoMP)[15], τεχνικών μη ορθογώνιας πολλαπλής πρόσβασης (NOMA) και διανεμημένης βελτιστοποίησης των πόρων για ακόμη μεγαλύτερη εξοικονόμηση ενέργειας[16].

## 1.2 Mobile Edge Computing (MEC)

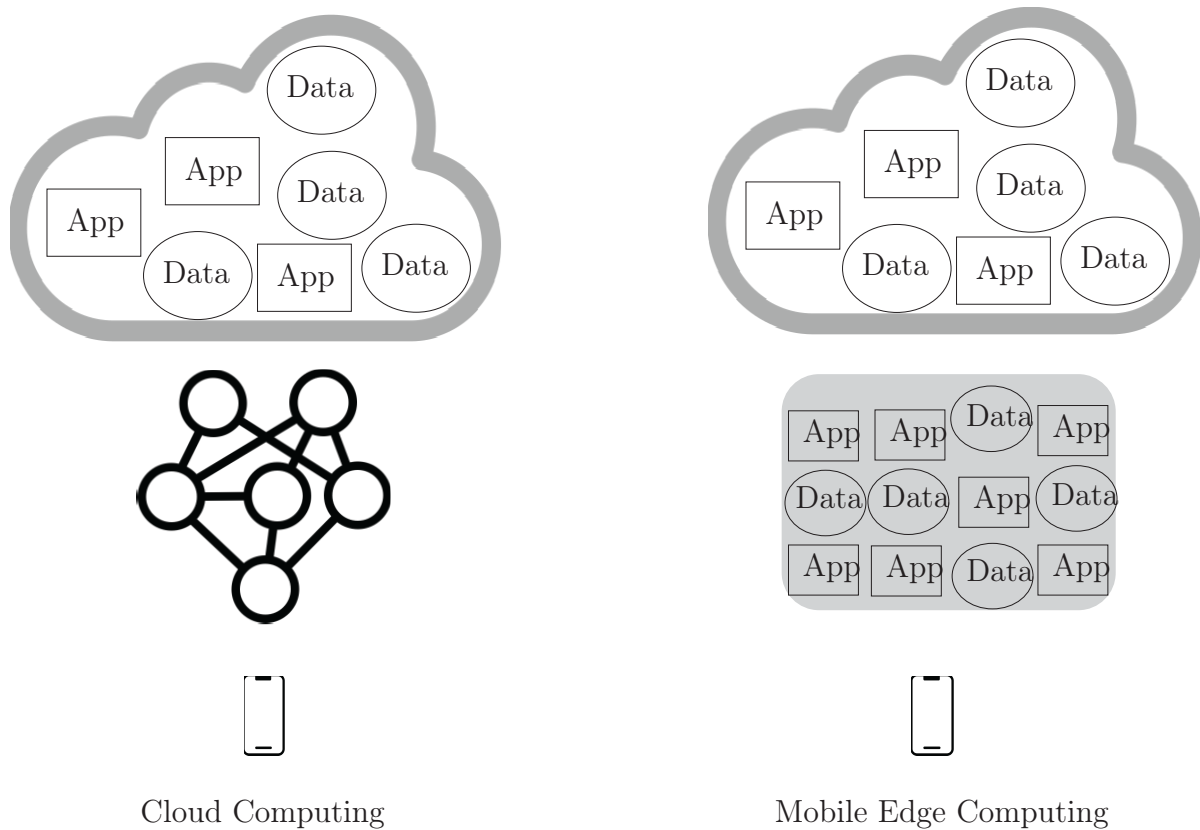
Βασικός στόχος των μελλοντικών γενεών δικτύων πέρα από το 5G αποτελεί η αξιόπιστη εξυπηρέτηση τεράστιου αριθμού συσκευών οι οποίες τρέχουν εφαρμογές με ετερογενή απαιτήσεις καθυστέρησης και δυνατότητες διασύνδεσης. Αναδυόμενες εφαρμογές στους τομείς της υγείας (smart medicine)[20][21], της μεταφοράς και κατασκευής αγαθών (smart manufacturing), της γεωργίας και ηλεκτροδότησης (smart grids)[22] προβάλλουν την ανάγκη επεξεργασίας, σε πραγματικό χρόνο, μεγάλου υπολογιστικού φόρτου με την παράλληλη ασφάλεια και αξιόπιστη μεταφορά και αποθήκευση του. Μία από τις αναδυόμενες λύσεις για αυτό το πλαίσιο τεχνολογικής ανάπτυξης αποτελεί το edge-centric computing[23], το οποίο συνδυάζει τις αρχές ανάπτυξης του cloud computing με τις δυνατότητες που παρέχουν τα σύγχρονα δίκτυα επικοινωνιών για να αξιοποιήσει τα άκρα του δικτύου edge προς υπολογισμό.

Για να γίνει κατανοητή η ανάγκη μετάβασης από το υπολογιστικό νέφος (cloud computing) στην υπολογιστική 'άκρη' edge computing, αξίζει μια αναφορά στις δυνατότητες και στους περιορισμούς του cloud. Συγκεκριμένα, το cloud computing αποτελεί ένα σύνολο τεχνολογιών και ένα πρότυπο ανάπτυξης το οποίο παρέχει στον τελικό χρήστη την δυνατότητα να αποθηκεύει και να επεξεργάζεται όγκους δεδομένων και σε πολύ μεγαλύτερες ταχύτητες απ' ό,τι του επιτρέπει η περιορισμένη υπολογιστική ισχύ και χώρος των τοπικών συσκευών του. Κεντρικά συστήματα και διακομιστές με επαυξημένη υπολογιστική ισχύ και δυνατότητες παράλληλης επεξεργασίας, τα οποία διαχειρίζονται οι πάροχοι υπηρεσιών cloud, αναλαμβάνουν την εξυπηρέτηση των tasks των τελικών χρηστών και αποστέλουν το τελικό αποτέλεσμα μέσω της δικτυακής υποδομής. Η εξάρτηση αυτή από την υποδομή του Διαδικτύου αποτελεί και τον βασικό περιορισμό της τεχνολογίας καθώς οι μεγάλες αποστάσεις μεταξύ των διακομιστών cloud και των χρηστών εισάγει χρονικές καθυστερήσεις και αυξάνουν την πιθανότητα σφάλματος, καθιστώντας το σύστημα αναξιόπιστο για εφάρμογες ευαίσθητες στην καθυστέρηση[24].

Το edge computing αναπτύχθηκε για να εκμεταλλευτεί την σημαντική ποσότητα αδρανούς επεξεργαστικής ισχύος και αποθηκευτικού χώρου των συσκευών στην άκρη του δικτύου, που μεσολαβεί μεταξύ του τελικού χρήστη και του κλάουδ σεερ, φέρνοντας τις υπηρεσίες του κλάουδ computing πιο κοντά στον χρήστη. Η εφαρμογή αυτής της ιδέας στο τηλεπικοινωνιακό δίκτυο οδήγησε στην ανάπτυξη του Mobile Edge Computing, μια ιδέα αρχιτεκτονικής δικτύου που προτάθηκε από το Ευρωπαϊκό Ινστιτούτο Τηλεπικοινωνιακών Προτύπων (European Telecommunications Standard Institute (ETSI))[25], το 2014. Σύμφωνα με τον ορισμό του ETSI, 'το Mobile Edge Computing παρέχει δυνατότητες Cloud Computing



εντός του Δικτύου Ασύρματης Πρόσβασης (*Radio Access Network, RAN*) και σε κοντινή απόσταση από τις κινητές συσκευές. Ο στόχος τους είναι η μείωση της καθυστέρησης, η διασφάλιση της αποδοτικής λειτουργίας του δικτύου, η παροχή υπηρεσιών και η βελτιωμένη εμπειρία χρήστη'.



**Σχήμα 1.3:** Εικονική αναπαράσταση του cloud computing vs MEC

Βασικό χαρακτηριστικό του MEC αποτελεί η δυνατότητα εκφόρτωσης υπολογιστικών διεργασιών των συσκευών άκρης σε διακομιστές άκρης (*edge server*), οι οποίοι εγκαθίστανται στους σταθμούς βάσης BS, σε κοντινή απόσταση από τους χρήστες[24],[26],[27]. Οι υπολογιστικές εργασίες (*tasks*) μπορούν να εκφορτωθούν και να εκτελεστούν είτε ολόκληρες (*full offloading*) είτε μερικώς (*partial offloading*), προσφέροντας ευελιξία στην διεκπεραίωση των υπολογισμών αλλά και αυξήσεις στην ενεργειακή αποδοτικότητα. Τέλος, οι *edge servers* μπορούν να εκτελέσουν διεργασίες βελτιστοποίησης της διαχείρισης των πόρων τόσο του RAN όσο και των χρηστών, επιτυγχάνοντας κοινή χρήση των υπολογιστικών και τηλεπικοινωνιακών πόρων[28],[29],[30] και παρέχοντας την απαραίτητη υποδομή για την υποστήριξη των αναδυόμενων εφαρμογών πραγματικού χρόνου (AR, streaming, self-driving cars, IoT)[31] και τεχνητής νοημοσύνης στα δίκτυα 5G/6G.

### 1.3 Τεχνολογίες αιχμής - σχετικές έρευνες

Τα τελευταία χρόνια, έχουν διεξαχθεί αρκετές μελέτες γύρω από το NOMA και τις δυνατότητες που παρέχει, σε συνδυασμό με άλλες τεχνικές, για τα δίκτυα πολλαπλής πρόσβασης. Συγκεκριμένα, έχουν προταθεί πολλές διαφορετικές εκδοχές NOMA, όπως το power domain NOMA, το code domain NOMA, το multi carrier, το sparse code multiple access NOMA και το distributed NOMA οι οποίες στηρίζονται στην ίδια βασική αρχή, σύμφωνα με την οποία περισσότεροι από ένας χρήστες μπορούν να εξυπηρετηθούν από ένα ορθογώνιο μπλοκ πόρων [10], [11], [12], [9]. Τα πλεονεκτήματα της μεθόδου για την πολλαπλή πρόσβαση των χρηστών στο κανάλι οφείλονται στην εξοικονόμηση πόρων, αποφεύγοντας την αποκλειστική ανάθεση των μπλοκ σε χρήστες με κακές συνθήκες καναλιού [12].

#### Power Domain NOMA

Το power domain NOMA, έχει λάβει μεγάλη προσοχή ανάμεσα στις διαφορετικές αυτές εκδοχές για τη δυνατότητα του να πετυχαίνει τη χωρητικότητα του καναλιού εκπομπής στην κάτω (downlink) ζεύξη και του καναλιού πολλαπλής πρόσβασης στην άνω (uplink) ζεύξη [12], [32] με την χρήση Successive Interference Cancellation (SIC) από την πλευρά του δέκτη [33]. Σημαντικός παράγοντας για την επιτυχημένη εξάλειψη των παρεμβολών, αποτελεί η επιλογή μιας σειράς αποκωδικοποίησης των χρηστών στο δέκτη, με βάση την ακριβή γνώση των κερδών που επιφέρει το κανάλι σε κάθε μετάδοση και στην τυπική εκδοχή του NOMA επιλέγεται σταθερή και φθίνουσα με βάση το πόσο ισχυρό είναι το κανάλι κάθε χρήστη.

Κατά την uplink μετάδοση, το power domain NOMA χρήση time sharing μεταξύ των πιθανών σειρών εκπομπής των χρηστών-πομπών μπορεί να οδηγήσει στην βελτίωση της 'δικαιοσύνης' (fairness) όσον αφορά την κατανομή ρυθμών μετάδοσης και ενέργειας και την επίτευξη οποιουδήποτε σημείου της περιοχής χωρητικότητας (capacity region) του καναλιού πολλαπλής πρόσβασης [34], [16]. Η επιλογή time sharing επιτρέπει σε χρήστες με μειωμένες απολαβές ρυθμού σε δεδομένη σειρά αποκωδικοποίησης να βιώνουν καλύτερη αξιοπιστία λήψης σε κάποια άλλη, λόγω της εναλλαγής των σειρών, μέσα στο διάστημα μετάδοσης  $T$ . Η αυξανόμενη πολυπλοκότητα του  $\nu$  SIC με την αύξηση του αριθμού χρηστών και άρα και των πιθανών σειρών αποκωδικοποίησης μπορεί να καταπολεμηθεί με τον χωρισμό των χρηστών σε ζευγάρια και την εφαρμογή power-domain NOMA μεταξύ τους [33] και έχει αποδειχθεί πως με αυτή τη μέθοδο, μπορεί να βελτιστοποιηθεί ο συνολικός ρυθμός μετάδοσης (Sum Rate) του συνόλου των χρηστών [35].

#### Distributed NOMA

Δεδομένου της κυριαρχίας των uplink ζευξεων στις επερχόμενες εφαρμογές IoT καθώς και της αρχιτεκτονικής C-RAN, κρίνεται ιδιαίτερα σημαντική μια διανεμημένη εκδοχή του

NOMA, το Distributed Uplink NOMA (DU-NOMA). Συγκεκριμένα, στο [16] διερευνάται το πρόβλημα μεγιστοποίησης του capacity region σε ένα σενάριο χρήσης DU-NOMA στην άνω ζεύξης μιας C-RAN αρχιτεκτονικής, με την παραδοχή μη σταθερών ρυθμών μετάδοσης των χρηστών και ικανότητα επικοινωνίας των RRHs του συστήματος μέσω ζεύξεων υψηλής χωρητικότητας ποιοτικών ινών για την εκτέλεση SIC χωρίς σφάλματα. Η περίπτωση της εκτέλεσης SIC μέσω ζεύξεων περιορισμένης χωρητικότητας και μη αλάνθαστης επικοινωνίας μεταξύ των RRH έχει διερευνηθεί στο [10].

Η εφαρμογή NOMA σε συστήματα MEC έχει αναγνωριστεί στα [34],[35],[36], ως ένας συνδυασμός τεχνικών που μπορεί να οδηγήσει σε σημαντικές μειώσεις στην καθυστέρηση και την ενεργειακή κατανάλωση. Η υπάρχουσα βιβλιογραφία έχει επικεντρωθεί στα προτερήματα που μπορεί να παρέχει το NOMA σε σχέση με τις παραδοσιακές τεχνικές OMA και στην πιθανότητα οι τεχνικές αυτές να πρέπει να συνδυαστούν στο πλαίσιο του MEC. Συγκεκριμένα, οι διαφορές στην ανοχή καθυστέρησης (latency) των διαφορετικών χρηστών μπορεί να καταστήσει την χρήση αποκλειστικά NOMA ανεπαρκή κάτι που οδήγησε, στα [37], [38], στην υιοθέτηση υβριδικών NOMA-OMA αρχιτεκτονικών. Αναλυτικότερα, στις δημοσιεύσεις αυτές, σε ένα σενάριο εκφόρτωσης δεδομένων 2 χρηστών, κρατώντας σταθερή την ανεκτή καθυστέρηση και την ισχύ εκπομπής ενός από τους δύο χρήστες και θεωρώντας ότι ο άλλος έχει μεγαλύτερη ανοχή στην καθυστέρηση, επιλύεται το πρόβλημα ελαχιστοποίησης καθυστέρησης και κατανάλωσης ενέργειας. Κάθε πλαίσιο μετάδοσης χωρίζεται σε μία φάση όπου οι χρήστες εκπέμπουν ταυτόχρονα κατά NOMA και σε μία φάση όπου εκπέμπει μόνο ο χρήστης με ανοχή στην καθυστέρηση, κατά OMA, ώστε να ολοκληρώσει την εκφόρτωση του. Κατά την διαδικασία SIC, επιλέγεται σταθερή σειρά αποκωδικοποίησης με το μήνυμα του χρήστη με σταθερή καθυστέρηση να αποκωδικοποιείται πάντα τελευταίο, κατά την φάση NOMA της διαδικασίας, και η εκφόρτωση να εκτελείται πλήρως (full offloading) στους edge servers.

## Partial Offloading

Στις έρευνες [40], [41], [42], [43] αναλύθηκαν μοντέλα με δυνατότητα απόφασης μερικής εκφόρτωσης (partial offloading) σε edge servers. Οι [40], [41] επικεντρώθηκαν στην ελαχιστοποίηση της κατανάλωσης ενέργειας, ενώ τα [42], [43] στην ελαχιστοποίηση του χρόνου ολοκλήρωσης των tasks των χρηστών με χρήση partial offloading σε συστήματα MEC.

Η σχέση των παραμέτρων της τοπικής επεξεργασίας στις συσκευές άκρης με τη δυνατότητα εκφόρτωσης αποτελούν μια ακόμη σημαντική κατεύθυνση της βιβλιογραφίας στα MEC συστήματα. Έτσι, στο [39], εισάγεται η δυνατότητα ελέγχου της ταχύτητας της κεντρικής μονάδας επεξεργασίας της συσκευής για τον υπολογισμό μέρους ενός task, μέσω της τεχνικής δυναμικής κλιμάκωσης της τάσης και στην συνέχεια επιλύεται ένα πρόβλημα από κοινού βελτιστοποίησης της υπολογιστικής ταχύτητας, της εκπεμπόμενης ισχύος της συσκευής και το λόγο εκφόρτωσης (offloading ratio) με αντικειμενικές συναρτήσεις (objective

functions) την ενεργειακή κατανάλωση και το συνολικό χρόνο εκτέλεσης των εργασιών. Παρατηρήθηκε πως οι δύο αυτές μετρικές παρουσιάζουν trade off και πως ο έλεγχος του ποσοστού εκφόρτωσης και της ταχύτητας της CPU μπορεί να τις εξισορροπήσει αποτελεσματικά.

Μια κοινή μετρική της ποιότητας επικοινωνίας Quality of Experience η οποία βασίζεται στο χρόνο εκτέλεσης διεργασιών και στην ενεργειακή κατανάλωση της εκάστοτε συσκευής, μεγιστοποιείται στο [44] μέσω ενός ευρετικού (heuristic) αλγορίθμου κυρτής βελτιστοποίησης που βρίσκει τμηματικά, την κατάλληλη χρονική στιγμή της απόφασης εκφόρτωσης και την βέλτιστη διαχείριση των υπολογιστικών και επικοινωνιακών πόρων του proximal cloud. Η διαχείριση του υπολογιστικού φόρτου του edge server ενός MEC συστήματος που χρησιμοποιεί NOMA για την εκφόρτωση, μπορεί να υποβοηθηθεί με την χρήση επικοινωνίας συσκευής-προς-συσκευή (D2D communications), όπως αναλύεται στο [45]. Στην συγκεκριμένη δημοσίευση, επιλέχθηκε το σταθμισμένο άθροισμα της ενέργειακής κατανάλωσης και συνολικής καθυστέρησης όλων των χρηστών του συστήματος, ενώ ως μέθοδος βελτιστοποίησης του μη κυρτού προβλήματος για την ανάθεση ισχύος στα ζεύγη D2D επιλέχθηκε ένας ευρετικός αλγόριθμος βασισμένος στο particle swarm optimization (PSO).

Σε αντίθεση με τις προαναφερθείσες εργασίες, στο [46] η βελτιστοποίηση των κοινών υπολογιστικών και επικοινωνιακών πόρων λαμβάνει υπόψη, τόσο την επίδραση της σειράς αποκωδικοποίησης των χρηστών κατά τη SIC διεργασία, όσο και την ικανότητα μεταβολής της υπολογιστικής ταχύτητας με ελεγχόμενη, πλέον, ταχύτητα ρολογιού (CPU). Παρόμοια διερεύνηση τελείται στο [27], σε ένα υβριδικό πρωτόκολλο NOMA-OMA, με την προσθήκη time-sharing μεταξύ διαφορετικών σειρών αποκωδικοποίησης κατά την εξάλειψη της παρεμβολής κατά τη φάση NOMA αλλά και χρησιμοποιώντας μεταβλητή ταχύτητα ρολογιού για την τοπική επεξεργασία. Αναλυτικότερα, όπως στις δημοσιεύσεις [37] και [38], όπου χρησιμοποιήθηκαν δυο φάσεις, NOMA για την ταυτόχρονη εκφόρτωση των χρηστών και OMA για τη εναπομένουσα εκφόρτωση του πιο αδύναμου, έτσι και στην [27], χρησιμοποιείται μια φάση με time sharing NOMA για επίτευξη οποιουδήποτε rate του καναλιού πολλαπλής πρόσβασης και μία φάση OMA για την ολοκλήρωση των εκφορτώσεων. Στην περίπτωση της πλήρους εκφόρτωσης εξετάζεται το πρόβλημα ελαχιστοποίησης της καθυστέρησης, περιορίζοντας την ενεργειακή κατανάλωση (delay-constrained energy minimization) και το πρόβλημα ελαχιστοποίησης της ενέργειας, περιορίζοντας την αποδεκτή καθυστέρηση (energy-constrained delay minimization). Στην περίπτωση του partial offloading, εξετάζεται το πρόβλημα ελαχιστοποίησης του σταθμισμένου αθροίσματος των καθυστερήσεων για την ολοκλήρωση των εργασιών των χρηστών, με ελεγχόμενη ταχύτητα ρολογιού της CPU.

