UNIVERSITY OF PIRAEUS
School of Information and
Communication Technologies
Department of Digital Systems

PH.D. THESIS

"Managing networks and services based on artificial intelligence in heterogeneous broadband environments for 5th generation and beyond"

BELIKAIDIS IOANNIS-PRODROMOS

Diplomate Engineer from the Department of Electrical and Computer Engineering of the Polytechnic School of the University of Thessaly

Master's degree from the Department of Electrical and Computer Engineering of the Polytechnic School of the University of Thessaly

Piraeus, 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Σχολή Τεχνολογιών
Πληροφορικής και Επικοινωνιών
Τμήμα Ψηφιακών Συστημάτων

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

" Διαχείριση δικτύων και υπηρεσιών βάσει τεχνητής νοημοσύνης σε ετερογενή ευρυζωνικά περιβάλλοντα 5ης γενιάς και ακόλουθων γενεών"

ΜΠΕΛΙΚΑΪΔΗΣ ΙΩΑΝΝΗΣ-ΠΡΟΔΡΟΜΟΣ

Πτυχιούχος Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών της Πολυτεχνικής Σχολής Πανεπιστημίου Θεσσαλίας

Κάτοχος Μεταπτυχιακού τίτλου σπουδών Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών της Πολυτεχνικής Σχολής Πανεπιστημίου Θεσσαλίας

Πειραιάς, 2021

UNIVERSITY OF PIRAEUS **ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Μπελικαΐδης Ιωάννης-Πρόδρομος

Πτυχιούχος Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών της Πολυτεχνικής Σχολής Πανεπιστημίου Θεσσαλίας

Κάτοχος Μεταπτυχιακού τίτλου σπουδών Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών της Πολυτεχνικής Σχολής Πανεπιστημίου Θεσσαλίας

ΔΙΑΧΕΙΡΙΣΗ ΔΙΚΤΥΩΝ ΚΑΙ ΥΠΗΡΕΣΙΩΝ ΒΑΣΕΙ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΕ ΕΤΕΡΟΓΕΝΗ ΕΥΡΥΖΩΝΙΚΑ ΠΕΡΙΒΑΛΛΟΝΤΑ 5ΗΣ ΓΕΝΙΑΣ ΚΑΙ ΑΚΟΛΟΥΘΩΝ ΓΕΝΕΩΝ

MANAGING NETWORKS AND SERVICES BASED ON ARTIFICIAL INTELLIGENCE IN HETEROGENEOUS BROADBAND ENVIRONMENTS FOR 5TH GENERATION AND BEYOND

Συμβουλευτική Επιτροπή:

Π. Δεμέστιχας, Καθηγητής Παν. Πειραιώς
Α. Ρούσκας, Καθηγητής Παν. Πειραιώς
Α. Μηλιώνης, Αναπληρωτής Καθηγητής Παν. Πειραιώς

Εγκρίθηκε από την επταμελή Εξεταστική Επιτροπή την 16η Ιουλίου 2021,

Π. Δεμέστιχας
Καθηγητής Πανεπιστημίου Πειραιώς

Α. Ρούσκας
Καθηγητής Πανεπιστημίου Πειραιώς

Α. Μηλιώνης
Αναπληρωτής Καθηγητής Πανεπιστημίου Πειραιώς

Α. Κανάτας
Καθηγητής Πανεπιστημίου Πειραιώς

Μ. Λούτα
Αναπληρώτρια Καθηγήτρια ΠΔΜ

Ν. Μήτρου
Καθηγητής ΕΜΠ

Σ. Παπαβασιλείου
Καθηγητής ΕΜΠ

Table of Contents

List of Figures

List of Tables

Acknowledgments

I offer my sincerest gratitude to Dr Panagiotis Demestichas, Dr Angelos Rouskas, Dr Apostolos Meliones, who have supported me throughout my thesis with patience, guidance and knowledge whilst allowing me the room to work in my own way.

In addition, I would like to thank my colleague Dr Andreas Georgakopoulos for the great collaboration that we had and all help he provided from day one.

A special thanks to my university and my department's faculty, staff and fellow students for their valuable assistance whenever needed and for creating a pleasant and creative environment during my studies.

Last but not least, I wish to thank my family and friends for their unconditional support and encouragement all these years.