## 1.4 Συνεισφορά

Οι κύριες συνεισφορές παρουσιάζονται συνοπτικά παρακάτω:

- Περιγράφεται το MEC με εφαρμογή πρωτοκόλλου DU-NOMA κατά το οποίο χρησιμοποιείται time sharing κατά τη διαδικασία SIC, μέσω της επικοινωνίας των RRH και περιγράφονται τα rates που επιτυγχάνουν οι χρήστες για κάθε πιθανή σειρά αποκωδικοποίησης των χρηστών.
- Διατυπώνονται τα προβλήματα ελαχιστοποίησης της καθυστέρησης, περιορίζοντας την ενεργειακή κατανάλωση (delay-constrained energy minimization) και το πρόβλημα της ελαχιστοποίησης της ενέργειας, περιορίζοντας την αποδεκτή καθυστέρηση (energy-constrained delay minimization) για την περίπτωση του full offloading, και το πρόβλημα ελαχιστοποίησης της καθυστέρησης εκφόρτωσης, λαμβάνοντας υπ' όψη τους περιορισμούς της τοπικής επεξεργασίας και τη δυνατότητα δυναμικού ελέγχου της ταχύτητας της.
- Επιλύονται τα προβλήματα βελτιστοποίησης αφού μετατραπούν σε κυρτές διατυπώσεις μέσω μιας συνδυαστικής μεθοδολογίας γεωμετρικού προγραμματισμού (geometric programming) και διαδοχικής κυρτής βελτιστοποίησης μέσω προγραμματισμού Διαφοράς Κυρτών συναρτήσεων (Difference-of-Convex (DC) optimization).
- Αναλύεται το αποτέλεσμα των διαφορετικών αποστάσεων των χρηστών σε σχέση με τους σταθμούς βάσης, στη βέλτιστη καθυστέρηση και ενεργειακή κατανάλωση του συστήματος και συγκρίνεται η αποτελεσματικότητα της μεθοδολογίας σε σχέση με μια απλούστερη μεθοδολογία επίλυσης των προβλημάτων, βασισμένη στην εξαντλητική αναζήτηση επιτρεπτών αναθέσεων ισχύος κατά το time sharing.

## 1.5 Δομή εργασίας

Το υπόλοιπο της διπλωματικής εργασίας οργανώνεται ως εξής. Στο **Κεφάλαιο 2** παρουσιάζεται το μοντέλο και η λειτουργία του MEC με εφαρμογή DU-NOMA. Πιο συγκεκριμένα, μελετώνται όλα τα πιθανά σενάρια επικοινωνίας των χρηστών με τα RRHs, εξάγονται οι αντίστοιχες μαθηματικές σχέσεις των rates των χρηστών και οι περιορισμοί ενέργειας, καθυστέρησης, υπολογιστικού φορτίου και τοπικής επεξεργασίας για τις περιπτώσεις της πλήρους και μερικής εκφόρτωσης. Στη συνέχεια, διατυπώνεται το πρόβλημα βελτιστοποίησης της ελαχιστοποίησης της καθυστέρησης εκτέλεσης διεργασίας και το πρόβλημα της ελαχιστοποίησης της ενεργειακής κατανάλωσης, στην περίπτωση της μεθόδου της πλήρους εκφόρτωσης (full offloading) και το πρόβλημα ελαχιστοποίησης της καθυστέρησης μερικής εκφόρτωσης. Στο **Κεφάλαιο 3** παρουσιάζεται η μεθοδολογία επίλυσης των προβλημάτων μέσω μετασχηματισμών των αρχικών διατυπώσεων σε μορφές επιλύσιμες από ρουτίνες βελτιστοποίησης. Στο **Κεφάλαιο 4**, παρουσιάζεται μια απλούστερη μεθοδολογία επίλυσης ως μέτρο σύγκρισης για τις νέες μεθόδους και αναλύονται τα αποτελέσματα των προσομοιώσεων των παραπάνω προβλημάτων για την πιστοποίηση της αποτελεσματικότητας των μεθόδων.

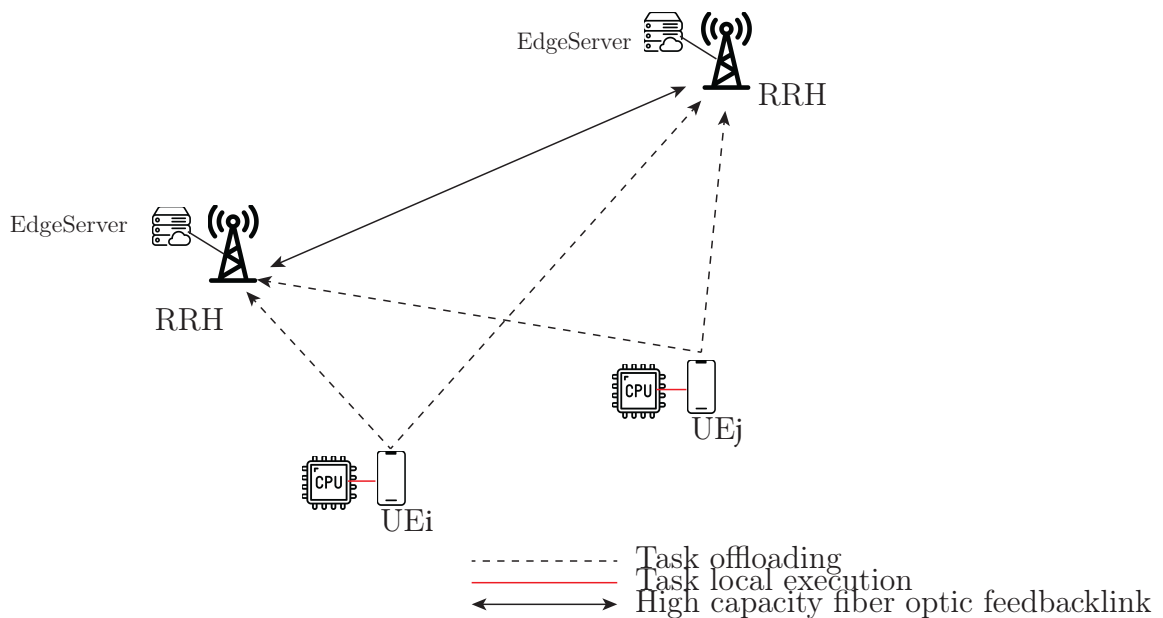
Τέλος, στο **Κεφάλαιο 5**, συνοψίζονται τα συμπεράσματα και προτείνονται πιθανές μελλοντικές προεκτάσεις της εργασίας.

## Κεφάλαιο 2

# Παρουσίαση της προτεινόμενης αρχιτεκτονικής - διατύπωση του προβλήματος

### 2.1 Μοντέλο και διάταξη συστήματος

Το μοντέλο MEC με DU-NOMA αποτελείται από 2 χρήστες (User Equipment):  $UE_i, UE_j$  οι οποίοι στέλνουν και δέχονται δεδομένα από δύο RRHs:  $RRH_a, RRH_b$ , όπως φαίνεται στην παρακάτω εικόνα:



Σχήμα 2.1: Εικονική αναπαράσταση του DU-NOMA MEC

Η διάταξη προτίθεται για την υλοποίηση ενός σεναρίου MEC, κατά το οποίο οι χρήστες επικοινωνούν με τα RRHs και εκφορτώνουν υπολογιστικό φόρτο προς εκτέλεση. Οι ραδιοσταθμοί εκτελούν την διεργασία βελτιστοποίησης της εκφόρτωσης μέσω ενός edge server,

προυποθέτωντας τέλεια πληροφορία κατάστασης του καναλιού (perfect channel state information) και υλοποίηση time-sharing μεταξύ διαφορετικών σεναρίων επικοινωνίας με τους χρήστες[10]. Έστω  $T$  η συνολική χρονική διάρκεια της φάσης του time-sharing και έστω  $g_{l,m} = |h_{l,m}|^2 w_{lm}$  το κέρδος καναλιού του χρήστη  $l$ ,  $\forall l \in i, j$  προς το  $RRH_m, \forall m \in \{a, b\}$  στο οποίο συμπεριλαμβάνονται οι διαλείψεις μικρής κλίμακας (small scale fading),  $h_{lm}$  και ο παράγοντας απώλειας διαδρομής  $w_{lm}$  για κάθε επιλογή χρήστη και ραδιοπομπού. Όπως θα γίνει αντιληπτό πιο διεξοδικά από την περιγραφή του πρωτοκόλλου στην επόμενη ενότητα, σε κάθε ένα από τα 8 σενάρια επικοινωνίας που προκύπτουν μεταξύ χρηστών και ραδιοπομπών, η ισχύς που καταναλώνεται από τον εκάστοτε χρήστη κατά τη διάρκεια του σεναρίου

$k$ , συμβολίζεται με  $P_l^{(k)}$  και το αντίστοιχο SNR ως  $p_l^{(k)} = \frac{P_l^{(k)}}{BN_o}, \forall l \in i, j$  και  $\forall k = 1, 2, \dots, 8$ . Το  $N_o$  αποτελεί την φασματική πυκνότητα του Γκαουσιανού Προσθετικού Θορύβου (AWGN) και  $B$ , το διαθέσιμο εύρος ζώνης.

Κάθε UE διαθέτει μονάδα CPU η οποία μπορεί να χρησιμοποιηθεί για την τοπική επεξεργασία δεδομένων, υποστηρίζοντας δύο σενάρια εκφόρτωσης: τόσο την πλήρη εκφόρτωση, όπου όλα τα bits της διεργασίας ανατίθενται στους edge servers των RRH και επιστρέφονται στους χρήστες [offloading ref], όσο και την μερική (partial offloading) στην οποία το ποσοστό τοπικής επεξεργασίας από τις CPU και της επεξεργασίας στο edge είναι μεταβλητό. Το κόστος αποστολής και λήψης σε χρόνο και ενέργεια καθώς και το ενεργειακό κόστος της επεξεργασίας σε κάθε RRH παραβλέπονται σε αυτή τη μελέτη.

## 2.2 Λειτουργία συστήματος - Rates χρηστών

Σύμφωνα με το πρωτόκολλο του DU-NOMA, οι χρήστες στέλνουν ταυτόχρονα τα μηνύματά τους και τα RRH συνεργάζονται για την αποκωδικοποίησή των μηνυμάτων, επικοινωνώντας μέσω ενός ιδανικού καναλιού οπτικών ινών, υψηλής χωρητικότητας (error free feedback link). Κατά την αποκωδικοποίηση των μηνυμάτων, εκτελείται SIC και μπορεί να εφαρμοστεί είτε σταθερή σειρά αποκωδικοποίησης των χρηστών (fixed decoding order) είτε time sharing, κατά το οποίο όλα τα πιθανά σενάρια αποκωδικοποίησης χρησιμοποιούνται για κάποιο ποσοστό της χρονικής διάρκειας  $T$ . Το ποσοστό αυτό συμβολίζεται με  $s_k, \forall k = 1, 2, \dots, 8$  και μαζί με την ισχύ κάθε σεναρίου  $P_l^{(k)}$  αποτελούν μεταβλητές βελτιστοποίησης για τον καθορισμό των εφικτών ρυθμών μετάδοσης και ενεργειακής κατανάλωσης των χρηστών.

Είναι σημαντικό να σημειωθεί πως η οπτική ίνα ανάδρασης (feedback) μεταξύ των RRH διασφαλίζει πώς σε κάθε φάση της time-sharing διαδικασίας, η καταστολή της παρεμβολής και η προβλεπόμενη από το NOMA, αποκωδικοποίηση μπορεί να πραγματοποιηθεί από οποιοδήποτε ραδιοπομπό και στην εντέλεια [16].

Με βάση την παραπάνω ανάλυση, σε κάθε σενάριο της διεργασίας αντιστοιχεί μία σειρά αποκωδικοποίησης  $z = (u, w), (v, x)$  όπου το μήνυμα  $u$  αποκωδικοποιείται πρώτο, θεωρώντας το μήνυμα  $v$  ως παρεμβολή, από το  $RRH_w$  και το μήνυμα  $v$  αποκωδικοποιείται δεύτερο από το  $RRH_x$  χρησιμοποιώντας SIC. Έχοντας 2 UE και 2 χρήστες, προκύπτουν 8 σειρές



αποκωδικοποίησης και για δεδομένη σειρά  $z$ , οι εφικτοί ρυθμοί δεδομένων δίνονται από τις εξής σχέσεις:

$$R_u^{(k)} = \log \left( 1 + \frac{g_{uw}p_u^{(k)}}{1 + g_{vw}p_v^{(k)}} \right) \quad (2.1)$$

$$R_v^{(k)} = \log \left( 1 + g_{vx}p_v^{(k)} \right) \quad (2.2)$$

$$\forall (u, v) \in i, j, \forall (w, x) \in a, b$$

Είναι σημαντικό να σημειωθεί πώς στις πρώτες 4 φάσεις του time-sharing, θεωρείται πως η σειρά αποκωδικοποίησης ξεκινά με την αποκωδικοποίησή του μηνύματος του ίδιου χρήστη, πάντα πρώτου, ανεξαρτήτως από τις συνθήκες του καναλιού ή τους περιορισμούς σε ενέργεια και ρυθμό δεδομένων. Για διευκόλυνση της κατανόησης, παρουσιάζονται οι ρυθμοί που προκύπτουν στην δεύτερη φάση της διαδικασίας. Οι χρήστες  $i, j$  εκπέμπουν με ισχύ  $P_i^{(k)}, P_j^{(k)}$  και 'βιώνουν' κέρδος καναλιού  $g_{ia}$  και  $g_{ja}$  προς το  $RRH_a$  και  $g_j b$  προς το  $RRH_b$ . Το μήνυμα του χρήστη  $i$  αποκωδικοποιείται πρώτο από το  $RRH_a$ , λαμβάνοντας το μήνυμά του  $j$  ως παρεμβολή και στην συνέχεια αποστέλλεται στο  $RRH_b$ , όπου η αποκωδικοποίηση του  $j$  γίνεται, μέσω SIC, χωρίς παρεμβολή. Τα data rates που επιτυγχάνουν τα UE σε αυτή την περίπτωση είναι:

$$R_i^{(1)} = \log \left( 1 + \frac{g_{ia}p_i^{(1)}}{1 + g_{ja}p_j^{(1)}} \right) \quad (2.3)$$

$$R_j^{(1)} = \log \left( 1 + g_{jb}p_j^{(1)} \right) \quad (2.4)$$

Κάθε ένα από τα σενάρια αποκωδικοποίησης, με τους αντίστοιχους ρυθμούς δεδομένων, χρησιμοποιούνται από το σύστημα για το ποσοστό  $s_k$  της διάρκειας  $T$  και επομένως οι συνολικοί ρυθμοί δίνονται από τις σχέσεις:

$$R_{i,total} = \sum_{k=1}^8 s_k R_i^{(k)} \quad (2.5)$$

$$R_{j,total} = \sum_{k=1}^8 s_k R_j^{(k)} \quad (2.6)$$

Από τις παραπάνω σχέσεις καθίσταται σαφές πως, όταν χρησιμοποιείται NOMA με time-sharing, οποιοδήποτε σημείο της περιοχής χωρητικότητας (capacity region) μπορεί να επιτευχθεί [25] και αντίστοιχα, ο αριθμός των bits τα οποία μπορούν να αποτεθούν από τους χρήστες  $i, j$  περιορίζεται από την ίδια. Αυτή η συνθήκη εκφράζεται με τις παρακάτω ανισότητες:

$$TBR_{i,total} \geq \tilde{N}_i \quad (2.7)$$

$$TBR_{i,total} \geq \tilde{N}_j \quad (2.8)$$

όπου ως  $N_i$ ,  $N_j$  συμβολίζονται τα *bits* τα οποία μπορεί κάθε χρήστης να εναποθέσει στους ραδιοπομπούς. Στην επόμενη ενότητα, παρουσιάζονται οι σχέσεις που περιγράφουν την ενεργειακή κατανάλωση και την καθυστέρηση στις περιπτώσεις της πλήρους και μερικής εναπόθεσης (full and partial offloading).

## 2.3 Περίπτωση Full Offloading

Στο σενάριο του full offloading, υποθέτουμε πως ο υπολογιστικός φόρτος κάθε χρήστη διατίθεται αποκλειστικά στους διακομιστές άκρης (edge server) που βρίσκεται σε κάθε RRH και επομένως, η μεμονωμένη καθυστέρηση καθορίζεται από την καθυστέρηση εκφόρτωσης:

$$\begin{aligned} T_{FO,i} &= T_{o,i} = T \\ T_{FO,j} &= T_{o,j} = T \end{aligned} \quad (2.9)$$

Επίσης, η ενέργεια που καταναλώνεται από κάθε χρήστη συμβολίζεται με  $E_{o,i}, E_{o,j}$  και μπορεί να εκφραστεί ως την ισχύ σε κάθε φάση του time sharing επί το χρονικό διάστημα που καταναλώνεται αυτή η ισχύς. Στο συγκεκριμένο σενάριο η ενέργεια καταναλώνεται εξ' ολοκλήρου για την ανάθεση των δεδομένων ( $E_{FO,i}, E_{FO,j}$ ):

$$E_{o,i} = E_{FO,i} = T \cdot p_i = T \cdot \sum_{k=1}^8 s_k \cdot p_i^{(k)} \quad (2.10)$$

$$E_{o,j} = E_{FO,j} = T \cdot p_j = T \cdot \sum_{k=1}^8 s_k \cdot p_j^{(k)} \quad (2.11)$$

Η ενεργειακή κατανάλωση πρέπει να περιορίζεται, σύμφωνα με τις απαιτήσεις ενεργειακής αποδοτικότητας για το εκάστοτε σύστημα και άρα, ισχύουν οι παρακάτω ανισότητες για μέγιστη κατανάλωση,  $E_i, E_j$ :

$$E_{o,i} \leq E_i \quad (2.12)$$

$$E_{o,j} \leq E_j \quad (2.13)$$

Αντίστοιχα, ο αριθμός των bits που μπορούν να ανατεθούν,  $\tilde{N}_i, \tilde{N}_j$  πρέπει να είναι τουλάχιστον ίσα με τον αριθμό των bits,  $N_i, N_j$  που χρειάζονται για την επίτευξη του υπολογισμού στον διακομιστή.

$$\tilde{N}_i \geq N_i \quad (2.14)$$

$$\tilde{N}_j \geq N_j \quad (2.15)$$

## 2.4 Περίπτωση Partial Offloading

Σε αυτή την περίπτωση, μέρος του υπολογιστικού φόρτου εναποτίθεται στον edge server ενώ το υπόλοιπο εναποτίθεται στην CPU του UE για τοπική επεξεργασία. Η καθυστέρηση

τοπικής επεξεργασίας για τον υπολογισμό της επιμέρους εργασίας(sub-task δίνεται από την εξής σχέση:

$$T_{loc,u} = \frac{L_u \cdot X_u}{f_u} \quad \forall u \in i, j \quad (2.16)$$

όπου  $L_i$  είναι το μέγεθος του sub-task σε bits,  $X_i$  το υπολογιστικό φόρτο σε CPU cycles ανα bit και  $f_i$ , η ταχύτητα του ρολογιού της τοπικής CPU. Η ταχύτητα αυτή μπορεί να ρυθμιστεί δυναμικά χρησιμοποιώντας την τεχνική δυναμικής κλιμάκωσης τάσης (dynamic voltage scaling) και η μετάδοση και τοπική επεξεργασία μπορούν να εκτελεστούν παράλληλα. Αντίστοιχα, η καθυστέρηση εκτέλεσης ολόκληρης της υπολογιστικής εργασίας του χρήστη  $u$  σε partial offloading(PO) μπορεί να εκφραστεί ως εξής:

$$T_{PO,u} = \max(T_{o,u}, T_{loc,u}) \quad \forall u \in i, j \quad (2.17)$$

καθώς η συνολική καθυστέρηση καθορίζεται είτε από την καθυστέρηση τοπικής επεξεργασίας,  $T_{o,i}$  είτε από την καθυστέρηση ανάθεσης του task,  $T_{loc,i}$ . Θέτοντας  $L_i = 0$ , το σενάριο του full offloading καλύπτεται καθώς δεν υπάρχουν bits προς τοπική επεξεργασία.

Κάθε κύκλος της CPU καταναλώνει ενέργεια ανάλογη της  $f_i^2$  και μπορεί να εκφραστεί ως  $k_i \cdot f_i^2$ , όπου το  $k_i$  είναι μια σταθερά η οποία σχετίζεται με την αρχιτεκτονική του υλικού (hardware architecture). Για μια υπο-εργασία μεγέθους  $L_i$  και υπολογιστικού φόρτου  $X_i$ , η ενέργεια τοπικής επεξεργασίας μπορεί να διατυπωθεί ως [8]:

$$E_{loc,u} = k_u L_u X_u f_u^2 \quad \forall u \in i, j \quad (2.18)$$

και η συνολική ενέργεια που καταναλώνεται δίνεται ως:

$$E_{PO,u} = E_{o,u} + E_{loc,u} \quad \forall u \in i, j \quad (2.19)$$

Τέλος, ο αριθμός των bits τα οποία υποβάλλονται σε τοπική επεξεργασία,  $L_u$  ή επιτυχημένα ανατίθενται στους διακομιστές άκρης,  $\tilde{N}_u$  περιορίζονται από την περιοχή χωρητικότητας του καναλιού πολλαπλής πρόσβασης και επομένως, ισχύει η παρακάτω ανισότητα:

$$\tilde{N}_u + L_u \geq N_u. \quad \forall u \in i, j \quad (2.20)$$

## 2.5 Πρόβλημα βελτιστοποίησης: Full offloading

Σε αυτή την ενότητα, διατυπώνονται τα προβλήματα βελτιστοποίησης για την ενεργειακά περιορισμένη, ελαχιστοποίηση της καθυστέρησης και την περιορισμένη απο πλευράς καθυστέρησης, ελαχιστοποίηση της ενέργειας του προτεινόμενου πρωτοκόλλου full offloading. Οι περιορισμοί επιτεύξιμων ρυθμών δεδομένων που δίνονται από τις σχέσεις (2.7), (2.8) καθώς και οι περιορισμοί ενεργειακής κατανάλωσης, (2.12), (2.13) λαμβάνονται υπόψη για την ολοκλήρωση της διατύπωσης.

### 2.5.1 Energy constrained delay minimization

Η μετρική απόδοσης που επιλέγεται για ελαχιστοποίηση είναι η συνολική διάρκεια του time-sharing για την πλήρη εκφόρτωση των bits των χρηστών,  $T_{FO,i} = T_{FO,j} = T$ . Να σημειωθεί πως και οι δύο χρήστες πρέπει να επιτύχουν την εκφόρτωση ολόκληρου του υπολογιστικού φορτίου τους μέσα στο ίδιο διάστημα  $T$ . Το πρόβλημα βελτιστοποίησης μπορεί να διατυπωθεί ως εξής:

$$\begin{aligned}
 & \min_{\mathbf{T}, \mathbf{p}, \tilde{\mathbf{N}}, \mathbf{s}} T \\
 & \text{s.t.} \quad C_1 : TBR_{i,total} \geq \tilde{N}_i, \\
 & \quad C_2 : TBR_{j,total} \geq \tilde{N}_j, \\
 & \quad C_3 : E_{o,i} \leq E_i, \\
 & \quad C_4 : E_{o,j} \leq E_j, \\
 & \quad C_5 : p_j^{(k)}, p_i^{(k)} \geq 0, \forall k = 1, \dots, 8,
 \end{aligned} \tag{2.21}$$

όπου ως  $\mathbf{T}, \mathbf{p}, \tilde{\mathbf{N}}$  συμβολίζονται τα διανύσματα που περιλαμβάνουν τις μεταβλητές  $T_{FO,i}, T_{FO,j}, p_i, p_j, \tilde{N}_i, \tilde{N}_j$ , αντίστοιχα. Οι  $C_1, C_2$  αντιπροσωπεύουν το κάτω όριο στον αριθμό των bits που πρέπει να εκφορτωθούν, όπως ορίζεται από τις (2.7), (2.8), ενώ οι  $C_3, C_4$  αντιστοιχούν στην μέγιστη διαθέσιμη ενέργεια για την επίτευξη της διεργασίας από κάθε χρήστη, όπως ορίζεται στις σχέσεις (2.10), (2.11). Ο περιορισμός  $C_5$  εξασφαλίζει ότι οι ισχύς σε κάθε φάση του χρονικού διαμοιρασμού θα είναι μεγαλύτερη ή ίση του μηδενός.

Σύμφωνα με την ανάλυση της ενότητας (2.2) και συγκεκριμένα τις σχέσεις (2.5), (2.6) μπορούμε να εισάγουμε την μεταβλητή  $s_k$  ως το ποσοστό του συνολικού rate που επιτυγχάνεται από την σειρά αποκωδικοποίησης της φάσης  $k$  του time-sharing. Το πρόβλημα βελτιστοποίησης μπορεί να διατυπωθεί ως εξής:

$$\begin{aligned}
 & \min_{T, \mathbf{p}, \mathbf{s}} T \\
 & \text{s.t.} \quad C_1 : s_1 TBR_i^{(1)} + s_2 TBR_i^{(2)} + \dots + s_7 TBR_i^{(7)} + s_8 TBR_i^{(8)} \geq \tilde{N}_i, \\
 & \quad C_2 : s_1 TBR_j^{(1)} + s_2 TBR_j^{(2)} + \dots + s_7 TBR_j^{(7)} + s_8 TBR_j^{(8)} \geq \tilde{N}_j, \\
 & \quad C_3 : T \sum_{k=1}^8 s_k P_i^{(k)} \leq E_i, \\
 & \quad C_4 : T \sum_{k=1}^8 s_k P_j^{(k)} \leq E_j, \\
 & \quad C_5 : T, s_k, P_i^{(k)}, P_j^{(k)} \geq 0, \\
 & \quad C_6 : \sum_{k=1}^8 s_k \leq 1, \forall k = 1, \dots, 8
 \end{aligned} \tag{2.22}$$

όπου προστίθεται περιορισμοί  $C_5, C_6$  οι οποίοι εξασφαλίζουν ότι οι μεταβλητές  $s_k$  εκφράζουν ποσοστά της μονάδας και το διάνυσμα αυτών των ποσοστών  $\mathbf{s}$  ως μεταβλητή βελτιστοποίησης για το πρόβλημα.

### 2.5.2 Delay constrained energy minimization

Στην περίπτωση της ελαχιστοποίησης της ενεργειακής κατανάλωσης, ένα αντίστοιχο πρόβλημα βελτιστοποίησης μπορεί να διατυπωθεί λαμβάνοντας υπ' όψη τους περιορισμούς στους

ρυθμούς δεδομένων των σχέσεων (2.7) και (2.8). Η μετρική απόδοσης προς βελτιστοποίηση είναι το σταθμισμένο άθροισμα της ενεργειακής κατανάλωσης των χρηστών, η οποία ορίζεται ως:

$$\mathcal{E}_{FO} = w_i E_{FO,i} + w_j E_{FO,j} \quad (2.23)$$

όπου  $w_l \in [0, 1] \forall \{i, j\}$  αποτελεί μια θετική σταθερά η οποία χρησιμοποιείται για την α-νάθεση διαφορετικών προτεραιοτήτων στους χρήστες από ανώτερα στρώματα του δικτύου, παρέχοντας διαφορετικές έννοιες δικαιοσύνης (fairness) μεταξύ των χρηστών [fairness ref]. Κατά την διάρκεια της διεργασίας, τα βάρη  $w_l$  μπορούν να προσαρμοστούν ώστε να αλλάξουν την επιτρεπόμενη από τους χρήστες, ενεργειακή κατανάλωση. Σε αντίθεση με το πρόβλημα βελτιστοποίησης της παραπάνω ενότητας, δεν λαμβάνονται υπόψη οι περιορισμοί των σχέσεων (2.12), (2.13) ενώ προστίθεται ένας περιορισμός για την επιτρεπτή καθυστέρηση  $T \leq T_{max}$ . Έτσι το πρόβλημα μπορεί με βάση τις σχέσεις  $E_{FO,i} = T \cdot \sum_{k=1}^8 s_k \cdot P_i^k$  και  $E_{FO,j} = T \cdot \sum_{k=1}^8 s_k \cdot P_j^k$  να εκφραστεί ως:

$$\begin{aligned} \min_{\mathbf{T}, \mathbf{p}, \mathbf{s}} \quad & w_i T \sum_{k=1}^8 s_k \cdot P_i^{(k)} + w_j T \sum_{k=1}^8 s_k \cdot P_j^{(k)} \\ \text{s.t.} \quad & C_1 : s_1 T B R_i^{(1)} + s_2 T B R_i^{(2)} + \dots + s_7 T B R_i^{(7)} + s_8 T B R_i^{(8)} \geq \tilde{N}_i \\ & C_2 : s_1 T B R_j^{(1)} + s_2 T B R_j^{(2)} + \dots + s_7 T B R_j^{(7)} + s_8 T B R_j^{(8)} \geq \tilde{N}_j, \\ & C_3 : T \leq T_{max}, \\ & C_4 : \sum_{n=1}^8 s_n \leq 1, \\ & C_5 : T, s_k, P_i^{(k)}, P_j^{(k)} \geq 0 \quad \forall k = 1, \dots, 8. \end{aligned} \quad (2.24)$$

Οι περιορισμοί  $C_1$  και  $C_2$  εκφράζουν τον αριθμό των bits που πρέπει να εκφορτωθούν σε διάστημα  $T$  και οι σχέσεις για τα  $R_i, j^{(k)}$  δίνονται από την ανάλυση της ενότητας 2.2. Οι  $C_4$  και  $C_5$  εξασφαλίζουν πως η λύση του προβλήματος θα είναι φυσικά υλοποιήσιμη, δηλαδή η καθυστέρηση  $T$  και οι ισχύς  $p^{(k)}$  είναι θετικές ποσότητες και τα  $s_k$  εκφράζουν ποσοστά της μονάδας.

## 2.6 Πρόβλημα βελτιστοποίησης: Partial offloading

Στην ενότητα αυτή, διατυπώνεται ένα πρόβλημα ελαχιστοποίησης της συνολικής καθυστέρησης  $T_{FO}$  όπως ορίζεται από την σχέση (2.17). Η μετρική αυτή συμπεριλαμβάνει τόσο την τοπική καθυστέρηση σύμφωνα  $T_{loc}$  και την καθυστέρηση εκφόρτωσης  $T_o$ . Η μεταβλητή  $T_{FO}$  μπορεί να χρησιμοποιηθεί για να εκφράσει περιορισμούς εκφόρτωσης bits. Σύμφωνα με τη σχέση (2.19) μπορούμε να εκφράσουμε περιορισμούς ενεργειακής κατανάλωσης που αφορούν τις συνολικές δαπάνες για τοπική επεξεργασία και εκφόρτωσης θέτοντας όρια και για τους

δύο χρήστες. Μια πιθανή διατύπωση του προβλήματος ελαχιστοποίησης είναι η εξής:

$$\begin{aligned}
 & \min_{\mathbf{T}_o, \mathbf{p}, \tilde{\mathbf{N}}, \mathbf{L}, \mathbf{f}, \mathbf{s}} T_{PO} \\
 & \text{s.t.} \quad C_1 : T_{o,i} BR_{i,total} \geq \tilde{N}_i, \\
 & \quad C_2 : T_{o,j} BR_{j,total} \geq \tilde{N}_j, \\
 & \quad C_3 : \tilde{N}_i + L_i \geq N_i, \\
 & \quad C_4 : \tilde{N}_j + L_j \geq N_j, \\
 & \quad C_5 : E_{PO,i} \leq E_i, \\
 & \quad C_6 : E_{PO,j} \leq E_j, \\
 & \quad C_7 : f_i \leq f_{max}, \\
 & \quad C_8 : f_j \leq f_{max}, \\
 & \quad C_9 : T_{o,i}, T_{o,j}, \tilde{N}_i, \tilde{N}_j, L_i, L_j, f_i, f_j, p_i^{(k)}, p_j^{(k)} \geq 0, \forall k = 1, \dots, 8
 \end{aligned} \tag{2.25}$$

όπου ως  $\mathbf{T}_o, \tilde{\mathbf{N}}, \mathbf{L}, \mathbf{f}, \mathbf{p}, \mathbf{s}$  συμβολίζονται τα διανύσματα που περιέχουν τις μεταβλητές βελτιστοποίησης  $T_{o,i}, T_{o,j}, \tilde{N}_i, \tilde{N}_j, L_i, L_j, f_i, f_j, p_i^{(k)}, p_j^{(k)}$ ,  $\forall k = 1, \dots, 8$ , αντίστοιχα. Οι περιορισμοί  $C_1$  και  $C_2$ , εκφράζουν το κάτω όριο στα bits που πρέπει να εκφορτωθούν κατά την περίοδο offloading,  $T_o$ , κάθε χρήστη και οι  $C_5, C_6$  εκφράζουν την μέγιστη δυνατή ενέργεια που μπορεί να δαπανηθεί απο κάθε χρήστη, συνολικά, τόσο για την εκφόρτωση όσο και για την τοπική επεξεργασία. Οι περιορισμοί  $C_3, C_4$  εξασφαλίζουν πώς ο αριθμός των bits που μπορούν να εκφορτωθούν,  $\tilde{N}_u$  και των bits που υποβάλλονται σε τοπική επεξεργασία,  $L_u$  καλύπτουν το υπολογιστικό φορτίο του sub-task,  $N_u$  κάθε χρήστη  $u$ . Τέλος, οι  $C_7, C_8$  εκφράζουν το άνω όριο στη συχνότητα λειτουργίας της CPU της συσκευής κάθε χρήστη και οι  $C_9$  εξασφαλίζουν πως οι μεταβλητές βελτιστοποίησης είναι φυσικά υλοποιήσιμες. Σύμφωνα με την ανάλυση που αφορά το full offloading, ο χρόνος εκφόρτωσης μπορεί να θεωρηθεί κοινός,  $T_{o,i} = T_{o,j} = T$  ενώ για την ελαχιστοποίηση του συνολικού χρόνου  $T_{PO}$  μπορούμε να λάβουμε το μέγιστο των τοπικών χρόνων επεξεργασίας  $T_{loc} = \max(T_{loc,i}, T_{loc,j})$  για να εκφράσουμε συνοπτικά την μετρική. Σύμφωνα με τις σχέσεις της ενότητας 2.2 και την μεταβλητή ποσοστών  $s_k$ :

$$\begin{aligned}
 & \min_{\mathbf{T}, \mathbf{p}, \tilde{\mathbf{N}}, \mathbf{L}, \mathbf{f}, \mathbf{s}} \max(T, T_{loc}) \\
 & \text{s.t.} \quad C_1 : T \sum_{k=1}^8 s_k R_i^{(k)} \geq \frac{\tilde{N}_i}{B}, \\
 & \quad C_2 : T \sum_{k=1}^8 s_k R_j^{(k)} \geq \frac{\tilde{N}_j}{B}, \\
 & \quad C_3 : \tilde{N}_i + L_i \geq N_i, \\
 & \quad C_4 : \tilde{N}_j + L_j \geq N_j, \\
 & \quad C_5 : T \sum_{k=1}^8 s_k P_i^{(k)} + k_i L_i X_i f_i^2 \leq E_i, \\
 & \quad C_6 : T \sum_{k=1}^8 s_k P_j^{(k)} + k_j L_j X_j f_j^2 \leq E_j, \\
 & \quad C_7 : f_i \leq f_{max}, \\
 & \quad C_8 : f_j \leq f_{max}, \\
 & \quad C_9 : T_{loc} = \max\left(\frac{L_i \cdot X_i}{f_i}, \frac{L_j \cdot X_j}{f_j}\right), \\
 & \quad C_{10} : T, T_{loc}, \tilde{N}_i, \tilde{N}_j, L_i, L_j, f_i, f_j, P_i^{(k)}, P_j^{(k)}, s_k \geq 0, \forall k = 1, \dots, 8
 \end{aligned} \tag{2.26}$$

Η διατύπωση αυτή παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς, μέσω των περιορισμών  $C_3, C_4$  στον αριθμό των bits και  $C_5, C_6$  στην ενεργειακή κατανάλωση, εμπλέκονται μεταβλητές εκφόρτωσης και τοπικής επεξεργασίας. Με την αύξηση της συχνότητας λειτουργίας της τοπικής CPU, η καθυστέρηση του τοπικού υπολογισμού μειώνεται, οδηγώντας και σε μείωση του αριθμού bits που πρέπει να εκφορτωθούν, με αποτέλεσμα την μείωση του χρόνου εκφόρτωσης, κάτω από συνθήκες παρεμβολής από άλλους χρήστες. Η μείωση αυτή συνεχίζεται μέχρι να η ταχύτητα επεξεργασίας να αυξήσει την ενεργειακή κατανάλωση σε μη επιτρεπτά επίπεδα. Η από-κοινού βελτιστοποίηση των τηλεπικοινωνιακών και υπολογιστικών πόρων αποσκοπεί στην εύρεση ικανοποιητικών λύσεων στο παραπάνω trade-off.

## Κεφάλαιο 3

### Επίλυση προβλημάτων

Στην παρακάτω ενότητα, θα παρουσιαστούν οι μεθοδολογίες επίλυσης των προβλημάτων βελτιστοποίησης που διατυπώθηκαν στο Κεφάλαιο 2 και θα καταστρωθούν προσομοιώσεις για την πιστοποίηση της εγκυρότητας των μεθόδων. Οι μεθοδολογίες θα βασιστούν σε τεχνικές από την θεωρία κυρτής βελτιστοποίησης (convex optimization) και συγκεκριμένα στο γεωμετρικό προγραμματισμό (geometric programming, στην βελτιστοποίηση Διαφοράς-Κυρτών συναρτήσεων (Difference-of-Convex, DC optimization), στην χρήση χαλάρωσης (relaxation) περιορισμών και προσθήκης μεταβλητών (slack variables). Η χρήση των τεχνικών αυτών αποσκοπεί στην διαχείριση των μη-κυρτών όρων που προκύπτουν στις διατυπώσεις και στην μετατροπή των προβλημάτων σε κυρτές διατυπώσεις, διαχειρίσιμες από εργαλεία βελτιστοποίησης του λογισμικού Matlab.

#### 3.1 Προτεινόμενες Λύσεις Full Offloading

##### 3.1.1 Επίλυση Energy constrained minimization

Σε συνέχεια της (2.22), για να μειώσουμε τον αριθμό των μεταβλητών και να αφαιρέσουμε τις συζευγμένες μεταβλητές  $s_k$  και  $T$ , ορίζουμε μία νέα μεταβλητή  $z_k = s_k \cdot T$ , η οποία εκφράζει το χρονικό διάστημα της υπό-φάσης  $k$  του time-sharing, κατά την οποία ισχύει μία από τις σειρές αποκωδικοποίησης που περιγράφεται στο μοντέλου του συστήματος. Επίσης, σύμφωνα με τις σχέσεις  $p_u^{(k)} = \frac{P_u^{(k)}}{NB_o}$  μπορούμε να εκφράσουμε τους περιορισμούς ενεργειακής κατανάλωσης ως προς το  $SNR = p_u$  για κάθε χρήστη  $u \in i, j$ . Επομένως το πρόβλημα



μπορεί να εκφραστεί ως:

$$\begin{aligned}
 \min_{T, \mathbf{p}, \mathbf{z}} \quad & T \\
 \text{s.t.} \quad & C_1 : z_1 BR_i^{(1)} + z_2 BR_i^{(2)} + \dots + z_7 BR_i^{(7)} + z_8 BR_i^{(8)} \geq \tilde{N}_i, \\
 & C_2 : z_1 BR_j^{(1)} + z_2 BR_j^{(2)} + \dots + z_7 BR_j^{(7)} + z_8 BR_j^{(8)} \geq \tilde{N}_j, \\
 & C_3 : \sum_{k=1}^8 z_k p_i^{(k)} \leq \frac{E_i}{NB_o}, \\
 & C_4 : \sum_{k=1}^8 z_k p_j^{(k)} \leq \frac{E_j}{NB_o}, \\
 & C_5 : T, z_k, p_i^{(k)}, p_j^{(k)} \geq 0, \\
 & C_6 : \sum_{k=1}^8 z_k \leq T. \forall k = 1, \dots, 8
 \end{aligned} \tag{3.1}$$

Οι μεταβλητές  $z_k$  και  $T$  δεν παρουσιάζουν σύζευξη στους περιορισμούς του προβλήματος και ο όρος  $\sum_{k=1}^8 z_k$  στο περιορισμό  $C_6$  αποτελεί αύξουσα συνάρτηση ως προς τις μεταβλητές  $z_k$ . Στην βέλτιστη κατάσταση, οποιαδήποτε μείωση στην τιμή των  $z_k$  θα οδηγούσε σε μείωση της τιμής  $T$  της αντικειμενικής συνάρτησης, λόγω της αύξουσας φύσης του αθροίσματος, και επομένως θα είχαμε μια καλύτερη τιμή ελαχίστου. Επομένως, στο πραγματικό ελάχιστο θα πρέπει ο περιορισμός να ισχύει ως ισότητα και άρα  $\sum_{k=1}^8 z_k = T$ . Στη συνέχεια θα προσπαθήσουμε να αφαιρέσουμε τους μη κυρτούς όρους από τις εκφράσεις των περιορισμών μέσω της εισαγωγής slack μεταβλητών[36]. Η εμφάνιση των όρων παρεμβολής  $p_j^{(k)}$  και  $p_i^{(k)}$  στις εκφράσεις των ρατες  $R_i^{(k)}$  και  $R_j^{(k)}$  αντίστοιχα, τις καθιστούν μη κυρτές ως προς τις μεταβλητές  $p_i^{(k)}$  και  $p_j^{(k)}$  μαζί. Εισάγουμε μεταβλητές  $r_i^{(k)}$  και  $r_j^{(k)}$  για τις οποίες ισχύουν οι εξής ανισότητες:

$$r_i^{(k)} \leq R_i^{(k)} \tag{3.2}$$

$$r_j^{(k)} \leq R_j^{(k)} \tag{3.3}$$

$$\forall k = 1, \dots, 8$$

όπου  $R_i^{(k)}$  και  $R_j^{(k)}$  δίνονται από τις σχέσεις (2.3), (2.4). Ενδεικτικά, θεωρούμαι τις σειρές αποκωδικοποίησης για  $k = 1 \dots 4$  κατά τις οποίες ο χρήστης  $i$  αποκωδικοποιείται πάντα πρώτος από το  $RRH_w$ , λαμβάνοντας το μήνυμα του χρήστη  $j$  ως παρεμβολή, ενώ ο χρήστης  $j$  αποκωδικοποιείται χωρίς παρεμβολή, μέσω SIC από το  $RRH_x$ . Το  $RRH_x$  μπορεί να είναι το ίδιο  $RRH$  στο οποίο έγινε η αποκωδικοποίηση του  $i$ , στις σειρές  $k \in 1, 3$  ή διαφορετικό στις σειρές  $k \in 2, 4$ . Οι παραπάνω ανισότητες μπορούν να γραφούν ως εξής:

$$r_i^{(k)} \leq \log \left( 1 + \frac{g_{iw} \cdot p_i^{(k)}}{1 + g_{jw} \cdot p_j^{(k)}} \right) \tag{3.4}$$

$$r_j^{(k)} \leq \log \left( 1 + g_{jx} \cdot p_j^{(k)} \right) \tag{3.5}$$

$$(w, x) \in \{a, b\}, \forall k = 1, \dots, 4$$

Εφαρμόζοντάς την αύξουσα συνάρτηση  $2^x$  και στα δύο μέλη των ανισοτήτων και φέρνο-