"If we knew what it was we were doing,

it would not be called research,

would it?"

— Albert Einstein

Abstract

The motivation of this research activity comes from the exponential increase in data demand observed in recent years as well as the approximation of the theoretical limits of network capacity. At the same time the process of licensing and acquiring new frequencies is expensive, time consuming and in some cases licensing is impossible for many providers. In combination with the support of the large volume of data caused by the new services and applications, it becomes necessary to research, implement and test the various solutions proposed for the optimal management of networks. The dissertation is structured in chapters and each chapter provides a detailed description and results of the conducted research activities for achieving the overall goal of managing networks and services based on artificial intelligence in heterogeneous broadband environments for 5th generation and beyond. Chapter 1 provides the main introduction and motivation of our work and sets the requirements of the necessary research for managing networks and services based on artificial intelligence. Chapter 2 investigates and evaluates mesh networks for achieving enhanced performance, even when connectivity is challenging. Chapter 3 focuses on the modeling and analysis of management for heterogeneous infrastructure. Chapter 4 elaborates on radio frequency resource management in a multi-provider environment with emphasis on hierarchical radio resource management scheme. Chapter 5 discusses the notion of network slicing and how slicing can lead to better service provisioning in demanding environments by blending different traffic types (e.g. URLLC, eMBB etc.). Chapter 6 provides useful insights on sharing and allocation of resources with emphasis on dynamic channel assignment for the new 5G services such as URLLC, eMBB and mMTC. Chapter 7 elaborates on RRM issues in a multi-connection environment with emphasis on 5G component carrier management and is being evaluated with system level simulations. In addition, the requirements of ultra-high connection density of IoT devices in mMTC environments are evaluated. The research of the proposed solutions resulted in various publications, conference papers, journals, books and in standards such as ITU for a study regarding evaluation of 5G. The last chapter provides the conclusions of this dissertation and suggestions for future research. Specifically, this thesis presents an in-depth analysis of solutions for better management of networks and services based on intelligent algorithms and techniques in heterogeneous broadband environments of 5th and next generations, taking into account the current situation and new challenges of networks. The aim was to analyze, propose and improve the state of the art techniques to bring 5G networks one step closer to what we think it could be. D2D / M2M, Mesh Networks, Ultra-densification, Dynamic selection channel, Network sharing, Network slicing, Multi-connectivity / Multi link, Carrier Aggregation and the utilization of new frequencies such as 3.5 GHz and narrow band are some of the techniques that were analyzed, improved and presented, achieving good results in all tested cases. These proposals could be further tested in a real environment and formulated by the research community and standards to be integrated into the new versions of 3GPP for 5G networks and future generations.