ντας όλους τους όρους στην αριστερή πλευρά, οι ανισότητες μπορούν να γραφούν ως:

$$\log_2 \left( \frac{1}{g_{iw}p_i^{(k)}} + \frac{g_{jw}p_j^{(k)}}{g_{iw}p_i^{(k)}} \right) + \log_2 (2^{r_i^{(k)}} - 1) \leq 0 \quad (3.6)$$

$$\log_2 (2^{r_j^{(k)}} - 1) - \log_2 (p_j^{(k)}) - \log_2 (g_{jx}) \leq 0 \quad (3.7)$$

$$(w, x) \in \{a, b\}, \forall k = 1, \dots, 4$$

Οι ίδιες ανισότητες ισχύουν για  $k = 5, \dots, 8$  με τους δείκτες  $i, j$  αντεστραμμένους καθώς οι σειρές αποκωδικοποίησης απαιτούν η αποκωδικοποίηση του χρήστη  $j$  να γίνεται πάντα πρώτη. Οι slack μεταβλητές διατηρούν την σχέση διάταξης των ανισοτήτων  $C_1, C_2$  και άρα, προσθέτοντας τις παραπάνω ανισότητες για τις  $r_i, j^{(k)}$ , το πρόβλημα διατυπώνεται ως εξής:

$$\begin{aligned} \min_{\mathbf{p}, \mathbf{z}, \mathbf{r}} \quad & \sum_{k=1}^8 z_k \\ \text{s.t.} \quad & C_1 : \sum_{k=1}^8 z_k r_i^{(k)} \geq \frac{\tilde{N}_i}{B}, \\ & C_2 : \sum_{k=1}^8 z_k r_j^{(k)} \geq \frac{\tilde{N}_j}{B}, \\ & C_3 : \sum_{k=1}^8 z_k p_i^{(k)} \leq \frac{E_i}{NB_o}, \\ & C_4 : \sum_{k=1}^8 z_k p_j^{(k)} \leq \frac{E_j}{NB_o}, \\ & C_6 : \log \left( \frac{1}{g_{iw}p_i^{(k)}} + \frac{g_{jw}p_j^{(k)}}{g_{iw}p_i^{(k)}} \right) + \log (2^{r_i^{(k)}} - 1) \leq 0, \\ & C_7 : \log (2^{r_j^{(k)}} - 1) - p_j^{(k)} - \log (g_{jx}) \leq 0, (w, x) \in \{a, b\}, \forall k = 1, \dots, 4 \\ & C_8 : \log \left( \frac{1}{g_{jw}p_j^{(k)}} + \frac{g_{iw}p_i^{(k)}}{g_{jw}p_j^{(k)}} \right) + \log (2^{r_j^{(k)}} - 1) \leq 0, \\ & C_9 : \log (2^{r_i^{(k)}} - 1) - p_i^{(k)} - \log (g_{ix}) \leq 0, (w, x) \in \{a, b\}, \forall k = 5, \dots, 8 \\ & C_{10} : r_k, z_k, p_i^{(k)}, p_j^{(k)} \geq 0, \forall k = 1, \dots, 8, \end{aligned} \quad (3.8)$$

Οι συζευγμένοι όροι  $z_k r_i^{(k)}$  και  $z_k r_j^{(k)}$  που εμφανίζονται στους περιορισμούς  $C_1, C_2$ , οι όροι  $z_k p_i^{(k)}, z_k p_j^{(k)}$  στους  $C_3, C_4$ , οι όροι  $\frac{1}{p_i^{(k)}}, \frac{p_j^{(k)}}{p_i^{(k)}}$  στον περιορισμό  $C_6$  και οι όροι  $\frac{1}{p_j^{(k)}}, \frac{p_i^{(k)}}{p_j^{(k)}}$  στον περιορισμό  $C_8$ , αποτελούν μονώνυμα (monomials) των μεταβλητών βελτιστοποίησης. Αυτός ο τύπος μη-κυρτών όρων παρακινεί την χρήση γεωμετρικού προγραμματισμού [37] και συγκεκριμένα, ενός μετασχηματισμού της μορφής  $x = e^{\tilde{x}}$  σε κάθε μεταβλητή των διανυσμάτων  $\mathbf{p}, \mathbf{z}, \mathbf{r}$ . Λογαριθμίζοντας κάθε πλευρά των ανισοτήτων  $C_1$  έως  $C_4$  και της αντικειμενικής

συνάρτησης, το πρόβλημα βελτιστοποίησης μπορεί να διατυπωθεί ως εξής:

$$\begin{aligned}
 \min_{\tilde{\mathbf{p}}, \tilde{\mathbf{z}}, \tilde{\mathbf{r}}} \quad & \log \left( \sum_{k=1}^8 e^{\tilde{z}_k} \right) \\
 \text{s.t.} \quad & C_1 : \log \left( \frac{\tilde{N}_i}{B} \right) - \log \left( \sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_i^{(k)}} \right) \leq 0 \\
 & C_2 : \log \left( \frac{\tilde{N}_j}{B} \right) - \log \left( \sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_j^{(k)}} \right) \leq 0 \\
 & C_3 : \log \left( \sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_i^{(k)}} \right) - \log \left( \frac{E_i}{BN_0} \right) \leq 0, \\
 & C_4 : \log \left( \sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_j^{(k)}} \right) - \log \left( \frac{E_j}{BN_0} \right) \leq 0, \\
 & C_5 : \log \left( \frac{1}{g_{ik}} e^{-\tilde{p}_i^{(k)}} + \frac{g_{jk}}{g_{ik}} e^{\tilde{p}_j^{(k)} - \tilde{p}_i^{(k)}} \right) + \log (2e^{\tilde{r}_i^{(k)}} - 1) \leq 0, \forall k = 1, \dots, 4 \\
 & C_6 : \log (2e^{\tilde{p}_j^{(k)}} - 1) - \tilde{p}_j^{(k)} - \log g_{jk} \leq 0, \forall k = 1, \dots, 4 \\
 & C_7 : \log \left( \frac{1}{g_{jk}} e^{-\tilde{p}_j^{(k)}} + \frac{g_{ik}}{g_{jk}} e^{\tilde{p}_i^{(k)} - \tilde{p}_j^{(k)}} \right) + \log (2e^{\tilde{r}_j^{(k)}} - 1) \leq 0, \forall k = 5, \dots, 8 \\
 & C_8 : \log (2e^{\tilde{p}_i^{(k)}} - 1) - \tilde{p}_i^{(k)} - \log g_{ik} \leq 0, \forall k = 5, \dots, 8
 \end{aligned} \tag{3.9}$$

Στην παραπάνω διατύπωση, οι περιορισμοί  $C_5$  έως  $C_8$  ξαναγράφονται με το φυσικό λογάριθμο ο οποίος διαφέρει από τον λογάριθμο με βάση το 2 μέσω μιας θετικής σταθεράς, η οποία δεν επηρεάζει την μορφή των ανισοτήτων. Επίσης, οι νέες μεταβλητές βελτιστοποίησης, ως εκθέτες του γεωμετρικού μετασχηματισμού λαμβάνουν τιμές σε όλο το  $\mathbf{R}$ . Στο συγκεκριμένο σημείο, πρέπει να διερευνηθεί η κυρτότητα του προβλήματος (3.9). Οι περιορισμοί

$C_1$  και  $C_4$  αποτελούνται από την πρόσθεση μίας log-sum-exp συνάρτησης των μεταβλητών  $\tilde{z}_k, \tilde{p}_i^{(k)}, \tilde{p}_j^{(k)}$ , η οποία είναι κυρτή κατά [36] και μια σταθεράς  $-\log \left( \frac{E}{BN_0} \right)$  η οποία δεν επηρεάζει την κυρτότητα. Επομένως, οι περιορισμοί αυτοί είναι κυρτοί ως προς τις μεταβλητές βελτιστοποίησης. Αντίστοιχα, η αντικειμενική συνάρτηση είναι κυρτή ως log-sum-exp συνάρτηση των μεταβλητών  $z_k$ . Η κυρτότητα των περιορισμών  $C_5$  έως  $C_8$  καλύπτεται από τις παρακάτω προτάσεις:

**Λεμμα 1.** Η συνάρτηση  $g(p_1, p_2, r) = \log \left( \frac{1}{g_1} e^{-p_1} + \frac{g_2}{g_1} e^{p_2 - p_1} \right) + \log (2e^r - 1)$  είναι κυρτή.

*Απόδειξη.* Ο πρώτος όρος  $g_1(p_1, p_2) = \log \left( \frac{1}{g_1} e^{-p_1} + \frac{g_2}{g_1} e^{p_2 - p_1} \right)$  μπορεί να γραφτεί ως εξής, χρησιμοποιώντας την σχέση  $x = e^{\log x}$  και τις ιδιότητες λογαρίθμων για τις σταθερές  $\frac{1}{g_1}$  και  $\frac{g_2}{g_1}$ :

$$g_1 = \log(e^{-p_1 - \log g_1} + e^{p_2 - p_1 + \log g_2 - \log g_1})$$

Κάθε ένας από τους εκθέτες,  $-p_1 - \log g_1$  και  $p_2 - p_1 + \log g_2 - \log g_1$  της έκφρασης αποτελεί μια affine συνάρτηση των μεταβλητών  $p_1$  και  $p_2$ . Η συνάρτηση  $g_1$  αποτελεί, λοιπόν, σύνθεση μιας log-sum-exp συνάρτησης,  $h(x_1, x_2) = \log(e^{x_1} + e^{x_2})$  και δύο affine συναρτήσεων,  $a_1(p_1) = -p_1 - \log g_1$  και  $a_2(p_1, p_2) = p_2 - p_1 + \log g_2 - \log g_1$ :

$$g_1(p_1, p_2) = h(a_1(p_1), a_2(p_1, p_2))$$

Οι log-sum-exp και affine συναρτήσεις είναι κυρτές[36] και άρα, η  $g_1$ , ως σύνθεση κυρτών συναρτήσεων, είναι και αυτή κυρτή.

Ο δεύτερος όρος  $g_2 = \log(2^{e^r} - 1)$  είναι κυρτή συνάρτηση της  $r$ . Αυτό αποδεικνύεται ελέγχοντας το πρόσημο της δεύτερης παραγώγου:

$$\frac{\partial^2 f}{\partial r^2} = \frac{2^z z \ln(2)(2^z - z \ln 2 - 1)}{(2^z - 1)^2}$$

όπου  $z = e^r$ . Ο όρος  $2^z z \log$  είναι θετικός καθώς  $z = e^r \geq 0$  και  $2^{e^r} \geq 2^0 = 1$  για κάθε  $r \in \mathbb{R}$ . Ο όρος  $(2^{e^r} - 1)^2$  είναι προφανώς μεγαλύτερος ή ίσος του μηδενός. Η συνάρτηση  $2^{e^r} - e^r \ln(2) - 1$  έχει παράγωγο  $(2^{e^r} - e^r)e^r \ln(2)$ , η οποία σύμφωνα με τις παραπάνω σχέσεις είναι και αυτή μεγαλύτερη ή ίση του μηδενός. Συνολικά, για τη δεύτερη παράγωγο ισχύει πως:

$$\frac{\partial^2 f}{\partial r^2} = \frac{2^z z \ln(2)(2^z - z \ln 2 - 1)}{(2^z - 1)^2} \geq 0$$

και άρα ο δεύτερος όρος  $g_2$  είναι κυρτός σε όλο το  $\mathbf{R}$ . Τέλος, η συνάρτηση  $g$  ως πρόσθεση κυρτών συναρτήσεων  $g_1$  και  $g_2$  είναι κυρτή.  $\square$

Οι περιορισμοί  $C_5, C_7$  μπορούν να γραφούν ως  $g(\tilde{p}_i^{(k)}, \tilde{p}_j^{(k)}, \tilde{r}_i^{(k)})$  και  $g(\tilde{p}_i^{(k)}, \tilde{p}_j^{(k)}, \tilde{r}_i^{(k)})$ , αντίστοιχα και άρα, σύμφωνα με το Λήμμα 1, αποτελούν κυρτές συναρτήσεις των μεταβλητών. Οι περιορισμοί  $C_6, C_6$  αποτελούν την πρόσθεση μίας affine συνάρτησης,  $-\tilde{p}_j^{(k)} - \log(g_{jk})$  και  $-\tilde{p}_i^{(k)} - \log(g_{ik})$  αντίστοιχα και μιας συνάρτησης  $g_2(\tilde{p}_j^{(k)})$ ,  $g_2(\tilde{p}_i^{(k)})$  κατά ακολουθία με το Λήμμα 1. Οι συναρτήσεις αυτές είναι κυρτές και άρα, οι περιορισμοί είναι κυρτοί ως προς τις μεταβλητές βελτιστοποίησης σε όλο το  $\mathbf{R}$ . Παρότι δείξαμε πως οι περιορισμοί  $C_3$  έως  $C_8$  είναι κυρτοί, οι περιορισμοί  $C_1$  και  $C_2$  παραμένουν μη-κυρτοί μετά τον γεωμετρικό μετασχηματισμό, καθώς αποτελούν άθροισμα ενός σταθερού όρου  $\log(\frac{\tilde{N}}{B})$  και ενός κοίλου  $-\log(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_i^{(k)}})$ .

Αναγνωρίζοντας τους περιορισμούς αυτούς ως τη διαφορά κυρτών συναρτήσεων (Difference-of-Convex functions) και συγκεκριμένα τη διαφορά μιας σταθερής και άρα και κυρτής συνάρτησης και μιας κυρτής log-sum-exp συνάρτησης των μεταβλητών  $z_k, r^{(k)}$ , το πρόβλημα μπορεί να έρθει σε επιλύσιμη μορφή μέσω μιας διαδικασίας διαδοχικών κυρτών προσεγγίσεων[38]. Σύμφωνα με τη μεθοδολογία του DC προγραμματισμού, ο δεύτερος όρος της διαφοράς μεταξύ κυρτών συναρτήσεων μπορεί να αντικατασταθεί με μια γραμμική προσέγγιση γύρω από ένα σημείο  $x_0$  και συγκεκριμένα, την προσέγγιση Taylor πολλών μεταβλητών ώστε να προσεγγιστεί επαναληπτικά το βέλτιστο (globally optimal) σημείο του προβλήματος βελτιστοποίησης. Έτσι, ορίζοντας το διάνυσμα  $\mathbf{x}_l = [\mathbf{z}, \mathbf{r}_l]$ ,  $\forall l \in \{i, j\}$  ως τη συνένωση των διανυσμάτων  $\mathbf{z}$  και  $\mathbf{r}_l$  ή  $\mathbf{r}_1$ , αντίστοιχα, οι δεύτεροι όροι μπορούν να γραφούν ως:

$$h_l(\mathbf{x}_l) = \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_l^{(k)}}\right), \forall l \in \{i, j\}$$

και η προσέγγιση γύρω από το σημείο  $x_0$  ως:

$$h_l(\mathbf{x}) \approx h_l(\mathbf{x}_{l_0}) + \nabla h_l(\mathbf{x}_{l_0})(\mathbf{x}_l - \mathbf{x}_{l_0}), \forall l \in \{i, j\} \quad (3.10)$$

όπου ισχύουν οι παρακάτω σχέσεις,

$$\begin{aligned} \nabla h_l(\mathbf{x}_{l_0}) &= [\mathbf{e}_z, \mathbf{e}_{r_l}], \forall l \in \{i, j\} \\ \text{s.t. } \mathbf{e}_z(k) &= \frac{\partial h}{\partial z_k}, \\ \mathbf{e}_{r_l}(k) &= \frac{\partial h_l}{\partial r_{l_k}}, \forall k = 1, \dots, 8 \\ \frac{\partial h_l}{\partial z_k} &= \frac{\partial h_l}{\partial r_{l_k}} = \frac{e^{\tilde{z}_k + \tilde{r}_l^{(k)}}}{\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_l^{(k)}}} \end{aligned}$$

Τα διανύσματα  $\mathbf{e}_z, \mathbf{e}_{r_l}$  περιέχουν τις μερικές παραγώγους ως προς τις μεταβλητές  $z_k$  και  $r_{l_k}, \forall l \in \{i, j\}$ . Αντικαθιστώντας τις προσεγγίσεις στους περιορισμούς  $C_1$  και  $C_2$ , το μη-κυρτό πρόβλημα αντικαθίσταται με μια κυρτή προσέγγιση της μορφής:

$$\begin{aligned} \min_{\mathbf{p}, \mathbf{x}_i, \mathbf{x}_j} \quad & \log \sum_{k=1}^8 e^{\tilde{z}_k} \\ \text{s.t.} \quad & C_1 : \log \left( \frac{\tilde{N}_i}{B} \right) - h_i(\mathbf{x}_{i_0}) - \nabla h_i(\mathbf{x}_{i_0})(\mathbf{x}_i - \mathbf{x}_{i_0}) \leq 0 \\ & C_2 : \log \left( \frac{\tilde{N}_j}{B} \right) - h_j(\mathbf{x}_{j_0}) - \nabla h_j(\mathbf{x}_{j_0})(\mathbf{x}_j - \mathbf{x}_{j_0}) \leq 0 \\ & C_3 : \log \left( \sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_i^{(k)}} \right) - \log \left( \frac{E_i}{BN_0} \right) \leq 0, \\ & C_4 : \log \left( \sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_j^{(k)}} \right) - \log \left( \frac{E_j}{BN_0} \right) \leq 0, \\ & C_5 : \log \left( \frac{1}{g_{ik}} e^{-\tilde{p}_i^{(k)}} + \frac{g_{jk}}{g_{ik}} e^{\tilde{p}_j^{(k)} - \tilde{p}_i^{(k)}} \right) + \log (2^{e^{\tilde{r}_i^{(k)}}} - 1) \leq 0, \forall k = 1, \dots, 4 \\ & C_6 : \log (2^{e^{\tilde{p}_j^{(k)}}} - 1) - \tilde{p}_j^{(k)} - \log g_{jk} \leq 0, \forall k = 1, \dots, 4 \\ & C_7 : \log \left( \frac{1}{g_{jk}} e^{-\tilde{p}_j^{(k)}} + \frac{g_{ik}}{g_{jk}} e^{\tilde{p}_i^{(k)} - \tilde{p}_j^{(k)}} \right) + \log (2^{e^{\tilde{r}_j^{(k)}}} - 1) \leq 0, \forall k = 5, \dots, 8 \\ & C_8 : \log (2^{e^{\tilde{p}_i^{(k)}}} - 1) - \tilde{p}_i^{(k)} - \log g_{ik} \leq 0, \forall k = 5, \dots, 8 \end{aligned} \tag{3.11}$$

Παρακάτω, αναπτύσσεται ένας απλό αλγόριθμος που προσεγγίζει τη λύση του αρχικού μη-κυρτού προβλήματος με μια ακολουθία προβλημάτων κυρτής βελτιστοποίησης, τα οποία μπορούν να λυθούν, σε κάθε επανάληψη, με τη μέθοδο interior point, σε πολυωνυμικό χρόνο[36]. Επιπλέον, το εσωτερικό του βρόχου έχει γραμμικό ρυθμό σύγκλισης και ο αλγόριθμος μπορεί να υλοποιηθεί ρεαλιστικά σε ένα πρακτικό σενάριο MEC, σε έναν διακομιστή άκρης(edge) με επαρκείς υπολογιστικές δυνατότητες. Στον παρακάτω αλγόριθμο, ως  $e^z$  και  $\sum \mathbf{z}$  ορίζεται το ύψωμα κάθε συνιστώσας του διανύσματος  $\mathbf{z}$ ,  $e^z = [e^{z_1}, \dots, e^{z_8}]$ , και το άθροισμα των συνιστωσών,  $\sum_{k=1}^8 z_k$ . Η μέθοδος επίλυσης αυτού του προβλήματος θα χρησιμοποιηθεί ως πρότυπο για την επίλυση των προβλημάτων στις παρακάτω ενότητες.

---

**Algorithm 1** Solution of delay minimization problem (25)
 

---

**Initialize:**  $A, e, \tilde{\mathbf{z}}, \tilde{\mathbf{r}}_i, \tilde{\mathbf{r}}_j$ **while**  $A > e$  **do****Solve:** optimization problem (3.11) with  $\mathbf{z}_0, \mathbf{r}_i \mathbf{0}, \mathbf{r}_j \mathbf{0}$ 

$$A = \|\tilde{\mathbf{z}}_0, \tilde{\mathbf{r}}_{i0}, \tilde{\mathbf{r}}_{j0} - [\mathbf{z}^*, \mathbf{r}_i^*, \tilde{\mathbf{r}}_j]\|_2^2$$

$$\tilde{\mathbf{z}}_0 \leftarrow \mathbf{z}^*$$

$$\tilde{\mathbf{r}}_{i0} \leftarrow \mathbf{r}_i^*$$

$$\tilde{\mathbf{r}}_{j0} \leftarrow \mathbf{r}_j^*$$

**end while**

$$\mathbf{z}^* \leftarrow e^{\mathbf{z}^*}$$

$$\mathbf{T}^* \leftarrow \sum \mathbf{z}^*$$

**Output:**  $\mathbf{T}^*$ 

### 3.1.2 Επίλυση Delay constrained minimization

Το πρόβλημα της ελαχιστοποίησης της ενεργειακής κατανάλωσης επιλύεται ακολουθώντας τη συλλογιστική του προβλήματος της προηγούμενης ενότητας. Συνοπτικά τα βήματα προς την επιλύσιμη μορφή είναι τα εξής:

1. Εισάγουμε τις μεταβλητές χρονικού διαστήματος  $z_k = s_k \cdot T, \forall k = 1 \dots 8$  για να αφαιρέσουμε τη σύζευξη των  $s_k$  και  $T$ . Η χρονική περίοδος  $T$  εκφράζεται ως  $\sum_{k=1}^8 z_k$ .
2. Εισάγουμε slack μεταβλητές  $r_l^{(k)}$  για τις οποίες ισχύει  $r_l^{(k)} \leq R_l^{(k)} \forall l \in \{i, j\}, k = 1 \dots 8$ . Εισάγουμε τις νέες αυτές ανισότητες ως περιορισμούς, εκφρασμένες ως προς τις μεταβλητές  $r_l^{(k)}$  και  $p_l^{(k)} \forall l \in \{i, j\}, k = 1 \dots 8$ .
3. Εκτελούμε το γεωμετρικό μετασχηματισμό  $x = e^{\tilde{x}}$  στις μεταβλητές  $\mathbf{z}, \mathbf{p}, \mathbf{r}$  και λογαριθμίζουμε την αντικειμενική συνάρτηση και τους περιορισμούς του προβλήματος.
4. Αντικαθιστούμε τις ανισότητες που παρουσιάζουν δομή DC με την πρώτη προσέγγιση Taylor πολλών μεταβλητών και λύνουμε προοδευτικά το πρόβλημα, μέσω ενός επαναληπτικού αλγορίθμου.

Οι περιορισμοί στον αριθμό των bits που πρέπει να εκφορτωθούν είναι κοινοί με το πρόβλημα ελαχιστοποίησης της καθυστέρησης και άρα, τους διαχειριζόμαστε με τον ίδιο τρόπο.

$$\begin{aligned}
 \min_{\mathbf{p}, \mathbf{x}_i, \mathbf{x}_j} \quad & \log(w_i \sum_{k=1}^8 e^{\tilde{p}_i^{(k)} + \tilde{z}_k} + w_j \sum_{k=1}^8 e^{\tilde{p}_j^{(k)} + \tilde{z}_k}) \\
 \text{s.t.} \quad & C_1 : \log\left(\frac{\tilde{N}_i}{B}\right) - h_i(\mathbf{x}_{i0}) - \nabla h_i(\mathbf{x}_{i0})(\mathbf{x}_i - \mathbf{x}_{i0}) \leq 0 \\
 & C_2 : \log\left(\frac{\tilde{N}_j}{B}\right) - h_j(\mathbf{x}_{j0}) - \nabla h_j(\mathbf{x}_{j0})(\mathbf{x}_j - \mathbf{x}_{j0}) \leq 0 \\
 & C_3 : \log\left(\sum_{k=1}^8 e^{\tilde{z}_k}\right) - \log(T_{max}) \leq 0, \\
 & C_4 : \log\left(\frac{1}{g_{ik}} e^{-\tilde{p}_i^{(k)}} + \frac{g_{jk}}{g_{ik}} e^{\tilde{p}_j^{(k)} - \tilde{p}_i^{(k)}}\right) + \log(2^{e^{\tilde{r}_i^{(k)}}} - 1) \leq 0, \forall k = 1, \dots, 4 \\
 & C_5 : \log(2^{e^{\tilde{p}_j^{(k)}}} - 1) - \tilde{p}_j^{(k)} - \log g_{jk} \leq 0, \forall k = 1, \dots, 4 \\
 & C_6 : \log\left(\frac{1}{g_{jk}} e^{-\tilde{p}_j^{(k)}} + \frac{g_{ik}}{g_{jk}} e^{\tilde{p}_i^{(k)} - \tilde{p}_j^{(k)}}\right) + \log(2^{e^{\tilde{r}_j^{(k)}}} - 1) \leq 0, \forall k = 5, \dots, 8 \\
 & C_7 : \log(2^{e^{\tilde{p}_i^{(k)}}} - 1) - \tilde{p}_i^{(k)} - \log g_{ik} \leq 0, \forall k = 5, \dots, 8
 \end{aligned} \tag{3.12}$$

όπου  $\tilde{\mathbf{x}}_l = [\tilde{\mathbf{z}}, \tilde{\mathbf{r}}_l], \forall l \in \{i, j\}$ . Μετά τον γεωμετρικό μετασχηματισμό, η αντικειμενική συνάρτηση είναι κυρτή ως σύνθεση affine συναρτήσεων των μεταβλητών  $p_i^{(k)}, p_j^{(k)}, z_k$  και μιας log-sum-exp συνάρτησης, ο οποίες είναι κυρτές[36]. Επιπλέον, ο περιορισμός  $C_3$  είναι κυρτός ως άθροισμα ενός σταθερού όρου,  $-\log(T_{max})$ , και μιας log-sum-exp συνάρτησης των μεταβλητών  $\tilde{z}_k$ .

Το παραπάνω πρόβλημα μπορεί να λυθεί επαναληπτικά, μέσω ενός απλού αλγορίθμου ο οποίος λύνει την κυρτή προσέγγιση χρησιμοποιώντας το βέλτιστο σημείο της προηγούμενης επανάληψης ως το αρχικό σημείο γύρω από το οποίο γίνεται η προσέγγιση των μη-κυρτών περιορισμών, μέχρι οι λύσεις κάθε επανάληψης να είναι επαρκώς κοντά.

---

**Algorithm 2** Solution of energy minimization problem (25)

---

**Initialize:**  $A, e, \tilde{\mathbf{z}}, \tilde{\mathbf{r}}_i, \tilde{\mathbf{r}}_j, \mathbf{w}_i, \mathbf{w}_j$

**while**  $A > e$  **do**

**Solve:** optimization problem (3.12) with  $\mathbf{z}_0, \mathbf{r}_{i0}, \mathbf{r}_{j0}$

$A = \|\tilde{\mathbf{z}}_0, \tilde{\mathbf{r}}_{i0}, \tilde{\mathbf{r}}_{j0} - [\mathbf{z}^*, \mathbf{r}_i^*, \mathbf{r}_j^*]\|_2^2$

$\tilde{\mathbf{z}}_0 \leftarrow \mathbf{z}^*$

$\tilde{\mathbf{r}}_{i0} \leftarrow \mathbf{r}_i^*$

$\tilde{\mathbf{r}}_{j0} \leftarrow \mathbf{r}_j^*$

**end while**

$\mathbf{E}_{\text{PO}i}^* \leftarrow e^{\tilde{\mathbf{z}}^* + \tilde{\mathbf{p}}_i^*}$

$\mathbf{E}_{\text{PO}j}^* \leftarrow e^{\tilde{\mathbf{z}}^* + \tilde{\mathbf{p}}_j^*}$

$\mathbf{E}_{\text{PO}}^* \leftarrow \sum \mathbf{w}_i \mathbf{E}_{\text{PO}i}^* + \mathbf{w}_j \mathbf{E}_{\text{PO}j}^*$

**Output:**  $\mathbf{E}_{\text{PO}}^*$

---

## 3.2 Προτεινόμενη Λύση Partial Offloading

Για την περίπτωση της μερικής εκφόρτωσης, θα ακολουθήσουμε τα βήματα εισαγωγής μεταβλητών και μετασχηματισμών που ακολουθήθηκαν στις παραπάνω μεθόδους επίλυσης, λαμβάνοντας υπόψη τους νέους μη-γραμμικούς όρους που εισάγει η μοντελοποίηση της τοπικής επεξεργασίας. Συγκεκριμένα, η διατύπωση (2.26) περιλαμβάνει τις μη κυρτές συναρτήσεις  $\max$  στην αντικειμενική συνάρτηση και στον περιορισμό ισότητας  $C_9$ . Εισάγοντας μεταβλη-

τές  $y$ ,  $u$  και 'χαλαρώνοντας' τον περιορισμό ισότητας, το πρόβλημα (2.26) γράφεται ως:

$$\begin{aligned}
 & \min_{\mathbf{T}, \mathbf{p}, \tilde{\mathbf{N}}, \mathbf{L}, \mathbf{f}, \mathbf{s}} && y \\
 & \text{s.t.} && C_1 : s_1 TBR_i^{(1)} + s_2 TBR_i^{(2)} + \dots + s_7 TBR_i^{(7)} + s_8 TBR_i^{(8)} \geq \tilde{N}_i, \\
 & && C_2 : s_1 TBR_j^{(1)} + s_2 TBR_j^{(2)} + \dots + s_7 TBR_j^{(7)} + s_8 TBR_j^{(8)} \geq \tilde{N}_j, \\
 & && C_3 : \tilde{N}_i + L_i \geq N_i, \\
 & && C_4 : \tilde{N}_j + L_j \geq N_j, \\
 & && C_5 : T \sum_{k=1}^8 s_k P_i^{(k)} + k_i L_i X_i f_i^2 \leq E_i, \\
 & && C_6 : T \sum_{k=1}^8 s_k P_j^{(k)} + k_j L_j X_j f_j^2 \leq E_j, \\
 & && C_7 : f_i \leq f_{max}, \\
 & && C_8 : f_j \leq f_{max}, \\
 & && C_9 : T \leq y, \\
 & && C_{10} : u \leq y, \\
 & && C_{11} : \frac{L_i \cdot X_i}{f_i} \leq u, \\
 & && C_{12} : \frac{L_j \cdot X_j}{f_j} \leq u, \\
 & && C_{13} : y, T, u, \tilde{N}_i, \tilde{N}_j, L_i, L_j, f_i, f_j, P_i^{(k)}, P_j^{(k)}, s_k \geq 0, \forall k = 1, \dots, 8
 \end{aligned} \tag{3.13}$$

Παρατηρούμε πώς οι μεταβλητές  $N_i$ ,  $N_j$  και  $u$  εμφανίζονται και στις δύο πλευρές ανισοτήτων των περιορισμών και άρα μπορούμε να τις αφαιρέσουμε και να συγχωνεύσουμε τους περιορισμούς, διατηρώντας την διάταξη των μεταβλητών. Εισάγοντας τις μεταβλητές χρονικών διαστημάτων,  $z_k = s_k T$ , και τις μεταβλητές SNR,  $p_l^{(k)}$  το πρόβλημα μπορεί να γραφεί ως:

$$\begin{aligned}
 & \min_{\mathbf{T}, \mathbf{p}, \tilde{\mathbf{N}}, \mathbf{L}, \mathbf{f}, \mathbf{z}} && y \\
 & \text{s.t.} && C_1 : \sum_{k=1}^8 z_k R_i^{(k)} + \frac{L_i}{B} \geq \frac{N_i}{B}, \\
 & && C_2 : \sum_{k=1}^8 z_k R_j^{(k)} + \frac{L_j}{B} \geq \frac{N_j}{B}, \\
 & && C_3 : \sum_{k=1}^8 z_k p_i^{(k)} + \frac{k_i L_i X_i f_i^2}{NB_o} \leq E_i / NB_o, \\
 & && C_4 : \sum_{k=1}^8 z_k p_j^{(k)} + \frac{k_j L_j X_j f_j^2}{NB_o} \leq E_j / NB_o, \\
 & && C_5 : f_i \leq f_{max}, \\
 & && C_6 : f_j \leq f_{max}, \\
 & && C_7 : \sum_{k=1}^8 z_k \leq y, \\
 & && C_8 : \frac{L_i \cdot X_i}{f_i} \leq y, \\
 & && C_9 : \frac{L_j \cdot X_j}{f_j} \leq y, \\
 & && C_{10} : y, T, u, L_i, L_j, f_i, f_j, p_i^{(k)}, p_j^{(k)} \geq 0, \forall k = 1, \dots, 8
 \end{aligned} \tag{3.14}$$

Η παραπάνω διατύπωση, χωρίς την ύπαρξη του  $\max$  όρου στην αντικειμενική συνάρτηση, αποτελεί την επιγραφική (epigraph) μορφή του προβλήματος.

Στη συνέχεια, εισάγουμε slack μεταβλητές  $r_l^{(k)} \leq R_l^{(k)}, \forall l \in \{i, j\}$ , και όπως στις παραπάνω μεθόδους βελτιστοποίησης, προσθέτουμε ανισώσεις εκφρασμένες ως προς τις μεταβλη-



τές  $r_k, p_i, p_j$ . Οι όροι  $\frac{L_i}{B}, \frac{L_j}{B}$  στους περιορισμούς  $C_1, C_2$ , οι όροι  $L_i f_i^2, L_j f_j^2$  στους περιορισμούς  $C_3, C_4$  και οι όροι  $\frac{L_i \cdot X_i}{f_i}, \frac{L_j \cdot X_j}{f_j}$  στους περιορισμούς  $C_9, C_{10}$  αποτελούν monomials των μεταβλητών βελτιστοποίησης. Ο γεωμετρικός μετασχηματισμός όλων των μεταβλητών του προβλήματος βοηθά στην διαχείριση των μη-κυρτών monomials που προκύπτουν στους περιορισμούς αυτούς, καθώς και των περιορισμών που παραμένουν κοινοί με τα προβλήματα της παραπάνω ενότητας. Λογαριθμίζοντας την αντικειμενική συνάρτηση και τους περιορισμούς το πρόβλημα λαμβάνει την εξής μορφή:

$$\begin{aligned}
& \min_{\tilde{\mathbf{z}}, \tilde{\mathbf{r}}, \tilde{\mathbf{p}}, \tilde{\mathbf{L}}, \tilde{\mathbf{f}}, \tilde{y}} \quad \tilde{y} \\
& \text{s.t.} \quad C_1 : \log\left(\frac{N_i}{B}\right) - \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_i^{(k)}} + \frac{1}{B} e^{\tilde{L}_i}\right) \leq 0, \\
& \quad C_2 : \log\left(\frac{N_j}{B}\right) - \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_j^{(k)}} + \frac{1}{B} e^{\tilde{L}_j}\right) \leq 0, \\
& \quad C_3 : \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_i^{(k)}} + \frac{k_i X_i}{N B_o} e^{\tilde{L}_i + 2\tilde{f}_i}\right) - \log\left(\frac{E_i}{N B_o}\right) \leq 0, \\
& \quad C_4 : \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_j^{(k)}} + \frac{k_j X_j}{N B_o} e^{\tilde{L}_j + 2\tilde{f}_j}\right) - \log\left(\frac{E_j}{N B_o}\right) \leq 0, \\
& \quad C_5 : \tilde{f}_i - \log(f_{max}) \leq 0, \\
& \quad C_6 : \tilde{f}_j - \log(f_{max}) \leq 0, \\
& \quad C_7 : \log\left(\sum_{k=1}^8 e^{\tilde{z}_k}\right) - \tilde{y} \leq 0, \\
& \quad C_8 : \tilde{L}_i - \tilde{f}_i + \log X_i - \tilde{y} \leq 0, \\
& \quad C_9 : \tilde{L}_j - \tilde{f}_j + \log X_j - \tilde{y} \leq 0, \\
& \quad C_{10} : \log\left(\frac{1}{g_{ik}} e^{-\tilde{p}_i^{(k)}} + \frac{g_{jk}}{g_{ik}} e^{\tilde{p}_j^{(k)} - \tilde{p}_i^{(k)}}\right) + \log(2^{e^{\tilde{r}_i^{(k)}}} - 1) \leq 0, \forall k = 1, \dots, 4 \\
& \quad C_{11} : \log(2^{e^{\tilde{p}_j^{(k)}}} - 1) - \tilde{p}_j^{(k)} - \log g_{jk} \leq 0, \forall k = 1, \dots, 4 \\
& \quad C_{12} : \log\left(\frac{1}{g_{jk}} e^{-\tilde{p}_j^{(k)}} + \frac{g_{ik}}{g_{jk}} e^{\tilde{p}_i^{(k)} - \tilde{p}_j^{(k)}}\right) + \log(2^{e^{\tilde{r}_j^{(k)}}} - 1) \leq 0, \forall k = 5, \dots, 8 \\
& \quad C_{13} : \log(2^{e^{\tilde{p}_i^{(k)}}} - 1) - \tilde{p}_i^{(k)} - \log g_{ik} \leq 0, \forall k = 5, \dots, 8
\end{aligned} \tag{3.15}$$

Ο περιορισμός  $C_7$  είναι κυρτός ως άθροισμα μιας γραμμικής συνάρτησης της επιγραφικής μεταβλητής,  $-\tilde{y}$ , και μιας log-sum-exp συνάρτησης των μεταβλητών  $\tilde{z}_k$ . Οι περιορισμοί  $C_5, C_6$  και  $C_8, C_9$  είναι κυρτοί ως affine συναρτήσεις των μεταβλητών  $\tilde{L}, \tilde{f}$  ενώ η κυρτότητα των  $C_{10}$  έως  $C_{13}$  έχει αποδειχθεί στις παραπάνω ενότητες. Οι περιορισμοί  $C_3, C_4$  έχουν την ίδια μορφή με τους αντίστοιχους περιορισμούς του προβλήματος delay minimization με την διαφορά ότι στο εσωτερικό του λογαρίθμου του πρώτου όρου προστίθεται μια εκθετική συνάρτηση,  $e^{\tilde{L}_l + 2\tilde{f}_l + \log(\frac{k_l X_l}{N B_o})} \forall l \in \{i, j\}$ . Η συνάρτηση αυτή αποτελεί σύνθεση μίας εκθετικής συνάρτησης και μια affine συνάρτησης,  $\tilde{L}_l + 2\tilde{f}_l + \log(\frac{k_l X_l}{N B_o}) \forall l \in \{i, j\}$ , και επομένως δεν επηρεάζει την κυρτότητα του όρου log-sum-exp κάθε περιορισμού. Τέλος, με ανάλογη λογική, οι όροι  $e^{\tilde{L}_l - \log B}$  στο εσωτερικό του δεύτερου, λογαριθμικού όρου δεν επηρεάζει την κυρτότητα του.

Οι  $C_1, C_2$  αναγνωρίζονται ως Difference of Convex και ακολουθείται η ίδια μεθοδολογία για την προσέγγισή τους με τις παραπάνω ενότητες. Συγκεκριμένα, οι περιορισμοί αντικαθίστανται από τις πρώτης τάξης, Taylor προσεγγίσεις τους όπως φαίνεται παρακάτω:

$$g_l(\mathbf{x}_l) = \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_l^{(k)}}\right), \mathbf{x}_l = [\tilde{\mathbf{z}}, \tilde{\mathbf{r}}_l, \tilde{L}_l] \forall l \in \{i, j\}$$

Η προσέγγιση γύρω από το σημείο  $\mathbf{x}_{\mathbf{l}_0}$  δίνεται ως:

$$g_l(\mathbf{x}) \approx h_l(\mathbf{x}_{\mathbf{l}_0}) + \nabla g_l(\mathbf{x}_{\mathbf{l}_0})(\mathbf{x} - \mathbf{x}_{\mathbf{l}_0}), \forall l \in \{i, j\} \quad (3.16)$$

όπου:

$$\begin{aligned} \nabla g_l(\mathbf{x}_{\mathbf{l}_0}) &= [\mathbf{e}_z, \mathbf{e}_{\mathbf{rl}}, \frac{\partial g}{\partial L_{lk}}], \forall l \in \{i, j\} \\ \mathbf{e}_z(k) &= \frac{\partial g}{\partial z_k}, \\ \mathbf{e}_{\mathbf{rl}}(k) &= \frac{\partial g}{\partial r_{lk}}, \forall k = 1, \dots, 8 \\ \frac{\partial g}{\partial z_k} &= \frac{\partial g}{\partial r_{lk}} = \frac{e^{\tilde{z}_k + \tilde{r}_l^{(k)}}}{\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_l^{(k)}} + \frac{1}{B} e^{\tilde{L}_j}} \\ \frac{\partial g}{\partial L_l} &= \frac{\frac{1}{B} e^{\tilde{L}_l}}{\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{r}_l^{(k)}} + \frac{1}{B} e^{\tilde{L}_l}} \end{aligned}$$

Το πλήρες πρόβλημα διατυπώνεται τελικά ως εξής:

$$\begin{aligned} \min_{\tilde{\mathbf{z}}, \tilde{\mathbf{p}}, \tilde{\mathbf{L}}, \tilde{\mathbf{f}}, \tilde{y}} \quad & \tilde{y} \\ \text{s.t.} \quad & C_1 : \log\left(\frac{\tilde{N}_i}{B}\right) - g_i(\mathbf{x}_{\mathbf{i}_0}) - \nabla g_i(\mathbf{x}_{\mathbf{i}_0})(\mathbf{x}_i - \mathbf{x}_{\mathbf{i}_0}) \leq 0 \\ & C_2 : \log\left(\frac{\tilde{N}_j}{B}\right) - g_j(\mathbf{x}_{\mathbf{j}_0}) - \nabla g_j(\mathbf{x}_{\mathbf{j}_0})(\mathbf{x}_j - \mathbf{x}_{\mathbf{j}_0}) \leq 0 \\ & C_3 : \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_i^{(k)}} + \frac{k_i X_i}{NB_o} e^{\tilde{L}_i + 2\tilde{f}_i}\right) - \log\left(\frac{E_i}{NB_o}\right) \leq 0, \\ & C_4 : \log\left(\sum_{k=1}^8 e^{\tilde{z}_k + \tilde{p}_j^{(k)}} + \frac{k_j X_j}{NB_o} e^{\tilde{L}_j + 2\tilde{f}_j}\right) - \log\left(\frac{E_j}{NB_o}\right) \leq 0, \\ & C_5 : \tilde{f}_i - \log(f_{max}) \leq 0, \\ & C_6 : \tilde{f}_j - \log(f_{max}) \leq 0, \\ & C_7 : \log\left(\sum_{k=1}^8 e^{\tilde{z}_k}\right) - \tilde{y} \leq 0, \\ & C_8 : \tilde{L}_i - \tilde{f}_i + \log X_i - \tilde{y} \leq 0, \\ & C_9 : \tilde{L}_j - \tilde{f}_j + \log X_j - \tilde{y} \leq 0, \\ & C_{10} : \log\left(\frac{1}{g_{ik}} e^{-\tilde{p}_i^{(k)}} + \frac{g_{jk}}{g_{ik}} e^{\tilde{p}_j^{(k)} - \tilde{p}_i^{(k)}}\right) + \log(2e^{\tilde{r}_i^{(k)}} - 1) \leq 0, \forall k = 1, \dots, 4 \\ & C_{11} : \log(2e^{\tilde{p}_j^{(k)}} - 1) - \tilde{p}_j^{(k)} - \log g_{jk} \leq 0, \forall k = 1, \dots, 4 \\ & C_{12} : \log\left(\frac{1}{g_{jk}} e^{-\tilde{p}_j^{(k)}} + \frac{g_{ik}}{g_{jk}} e^{\tilde{p}_i^{(k)} - \tilde{p}_j^{(k)}}\right) + \log(2e^{\tilde{r}_j^{(k)}} - 1) \leq 0, \forall k = 5, \dots, 8 \\ & C_{13} : \log(2e^{\tilde{p}_i^{(k)}} - 1) - \tilde{p}_i^{(k)} - \log g_{ik} \leq 0, \forall k = 5, \dots, 8 \end{aligned} \quad (3.17)$$

Το πρόβλημα μπορεί να λυθεί με τη χρήση επαναληπτικού αλγορίθμου που χρησιμοποιεί ως αρχικό σημείο προσέγγισης της εκάστοτε επανάληψης το βέλτιστο σημείο της προηγούμενης, μέχρι να συγκλίνουν επαρκώς.

---

**Algorithm 3** Solution of DC optimization problem (32)

---

**Initialize:**  $A, e, \tilde{\mathbf{z}}, \tilde{\mathbf{r}}$

**while**  $A > e$  **do**

**Solve:** optimization problem (9) for  $\mathbf{z}^*, \mathbf{r}^*, \mathbf{L}^*$

$\tilde{\mathbf{x}} = [\tilde{\mathbf{z}}, \tilde{\mathbf{r}}, \mathbf{L}]$

$A = \|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2^2$

$\tilde{\mathbf{z}} \leftarrow \mathbf{z}^*$

$\tilde{\mathbf{r}} \leftarrow \mathbf{r}^*$

**end while**

**Output:**  $\tilde{\mathbf{z}}^*, \tilde{\mathbf{p}}^*, \tilde{\mathbf{L}}^*, \tilde{\mathbf{f}}^*, \tilde{\mathbf{y}}^*$

---

## Κεφάλαιο 4

# Αποτελέσματα Προσομοιώσεων

Για τις προσομοιώσεις υποθέτουμε ότι  $B = 1$  MHz, ο αριθμός των βιτς για κάθε υπολογιστική εργασία είναι  $N_i = N_j = 0.5$  Mbits και  $E_i = E_j$  για κάθε χρήστη. Οι παράμετροι που ποικίλλουν στην προσομοίωση είναι ο SNR(db) και η μετρική  $K = \frac{E}{BN_0}$  όπου  $p$  είναι το μέγιστο SNR που χρησιμοποιείται. Αυτή η μετρική μετριέται σε δευτερόλεπτα και μπορεί να ερμηνευθεί ως ο χρόνος που το σύστημα μπορεί να λειτουργήσει με ένα δοσμένο SNR (συμβολίζεται ως  $p$ ), με τη διαθέσιμη ενέργεια  $E$ .

### 4.1 Περιγραφή benchmark

Ως μέτρο σύγκρισης για τα προβλήματα της πλήρους εκφόρτωσης θα χρησιμοποιηθεί μία απλούστερη μέθοδος επίλυσης η οποία βασίζεται στην εξής παρατήρηση: για τα πρώτα τέσσερα σενάρια λειτουργίας του συστήματος, αποκωδικοποιείται πρώτος ο χρήστης  $i$  και για τα υπόλοιπα τέσσερα αποκωδικοποιείται πρώτος ο χρήστης  $j$ . Δηλαδή κατά τη διάρκεια  $T$  του time sharing, ένα από τα μηνύματα των χρηστών αποκωδικοποιείται πάντα πρώτο, ανεξάρτητα από τις συνθήκες του καναλιού. Σύμφωνα με αυτή τη λογική και κάνοντας χρήση μιας παραμέτρου  $c_1$ , όπου  $c_1 \in [0, 1]$ , μπορεί να διατυπωθεί ότι για τον χρήστη  $i$ , για τα πρώτα 4 σενάρια όπου αποκωδικοποιείται πρώτος, η ισχύς εκπομπής του και το αντίστοιχο SNR είναι  $c_1 P_{max}$ ,  $c_1 p_{max}$  και για τα υπόλοιπα 4 σενάρια που αποκωδικοποιείται δεύτερος, κάνοντας χρήση μιας ακόμη παραμέτρου  $c_2$ , με  $c_2 \in [0, 1]$  είναι  $c_2 P_{max}$  και  $c_2 p_{max}$ .

Με την ακριβώς αντίστοιχη λογική, για τον χρήστη  $j$ , για τα πρώτα 4 σενάρια που αποκωδικοποιείται δεύτερος η ισχύς εκπομπής του και το αντίστοιχο SNR είναι  $c_2 P_{max}$ ,  $c_2 p_{max}$  και για τα υπόλοιπα 4 που αποκωδικοποιείται πρώτος είναι  $c_1 P_{max}$  και  $c_1 p_{max}$ . Το πρόβλημα διατυπώνεται ως προς τις μεταβλητές  $z_k$  και τις παραμέτρους  $c_1, c_2$  και λύνεται επιλέγοντας τις καλύτερες τιμές καθυστέρησης  $\sum_{k=1}^8 z_k$  για  $c \in [0, 1]$  με διακριτοποίηση

βήματος 0.1.

$$\begin{aligned}
 \min_{\mathbf{z}} \quad & \sum_{k=1}^8 z_k \\
 \text{s.t.} \quad & C_1 : z_1 BR_i^{(1)} + \dots + z_4 BR_i^{(4)} + \dots + z_5 BR_i^{(7)} + \dots + z_8 BR_i^{(8)} \geq \tilde{N}_i, \\
 & C_2 : z_1 BR_j^{(1)} + \dots + z_4 BR_j^{(4)} + \dots + z_5 BR_j^{(5)} + \dots + z_8 BR_j^{(8)} \geq \tilde{N}_j, \\
 & C_3 : z_1 c_1 p_{max} + \dots + z_4 c_1 p_{max} + \dots + z_5 c_2 p_{max} + \dots + z_8 c_2 p_{max} \leq \frac{E_i}{BN_0}, \\
 & C_4 : z_1 c_2 p_{max} + \dots + z_4 c_2 p_{max} + \dots + z_5 c_1 p_{max} + \dots + z_8 c_1 p_{max} \leq \frac{E_j}{BN_0}, \\
 & C_5 : z_k, c_1, c_2 \geq 0,
 \end{aligned} \tag{4.1}$$

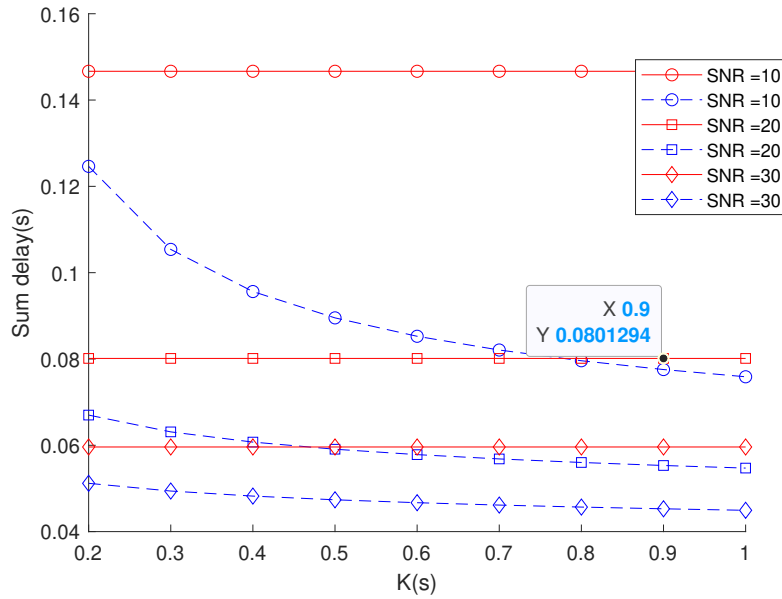
Για την delay constrained energy minimization, ακολουθώντας αντίστοιχη λογική το πρόβλημα διατυπώνεται ως:

$$\begin{aligned}
 \min_{\mathbf{z}} \quad & (c_1 * w_i + c_2 * w_j) * p_{max} * \sum_{k=1}^4 z_k + (c_2 * w_i + c_1 * w_j) * p_{max} * \sum_{k=5}^8 z_k \\
 \text{s.t.} \quad & C_1 : z_1 BR_i^{(1)} + \dots + z_4 BR_i^{(4)} + \dots + z_5 BR_i^{(7)} + \dots + z_8 BR_i^{(8)} \geq \tilde{N}_i, \\
 & C_2 : z_1 BR_j^{(1)} + \dots + z_4 BR_j^{(4)} + \dots + z_5 BR_j^{(5)} + \dots + z_8 BR_j^{(8)} \geq \tilde{N}_j, \\
 & C_3 : \sum_{k=1}^8 z_k \leq T_{max}
 \end{aligned} \tag{4.2}$$

## 4.2 Προσομοιώσεις delay minimization

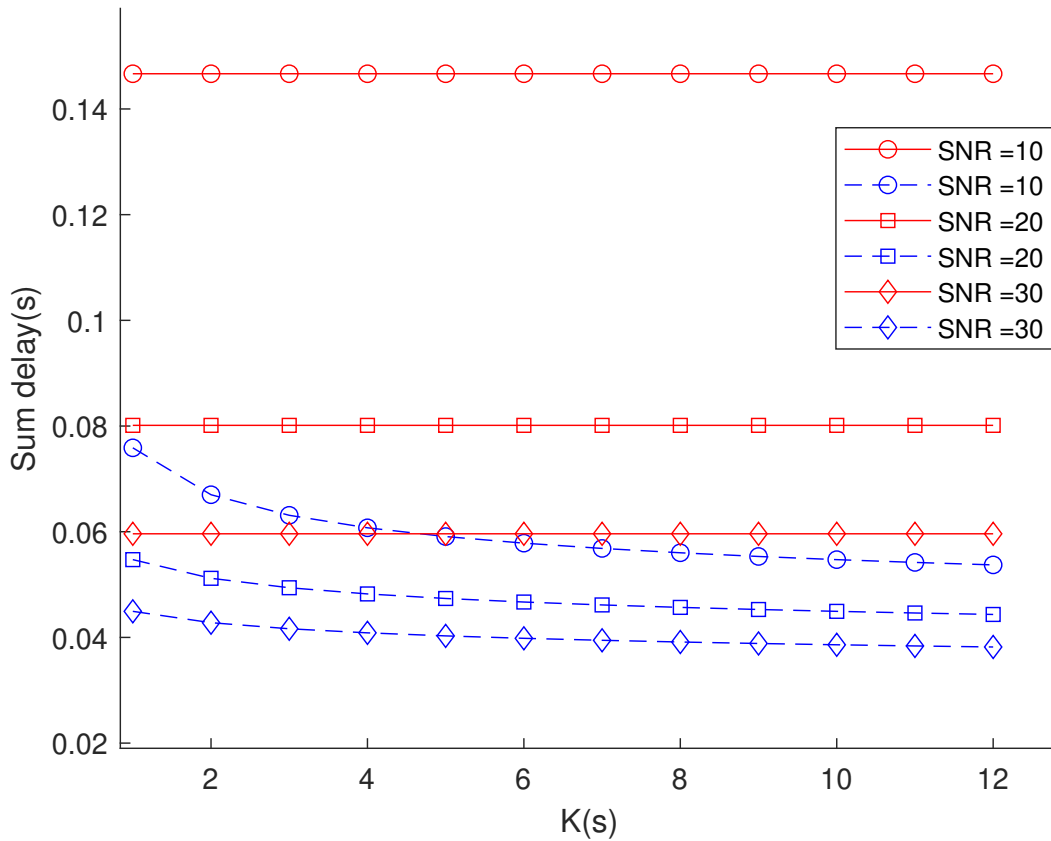
### Επίδραση μέγιστου SNR

Για δεδομένες τιμές  $K$  στο διάστημα  $[0.2, 1]$  και τιμές  $SNR = 10, 20, 30$  έχουμε:



Σχήμα 4.1: Μεταβολή  $K$

και για  $K$  στο διάστημα  $[1, 12]$

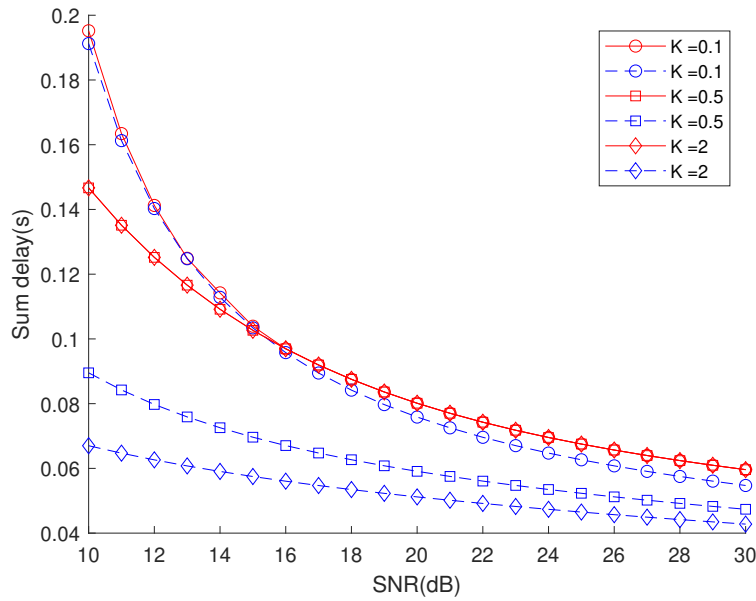


Σχήμα 4.2: Μεταβολή K

Παρατηρούμε πως για κάθε τιμή μέγιστου SNR, η μεθοδολογία γεωμετρικού και DC προγραμματισμού πετυχαίνει μικρότερες τιμές συνολικής καθυστέρησης από τη μεθοδολογία γραμμικού προγραμματισμού όσο αυξάνεται ο διαθέσιμος χρόνος σε μέγιστη εκπομπή, με μεγαλύτερες απολαβές για μικρότερα SNR. Το benchmark επιτυγχάνει βέλτιστες λύσεις που δεν επηρεάζονται από την αλλαγή K, καθώς ο περιορισμός ενέργειας παρουσιάζει γραμμική συσχέτιση με το K και μη γραμμική, φθίνουσα με το SNR ως  $\frac{E}{BN_0} = K * p_{max}$ . Η μεθοδολογία της διπλωματικής εκμεταλλεύεται καλύτερα την διαθέσιμη ενέργεια και παρατηρούμε πως για K μεγαλύτερο του 0.6, για εκάστοτε SNR, η βέλτιστη καθυστέρηση συγκλίνει.

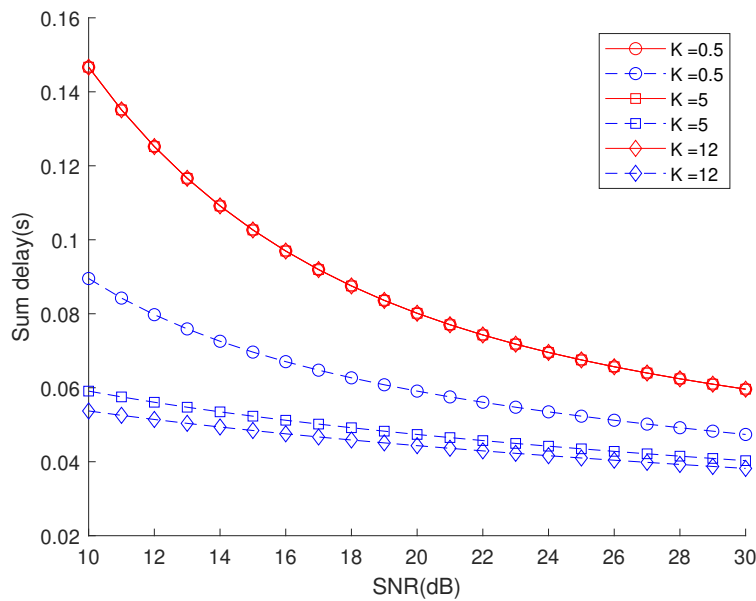
## Επίδραση SNR

Για δεδομένες τιμές  $K \in \{0.1, 0.2, 0.5\}$  και τιμές  $SNR \in [1030]$  έχουμε:



Σχήμα 4.3: Μεταβολή SNR

και για  $K \in \{0.5, 5, 12\}$  seconds



Σχήμα 4.4: Μεταβολή SNR

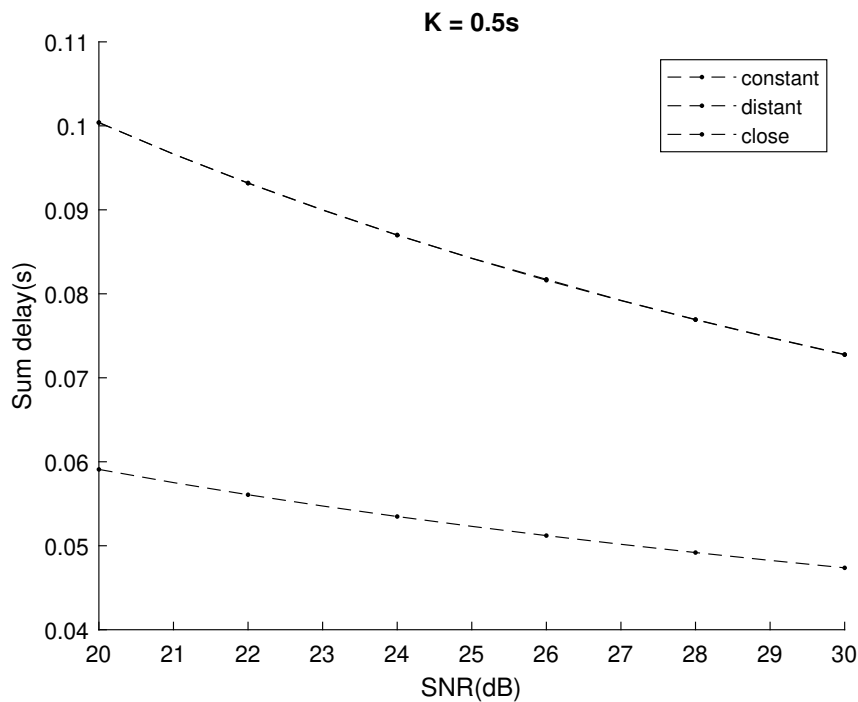
Παρόμοια με τις προσομοιώσεις με σταθερό SNR, η επιλογή  $K$  δεν επηρεάζει τη βέλτιστη λύση της γραμμικής μεθοδολογίας και παρατηρούμε πως με μεγαλύτερο  $K$  και για SNR μεγαλύτερο των 22 dB, έχουμε σύγκλιση της καθυστέρησης σε τιμή έως και 30% μεγαλύτερη της γραμμικής λύσης. Για  $K = 0.1$ , οι λύσεις των δύο μεθόδων συγκλίνουν, ειδικά για μικρές τιμές SNR.

## Αποστάσεις

Τα παρακάτω σχήματα προκύπτουν με σταθερά κέρδη καναλιού, υποθέτοντας διαλείψεις Rayleigh , και αρχικά συγκρίνονται 3 διαφορετικά σενάρια για την τοποθέτηση των χρηστών. Η απόσταση μεταξύ του χρήστη  $l$  και του PPH  $m$  συμβολίζεται με  $d_{lm} \forall l \in i, j$  και  $\forall m \in a, b$ , ο παράγοντας απώλειας πορείας (path loss factor) ως  $w_{lm} \forall l \in i, j$  και  $\forall m \in a, b$ , και τα διαφορετικά σενάρια διατυπώνονται ως εξής:

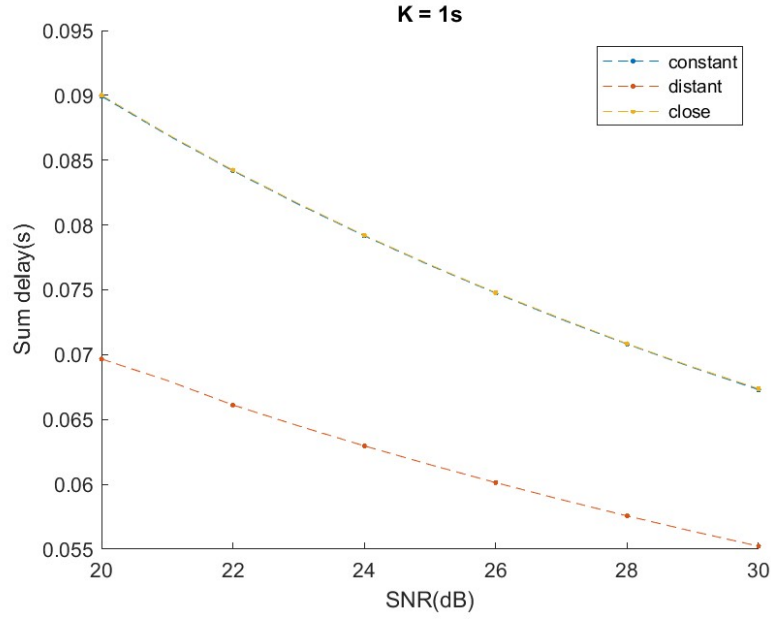
- η constant περίπτωση στην οποία όλες οι αποστάσεις  $d_{lm}$  είναι ίδιες και κανονικοποιούνται σε 1
- η distant περίπτωση στην οποία οι λόγοι των αποστάσεων είναι  $\frac{d_{ia}}{d_{ib}} = \frac{d_{jb}}{d_{ja}} = 1/5$ ,  $d_{ia} = d_{jb}$  και οι λόγοι των παραγόντων απώλειας πορείας  $\frac{w_{im}}{w_{jm}} = \left(\frac{d_{im}}{d_{jm}}\right)^3, \forall m \in a, b$ , που ισοδυναμεί με path loss exponent ίσο με 3.
- η constant περίπτωση , στην οποία οι αποστάσεις είναι  $\frac{d_{ia}}{d_{ib}} = \frac{d_{ja}}{d_{jb}} = 1/5$ ,  $d_{ia} = d_{ja}$ , με path loss exponent ίσο με 3.

Στα Σχήματα , και , μπορεί να παρατηρηθεί η επίδραση της απόστασης των χρηστών στη συνολική καθυστέρηση, καθορίζοντας το  $K$  και μεταβάλλοντας το μέγιστο  $SNR \in \mu$  που επιτυγχάνεται από τους χρήστες, από 20 έως 30 dB.

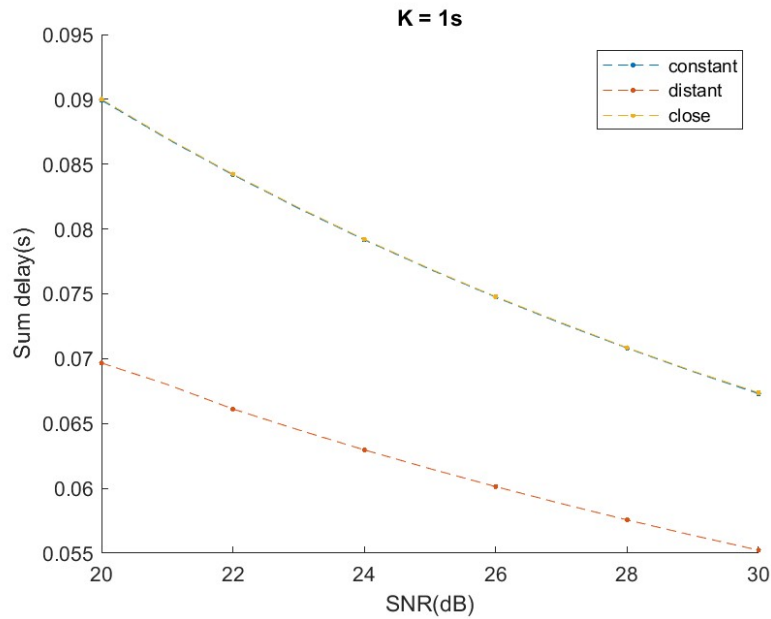


Σχήμα 4.5: Σύγκριση αποστάσεων





Σχήμα 4.6: Σύγκριση αποστάσεων

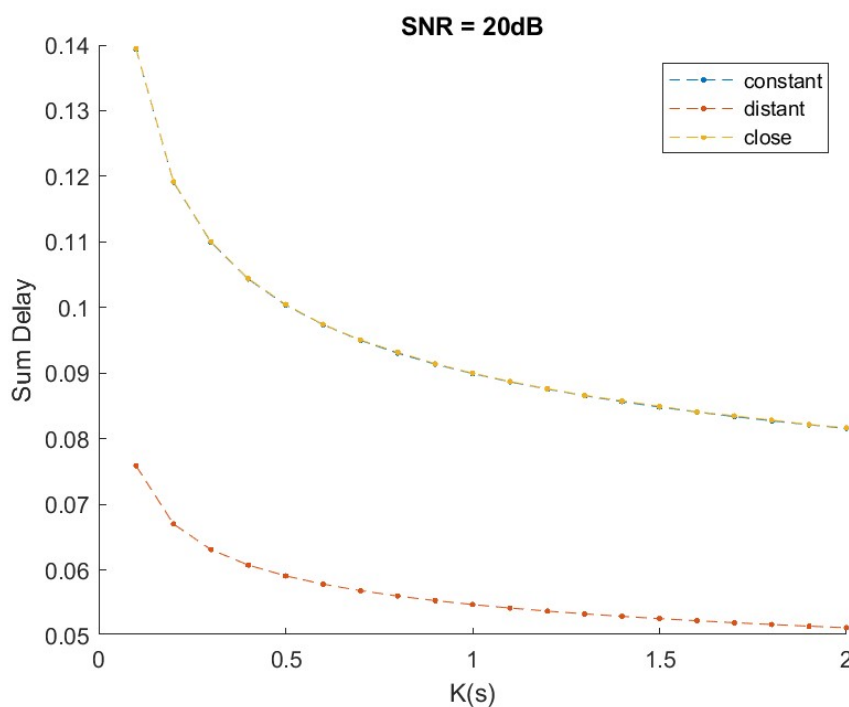


Σχήμα 4.7: Σύγκριση αποστάσεων

Αυξάνοντας τη μέγιστη κανονικοποιημένη ισχύ κάθε χρήστη, η συνολική καθυστέρηση μειώνεται, το οποίο αναμένεται επειδή, στο βέλτιστη επιλογή σειρών αποκωδικοποίησης κατά το time sharing, η διεργασία βελτιστοποίησης μπορεί να ικανοποιήσει τους περιορισμούς ρυθμαπόδοσης με μεγαλύτερη ευκολία, διατηρώντας παράλληλα χαμηλή την συνολική ενέργεια. Επιπλέον, συγκρίνοντας τα διαγράμματα παρατηρούμε ότι για μεγαλύτερες τιμές  $K$ , επιτυγχάνεται μικρότερη καθυστέρηση σε όλες τις πιθανές διατάξεις.

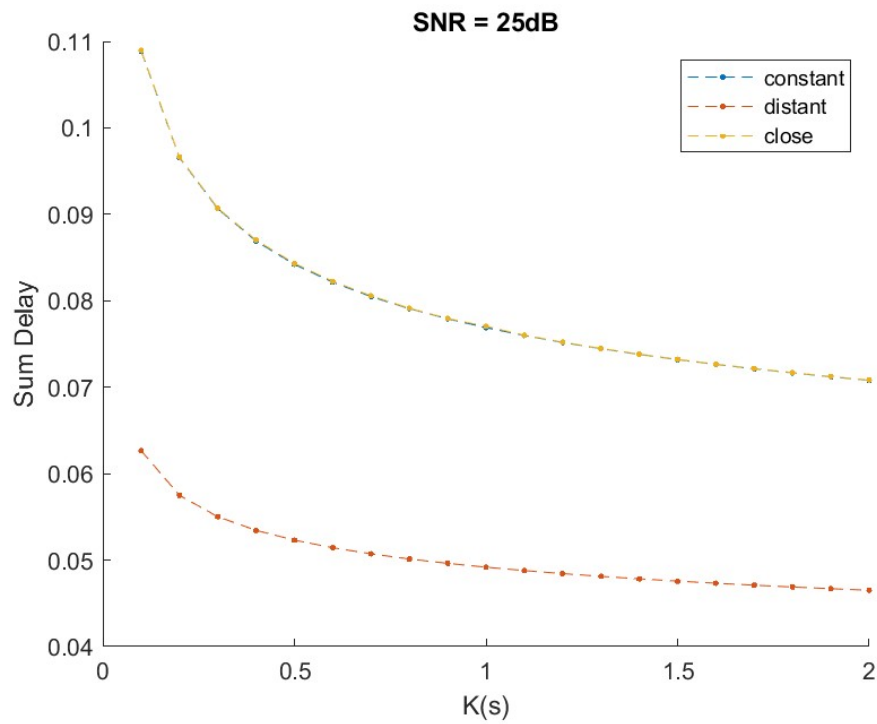
Είναι ενδιαφέρον να σημειωθεί ότι, στην απομακρυσμένη διάταξη των χρηστών (distant), η συνολική καθυστέρηση που επιτυγχάνεται είναι μικρότερη από την αντίστοιχη καθυστέρηση

ση για τις σταθερές(constant) και κοντινές(close) διατάξεις, ενώ στις περιπτώσεις σταθερής και κοντινής διάταξης η συνολική καθυστέρηση είναι ταυτόσημη, για όλες τις τιμές του SNR. Μια πιθανή εξήγηση για την διαφορά της βέλτιστης λύσης όσον αφορά τις αποστάσεις, θα ήταν το συγκριτικό πλεονέκτημα στην απόδοση που παρουσιάζουν τα NOMA συστήματα σε σχέση με τα OMA, όταν οι χρήστες έχουν σημαντική διαφορά συνθηκών καναλιών[16]. Αντίστοιχη συμπεριφορά της βέλτιστης λύσης παρατηρούμε για την περίπτωση όπου μεταβάλλουμε το  $K$ , κρατώντας το SNR σταθερό. Στα σχήματα παρουσιάζεται το αποτέλεσμα της αύξησης του παρεχόμενου χρόνου λειτουργίας σε μέγιστη ισχύ,  $K$ , από 0.1 σε 2 s με SNR σταθερό σε 20,25 και 30 dB, αντίστοιχα.

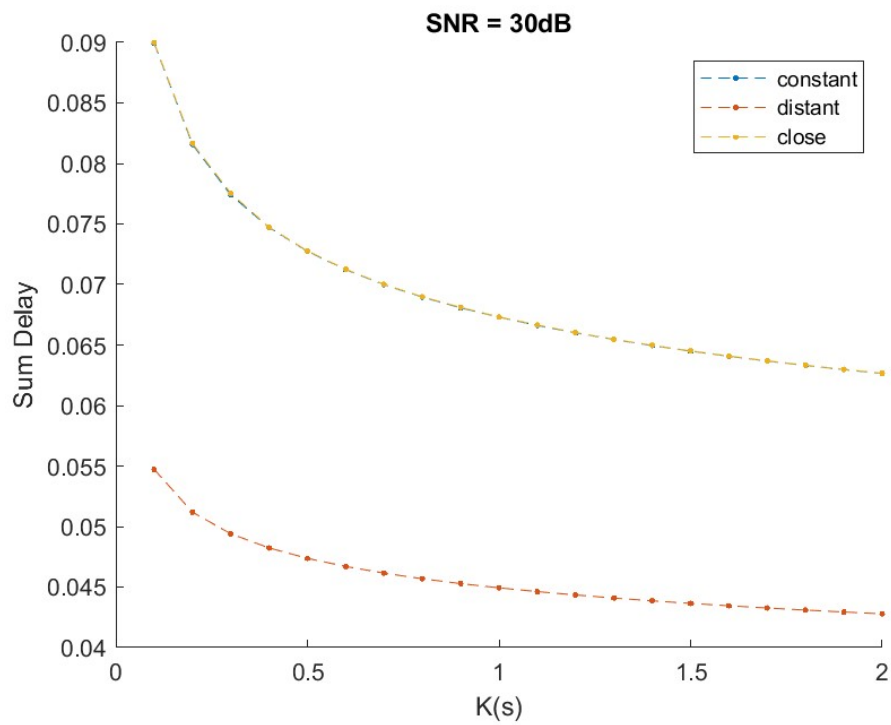


Σχήμα 4.8: Σύγκριση αποστάσεων

Είναι εμφανές πώς για μικρές τιμές της παραμέτρου  $K$  δηλαδή για λιγότερο διαθέσιμος χρόνο εκπομπής σε μέγιστη ενέργεια, η ανάθεση είναι πιο δύσκολη και απαιτεί σημαντικά περισσότερο χρόνο, σε milliseconds από την ελάχιστη δυνατή σε αυτές τις προσομοιώσεις. Η σχέση της διάταξης με την βέλτιστη λύση παραμένει η ίδια και παρατηρούμε πως για  $K > 0.6s$ , η αύξηση του  $K$  δεν επηρεάζει σημαντικά την μέγιστη καθυστέρηση. Συγκρίνοντας τα διαγράμματα μεταξύ τους παρατηρούμε πως η ελάχιστη καθυστέρηση επιτυγχάνεται με την αύξηση του SNR, κάτι που επιβεβαιώνει και τα αποτελέσματα της παραπάνω ανάλυσης.



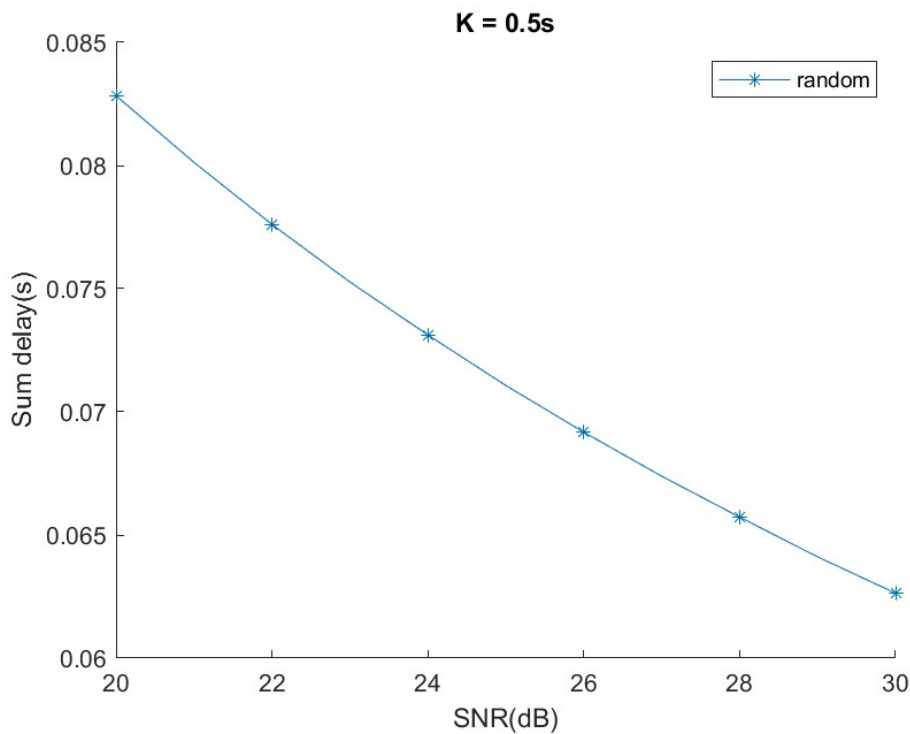
Σχήμα 4.9: Σύγκριση αποστάσεων



Σχήμα 4.10: Σύγκριση αποστάσεων

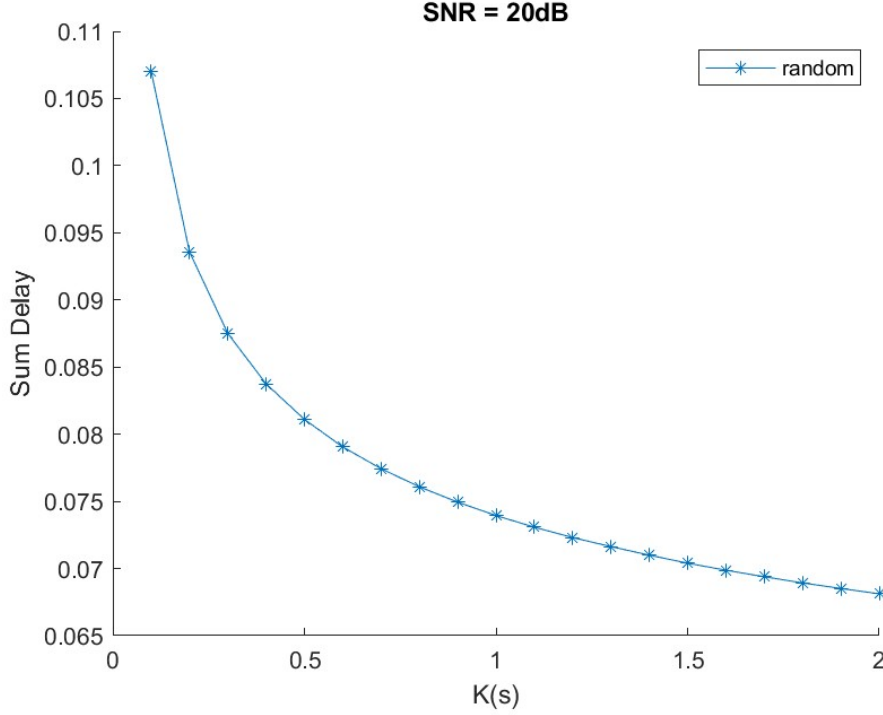
## Monte Carlo

Οι γενικές παρατηρήσεις για την energy constrained delay minimization μπορούν να επιβεβαιωθούν μέσω μιας προσομοίωσης Monte Carlo για την οποία θεωρούμε ότι, σε κάθε δοκιμή, η απόσταση των χρηστών από τα RRHs είναι διαφορετική με αποτέλεσμα διαφορετικά κέρδη καναλιών, όπου οι διαλείψεις μικρής κλίμακας (small scale fading) λαμβάνονται από την τυχαία μεταβλητή  $h_l \sim \mathcal{CN}(0, 1)$ ;  $\forall l \in \{i, j\}$ . Το σχήμα δίνεται για 50 δοκιμές, και σταθερό  $K = 0.5$  που όπως είδαμε αρκεί για να χαρακτηρίσει την συμπεριφορά του συστήματος για τις περισσότερες τιμές:



Σχήμα 4.11: Σύγκριση αποστάσεων

Παρατηρούμε πως η 'μέση' καθυστέρηση για αυθαίρετες τιμές κερδών καναλιού βρίσκεται κοντά στην καθυστέρηση για την distant διάταξη των χρηστών. Στο σχήμα παρουσιάζεται η μέση συμπεριφορά του συστήματος για 50 δοκιμές με  $SNR = 20$ :



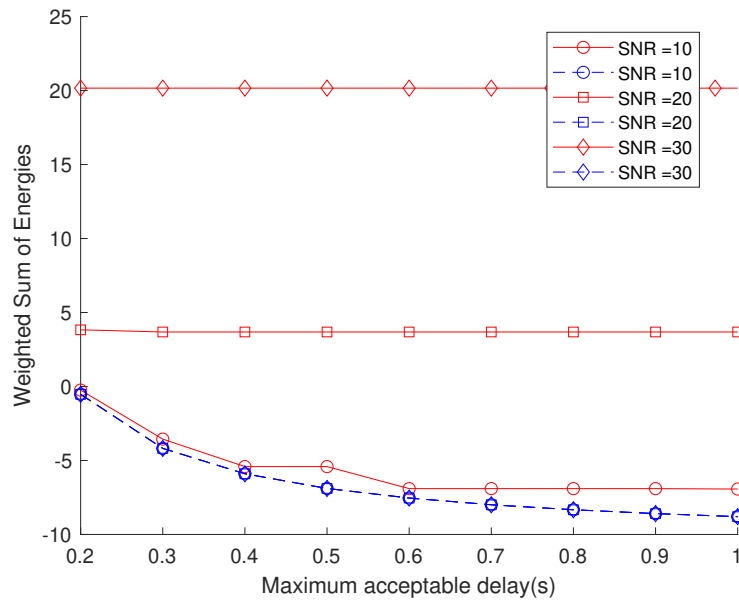
Σχήμα 4.12: Σύγκριση αποστάσεων

### 4.3 Προσομοιώσεις energy minimization

Στην περίπτωση της ελαχιστοποίησης της ενεργειακής κατανάλωσης, οι προσομοιώσεις εκτελούνται με σταθερές τιμές αποδεκτής καθυστέρησης  $T_{max}$ , διαφορετικές τιμές των σταθερών βάρους  $w_i, w_j$  που ελέγχουν το ενεργειακό ισοζύγιο των χρηστών στην συνολική κατανάλωση του συστήματος και πανομοιότυπες διατάξεις χρηστών με την προηγούμενη ενότητα. Σε όλα τα παρακάτω διαγράμματα, ο άξονας της ενέργειας αντιπροσωπεύει την ενέργεια κανονικοποιημένη με την φασματική πυκνότητα του θορύβου,  $\frac{E}{B \cdot N_0}$ .

### Επίδραση αποδεκτής καθυστέρησης $D_{\max}$

Για τιμές του μέγιστου  $SNR \in \{10, 20, 30\}$ , αποδεκτή καθυστέρηση  $D_{\max}$  στο διάστημα  $[0.2 \ 1]$  και  $w_i = w_j = 0.5$  έχουμε πώς:



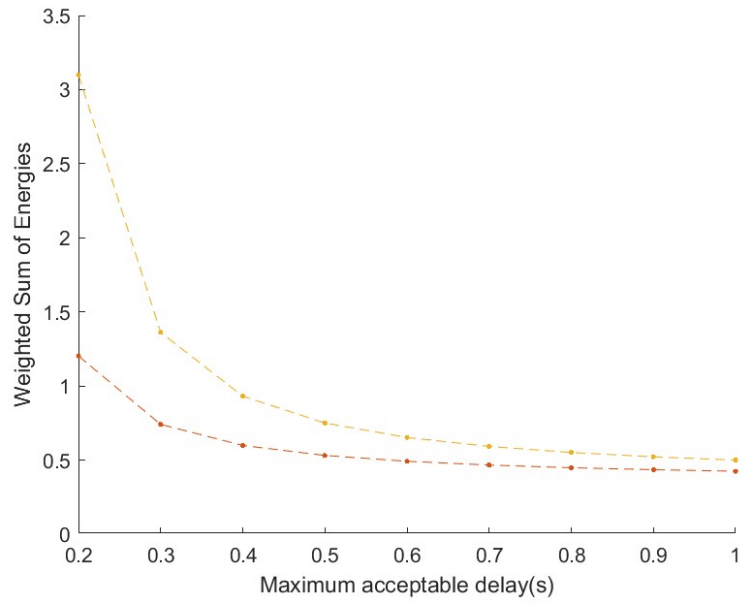
Σχήμα 4.13: Σύγκριση αποστάσεων

Παρατηρούμε πως, για οποιαδήποτε τιμή του μέγιστου SNR, η μεθοδολογία γεωμετρικού-DC προγραμματισμού επιτυγχάνει πολύ μικρή συνολική ενεργειακή κατανάλωση κατά τάξεις μεγέθους σε σχέση με τη μεθοδολογία γραμμικού προγραμματισμού. Η αποδεκτή καθυστέρηση βελτιώνει την ενεργειακή κατανάλωση, όπως ήταν αναμενόμενο.

### Επίδραση αποστάσεων

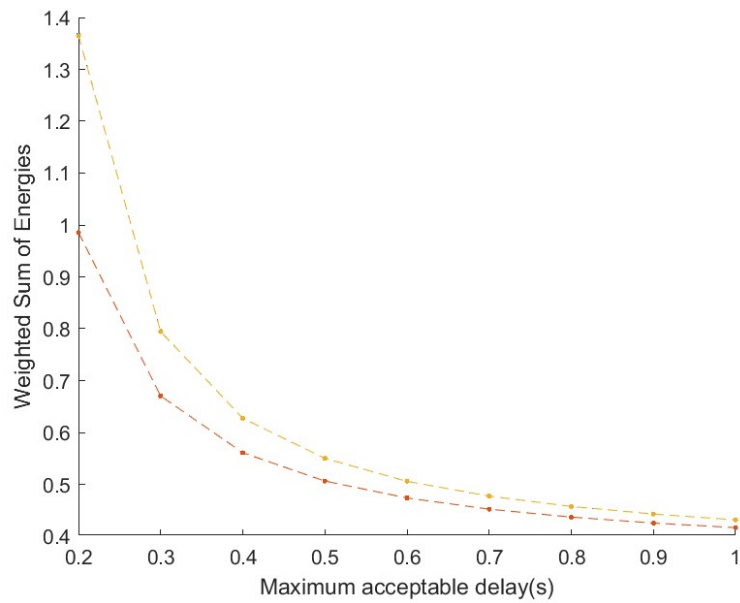
Οι διατάξεις constant και close παρουσιάζουν ίδιες βέλτιστες τιμές και απεικονίζονται σε κοινή καμπύλη.

Συγκεκριμένα, για  $T_{\max} \in [0.21]$ , στο σχήμα 4.9 απεικονίζεται η συνολική ενεργειακή κατανάλωση  $E_{FO}$  για  $w_i = w_j = 0.5$  και τιμές καθυστέρησης  $T_{\max}$  από 0.2 έως 2 s:



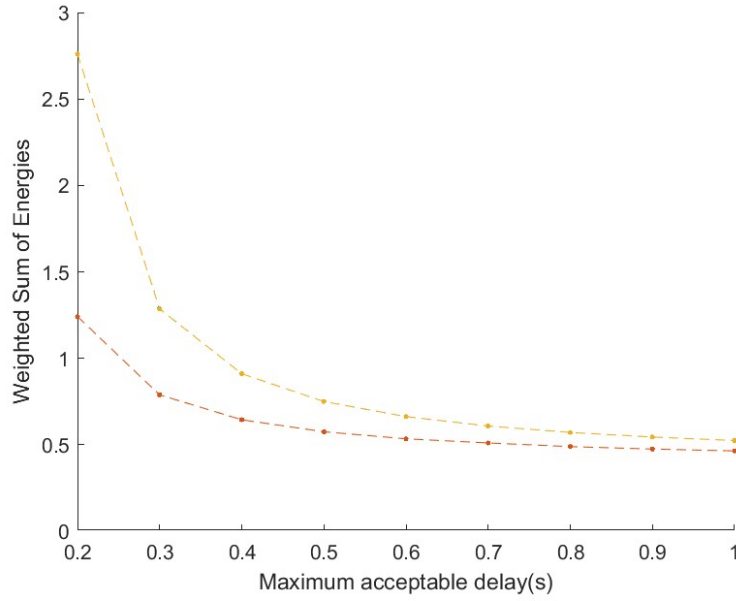
Σχήμα 4.14: Σύγκριση αποστάσεων

Για  $w_i = 0.1, w_j = 0.9$ , δηλαδή δίνοντας προτεραιότητα στις ανάγκες κατανάλωσης μόνο του ενός χρήστη (στην προκειμένη του  $j$ ):



Σχήμα 4.15: Σύγκριση αποστάσεων

Τέλος, για  $w_i = 0.4, w_j = 0.7$ :



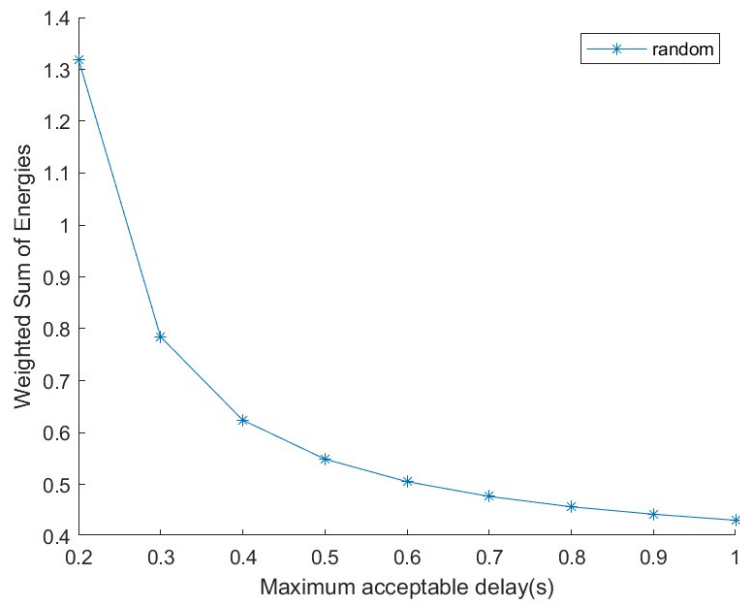
Σχήμα 4.16: Σύγκριση αποστάσεων

Η κόκκινη γραμμή αναπαριστά τη distant διάταξη και η κίτρινη την κοινή close/constant. Στα παραπάνω σχήματα είναι εμφανές πως η distant διάταξη, λόγω της μεγάλης διαφοράς των καναλιών προς τα PPHs επωφελείται από την τεχνική NOMA για να μειώσει την συνολική ενεργειακή κατανάλωση των χρηστών περισσότερο από τις υπόλοιπες διατάξεις. Αυτό είναι πιο εμφανές για  $T_{max}$  μικρότερο του 0.4s. Συγκρίνοντας, τα διαγράμματα μεταξύ τους, παρατηρούμε πως δίνοντας μεγαλύτερη βαρύτητα σε έναν χρήστη ( $w_i = 0.1, w_j = 0.9$ , Σχήμα 4.9) η ελάχιστη ενεργειακή κατανάλωση είναι σημαντικά μικρότερη σε σχέση με τις περιπτώσεις όπου κατανάλωση μοιράζεται πιο ισομερώς στους χρήστες και ταυτόχρονα, μειώνεται η επίδραση της διάταξης των χρηστών στο χώρο καθώς αυξάνεται η αποδεκτή καθυστέρηση (τα διαγράμματα έχουν πολύ πιο μικρή απόκλιση). Μια πιθανή εξήγηση για την παρατήρηση αυτή, θα μπορούσε να είναι πως η διαδικασία βελτιστοποίησης δίνει βαρύτητα στην ικανοποίηση των περιορισμών του κυρίαρχου ( $w_j = 0.9$ ) και στην συνέχεια, βρίσκει ικανοποιητικές λύσεις για τον ενεργειακή κατανάλωση του άλλου ( $w_i = 0.1$ ).

## Monte Carlo

Οι γενικές παρατηρήσεις για την delay constrained energy minimization μπορούν να επιβεβαιωθούν μέσω μιας προσομοίωσης Monte Carlo για την οποία θεωρούμε ότι, σε κάθε δοκιμή, η απόσταση των χρηστών από τα RRHs είναι διαφορετική με αποτέλεσμα διαφορετικά κέρδη καναλιών, όπου οι διαλείψεις μικρής κλίμακας (small scale fading) λαμβάνονται από την τυχαία μεταβλητή  $h_l \sim \mathcal{CN}(0, 1)$ ;  $\forall l \in \{i, j\}$ . Το σχήμα δίνεται για 50 δοκιμές,  $w_i = w_j = 0.5$  που όπως είδαμε αρκεί για να χαρακτηρίσει την συμπεριφορά του συστήματος για τις περισσότερες τιμές:





Σχήμα 4.17: Monte Carlo

Η μέση συμπεριφορά ως προς την ενεργειακή κατανάλωση είναι φθίνουσα, όπως ήταν αναμενόμενο και κοντά στις τιμές καθυστέρησης της distant διάταξης.

## Κεφάλαιο 5

# Συμπεράσματα - Μελλοντική Έρευνα

### 5.1 Επίλογος - Συμπεράσματα

Η παρούσα διπλωματική εργασία επικεντρώθηκε στη μελέτη, τη δομή και τη λειτουργία ενός συστήματος MEC με χρήση DU-NOMA, του οποίου η φιλοσοφία ανταποκρίνεται στις απαιτήσεις των ασύρματων δικτύων μελλοντικών γενιών. Αναφέρεται όλο το απαραίτητο θεωρητικό υπόβαθρο που χρειάζεται για τη κατανόησή της συγκεκριμένης εφαρμογής. Προτείνεται, η κατάλληλη διάταξη του συστήματος και εξετάζεται διεξοδικά η λειτουργία επικοινωνίας των χρηστών με τα RRHs. Η ενορχήστρωση του συστήματος βελτιστοποιείται για τη μέθοδο της πλήρους εκφόρτωσής (full offloading) για τις περιπτώσεις της ελαχιστοποίησης της καθυστέρησης εκφόρτωσης, δεδομένου περιορισμών ενεργειακής κατανάλωσης και ελαχιστοποίησης της ενεργειακής κατανάλωσης των χρηστών, δεδομένου περιορισμένης αποδεκτής καθυστέρησης. Επιπλέον, διατυπώνεται το πρόβλημα ελαχιστοποίησης της καθυστέρησης για τη μέθοδο partial offloading και καταstrώνονται προσομοιώσεις για να επιβεβαιώσουν την αποτελεσματικότητα της αρχιτεκτονικής. Συγκεκριμένα, παρατηρείται επίτευξη μικρότερης καθυστέρησης σε περιπτώσεις καλύτερου διαχωρισμού των καναλιών των χρηστών, επίτευξη μικρότερης συνολικής ενεργειακής κατανάλωσης σε περιπτώσεις όπου η ισχύς κατανέμεται ανόμοια μεταξύ τους και συνολικά καλύτερη επίδοση με αύξηση του μέγιστου SNR και του χρόνου εκπομπής σε αυτό.

### 5.2 Πρόταση για Μελλοντική Έρευνα

Η μεθοδολογία που αναπτύχθηκε στην παρούσα διπλωματική θα μπορούσε να χρησιμοποιηθεί για την διαχείριση της μη-κυρτότητας(non convexity) που προκύπτει σε προβλήματα resource allocation με πολλαπλούς όρους παρεμβολής, σε συστήματα MEC με άλλες τεχνικές πολλαπλής πρόσβασης. Η ανάλυση με 2 χρήστες και 2 RRHs μπορεί να χρησιμοποιηθεί

για την επέκταση του συστήματος σε διατάξεις με πολλαπλές κεραίες(multi-antenna) και πολλαπλούς σταθμούς βάσης, μέσω user pairing, μια δυνατότητα που υποστηρίζεται στα 4G/LTE δίκτυα, για τη δημιουργία παράλληλων MEC διατάξεων. Επιπλέον, η μεθοδολογία αυτή θα μπορούσε να χρησιμοποιηθεί για την βελτίωση uplink εφαρμογών σε συστήματα ταυτόχρονης επικοινωνίας και ανιχνεύσης(joint communication and sensing) ή συλλογής ενέργειας(energy harvesting) και της επακόλουθης σύγκλισης τους, στα δίκτυα 6G.

# Βιβλιογραφία

- [1] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, “A survey on 5g networks for the internet of things: Communication technologies and challenges,” *IEEE Access*, vol. 6, no. 10, pp. 3619–3647, 2018.
- [2] N. Al-Falahy and O. Y. Alani, “Technologies for 5g networks: Challenges and opportunities,” *IT Professional*, vol. 19, no. 1, pp. 12–20, Jan.-Feb. 2017.
- [3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5g be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [4] K.-C. Chen, T. Zhang, R. D. Gitlin, and G. Fettweis, “Ultra-low latency mobile networking,” *IEEE Network*, vol. 33, no. 2, pp. 181–187, 2019.
- [5] D. Soldani and A. Manzalini, “Horizon 2020 and beyond: On the 5g operating system for a true digital society,” *IEEE Vehicular Technology Magazine*, vol. 10, no. 1, pp. 32–42, 2015.
- [6] M. Alsabah, M. A. Naser, B. M. Mahmmoud, S. H. Abdulhussain, M. R. Eissa, A. Al-Baidhani, N. K. Noordin, S. M. Sait, K. A. Al-Utaibi, and F. Hashim, “6g wireless communications networks: A comprehensive survey,” *IEEE Access*, vol. 9, pp. 148 191–148 243, 2021.
- [7] S. Zhang, C. Xiang, and S. Xu, “6g: Connecting everything by 1000 times price reduction,” *IEEE Open Journal of Vehicular Technology*, vol. 1, pp. 107–115, 2020.
- [8] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, and L. Hanzo, “On the road to 6g: Visions, requirements, key technologies, and testbeds,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.
- [9] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, “Power-domain non-orthogonal multiple access (noma) in 5g systems: Potentials and challenges,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2017.

- [10] P. D. Diamantoulakis and G. K. Karagiannidis, “Performance analysis of distributed uplink noma,” *IEEE Communications Letters*, vol. 25, no. 3, pp. 788–792, 2021.
- [11] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, “A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, 2017.
- [12] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, “Application of non-orthogonal multiple access in lte and 5g networks,” *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, 2017.
- [13] Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi, “Wireless network cloud: Architecture and system requirements,” *IBM Journal of Research and Development*, vol. 54, no. 1, pp. 4:1–4:12, 2010.
- [14] “The road towards green ran,” *China Mobile Res. Inst., Beijing, China, White Paper*, Oct.2011.
- [15] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, “A survey of noma: Current status and open research challenges,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.
- [16] K. N. Pappi, P. D. Diamantoulakis, and G. K. Karagiannidis, “Distributed uplink-noma for cloud radio access networks,” *IEEE Communications Letters*, vol. 21, no. 10, pp. 2274–2277, 2017.
- [17] C.-L. I, J. Huang, R. Duan, C. Cui, J. Jiang, and L. Li, “Recent progress on c-ran centralization and cloudification,” *IEEE Access*, vol. 2, pp. 1030–1039, 2014.
- [18] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, “A comprehensive survey of ran architectures toward 5g mobile communication system,” *IEEE Access*, vol. 7, pp. 70 371–70 421, 2019.
- [19] C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, “Toward green and soft: a 5g perspective,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 66–73, 2014.
- [20] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, “The roadmap to 6g: Ai empowered wireless networks,” *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.
- [21] H. Li, K. Ota, and M. Dong, “Learning iot in edge: Deep learning for the internet of things with edge computing,” *IEEE Network*, vol. 32, no. 1, pp. 96–101, 2018.

- [22] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, “6g wireless networks: Vision, requirements, architecture, and key technologies,” *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.
- [23] H. Li, K. Ota, and M. Dong, “Eccn: Orchestration of edge-centric computing and content-centric networking in the 5g radio access network,” *IEEE Wireless Communications*, vol. 25, no. 3, pp. 88–93, 2018.
- [24] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, “A survey on mobile edge computing: The communication perspective,” *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.
- [25] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, A. Neal *et al.*, “Mobile-edge computing introductory technical white paper,” *White paper, mobile-edge computing (MEC) industry initiative*, pp. 1089–7801, 2014.
- [26] C. You, K. Huang, H. Chae, and B.-H. Kim, “Energy-efficient resource allocation for mobile-edge computation offloading,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397–1411, 2017.
- [27] K. Li, “Computation offloading strategy optimization with multiple heterogeneous servers in mobile edge computing,” *IEEE Transactions on Sustainable Computing*, pp. 1–1, 2019.
- [28] P. D. Diamantoulakis, P. S. Bouzinis, P. G. Sarigiannidis, Z. Ding, and G. K. Karagiannidis, “Optimal design and orchestration of mobile edge computing with energy awareness,” *IEEE Transactions on Sustainable Computing*, vol. 7, no. 2, pp. 456–470, 2022.
- [29] A. A. Al-Habob, O. A. Dobre, A. G. Armada, and S. Muhaidat, “Task scheduling for mobile edge computing using genetic algorithm and conflict graphs,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8805–8819, 2020.
- [30] A. A. Al-Habob, A. Ibrahim, O. A. Dobre, and A. G. Armada, “Collision-free sequential task offloading for mobile edge computing,” *IEEE Communications Letters*, vol. 24, no. 1, pp. 71–75, 2020.
- [31] Q.-V. Pham, F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W.-J. Hwang, and Z. Ding, “A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art,” *IEEE Access*, vol. 8, pp. 116 974–117 017, 2020.
- [32] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, “On the performance gain of noma over oma in uplink communication systems,” *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 536–568, 2020.

- [33] P. Wang, J. Xiao, and L. Ping, “Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems,” *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 4–11, 2006.
- [34] P. D. Diamantoulakis, K. N. Pappi, Z. Ding, and G. K. Karagiannidis, “Wireless-powered communications with non-orthogonal multiple access,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8422–8436, 2016.
- [35] Z. Ding, P. Fan, and H. V. Poor, “Impact of user pairing on 5g nonorthogonal multiple-access downlink transmissions,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2016.
- [36] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [37] S. Boyd, S.-J. Kim, L. Vandenberghe, and et al., “A tutorial on geometric programming,” *Optimization and Engineering*, vol. 8, pp. 67–127, 2007. [Online]. Available: <https://doi.org/10.1007/s11081-007-9001-7>
- [38] H. A. Le Thi and T. Pham Dinh, “Dc programming in communication systems: challenging problems and methods,” *Vietnam Journal of Computer Science*, vol. 1, pp. 15–28, 2014. [Online]. Available: <https://doi.org/10.1007/s40595-013-0010-5>
- [39] N. Ye, H. Han, L. Zhao, and A.-h. Wang, “Uplink nonorthogonal multiple access technologies toward 5g: A survey,” *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
- [40] Z. Ding, P. Fan, and H. V. Poor, “Impact of non-orthogonal multiple access on the offloading of mobile edge computing,” *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 375–390, 2019.
- [41] A. Kiani and N. Ansari, “Edge computing aware noma for 5g networks,” *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, 2018.
- [42] J. Du, W. Liu, G. Lu, J. Jiang, D. Zhai, F. R. Yu, and Z. Ding, “When mobile-edge computing (mec) meets nonorthogonal multiple access (noma) for the internet of things (iot): System design and optimization,” *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7849–7862, 2021.
- [43] Z. Ding, D. W. K. Ng, R. Schober, and H. V. Poor, “Delay minimization for noma-mec offloading,” *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1875–1879, 2018.
- [44] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, “Joint power and time allocation for noma-mec offloading,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6207–6211, 2019.

- [45] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, “Mobile-edge computing: Partial computation offloading using dynamic voltage scaling,” *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4268–4282, 2016.
- [46] Y. Pan, M. Chen, Z. Yang, N. Huang, and M. Shikh-Bahaei, “Energy-efficient noma-based mobile edge computing offloading,” *IEEE Communications Letters*, vol. 23, no. 2, pp. 310–313, 2019.
- [47] H. Li, F. Fang, and Z. Ding, “Joint resource allocation for hybrid noma-assisted mec in 6g networks,” *Digital Communications and Networks*, vol. 6, no. 3, pp. 241–252, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352864819304274>
- [48] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis, “Optimal task partition and power allocation for mobile edge computing with noma,” in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.
- [49] Y. Wu, L. P. Qian, K. Ni, C. Zhang, and X. Shen, “Delay-minimization nonorthogonal multiple access enabled multi-user mobile edge computation offloading,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 392–407, 2019.
- [50] X. Lyu, H. Tian, C. Sengul, and P. Zhang, “Multiuser joint task offloading and resource optimization in proximate clouds,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3435–3447, 2017.
- [51] X. Diao, J. Zheng, Y. Wu, and Y. Cai, “Joint computing resource, power, and channel allocations for d2d-assisted and noma-based mobile edge computing,” *IEEE Access*, vol. 7, pp. 9243–9257, 2019.
- [52] F. Wang, J. Xu, and Z. Ding, “Multi-antenna noma for computation offloading in multiuser mobile edge computing systems,” *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2450–2463, 2019.