Απόσπασμα

Τα κίνητρα αυτής της ερευνητικής δραστηριότητας προέρχονται από το γεγονός της εκθετικής αύξησης της ζήτησης δεδομένων που παρατηρείται τα τελευταία χρόνια καθώς και στην προσέγγιση των θεωρητικών ορίων του capacity των δικτύων. Ταυτόχρονα η διαδικασία της αδειοδότησης και της απόκτησης νέων συχνοτήτων είναι ακριβή, χρονοβόρα και σε ορισμένες περιπτώσεις η αδειοδότηση είναι αδύνατη για πολλούς παρόχους. Σε συνδυασμό με την υποστήριξη του μεγάλου όγκου δεδομένων που προκαλούνται από τις καινούριες υπηρεσίες και εφαρμογές, πραγματοποιήθηκε έρευνα, βρέθηκαν, υλοποιήθηκαν και δοκιμάστηκαν λύσεις για την βέλτιστη διαχείριση των δικτύων. Η παρούσα διπλωματική εργασία διαρθρώνεται σε κεφάλαια και κάθε κεφάλαιο παρέχει μια λεπτομερή περιγραφή και τα αποτελέσματα των διεξαγόμενων ερευνητικών δραστηριοτήτων για την επίτευξη του γενικού στόχου διαχείρισης δικτύων και υπηρεσιών βασισμένων στην τεχνητή νοημοσύνη σε ετερογενή ευρυζωνικά περιβάλλοντα για 5η γενιά και μετά. Το Κεφάλαιο 1 παρέχει την κύρια εισαγωγή και το κίνητρο της εργασίας μας και θέτει τις απαιτήσεις της απαραίτητης έρευνας για τη διαχείριση δικτύων και υπηρεσιών που βασίζονται στην τεχνητή νοημοσύνη. Το Κεφάλαιο 2 διερευνά και αξιολογεί δίκτυα πλέγματος για την επίτευξη βελτιωμένης απόδοσης, ακόμη και όταν η συνδεσιμότητα είναι δύσκολο να πραγματοποιηθεί. Το Κεφάλαιο 3 επικεντρώνεται στη μοντελοποίηση και ανάλυση της διαχείρισης για ετερογενείς υποδομές. Το Κεφάλαιο 4 επεξεργάζεται τη διαχείριση πόρων ραδιοσυχνοτήτων σε περιβάλλον πολλαπλών παρόχων, με έμφαση στο ιεραρχικό σχήμα διαχείρισης πόρων ραδιοφώνου. Το Κεφάλαιο 5 αναλύει την έννοια του τεμαχισμού δικτύου και πώς ο τεμαχισμός μπορεί να οδηγήσει σε καλύτερη παροχή υπηρεσιών σε απαιτητικά περιβάλλοντα συνδυάζοντας διαφορετικούς τύπους επισκεψιμότητας (π.χ. URLLC, eMBB, mMTC, κ.λπ.). Το Κεφάλαιο 6 παρέχει χρήσιμες πληροφορίες σχετικά με την κοινή χρήση και την κατανομή πόρων με έμφαση στην εκχώρηση δυναμικού καναλιού για τις νέες υπηρεσίες 5G όπως το URLLC, το eMBB και το mMTC. Το Κεφάλαιο 7 επεξεργάζεται θέματα RRM σε περιβάλλον πολλαπλών συνδέσεων με έμφαση στη διαχείριση φορέα 5G και αξιολογείται με προσομοιώσεις σε επίπεδο συστήματος. Επιπλέον, αξιολογούνται οι απαιτήσεις για εξαιρετικά υψηλή πυκνότητα σύνδεσης συσκευών IoT σε περιβάλλον mMTC. Η έρευνα των προτεινόμενων λύσεων κατέληξε σε διάφορες δημοσιεύσεις, εργασίες συνεδρίων, περιοδικά, βιβλία και σε πρότυπα όπως το ITU για μελέτη σχετικά με την αξιολόγηση του 5G. Το τελευταίο κεφάλαιο παρέχει το συμπέρασμα αυτής της διπλωματικής εργασίας και προτάσεις για μελλοντική έρευνα. Συγκεκριμένα αυτή η διατριβή παρείχε μια σε βάθος ανάλυση λύσεων για τη καλύτερη διαχείριση δικτύων και υπηρεσιών που βασίζονται σε ευφυείς αλγόριθμους και τεχνικές σε ετερογενή ευρυζωνικά περιβάλλοντα 5ης και επόμενων γενεών, λαμβάνοντας υπόψη την τρέχουσα κατάσταση και προσβλέποντας τις νέες προκλήσεις των δικτύων. Στόχος ήταν να αναλύσουμε, να προτείνουμε και να βελτιώσουμε τις state-of-the-art τεχνικές για να φέρουμε τα δίκτυα 5G ένα βήμα πιο κοντά σε αυτό που θεωρούμε ότι θα μπορούσε να είναι. Οι D2D / M2M, Mesh Networks, Ultra-densification, Dynamic channel selection, Network sharing, Network slicing, Multi-connectivity / Multi link, Carrier Aggregation και χρήση νέων συχνοτήτων 3.5GHz και narrowband είναι μερικές από τις καινοτομίες που αναλύθηκαν και παρουσιάστηκαν σε αυτή τη διατριβή πετυχαίνοντας πολύ καλά αποτελέσματα σε όλες τις περιπτώσεις που δοκιμάστηκαν. Οι λύσεις και οι τεχνικές που προτείνονται σε αυτή τη διατριβή θα μπορούσαν να δοκιμαστούν σε πραγματικά περιβάλλοντα και να διαμορφωθούν από την ερευνητική κοινότητα ώστε να ενσωματωθούν στις νέες εκδόσεις της 3GPP για δίκτυα 5G αλλά και επόμενων γενεών.

Περίληψη

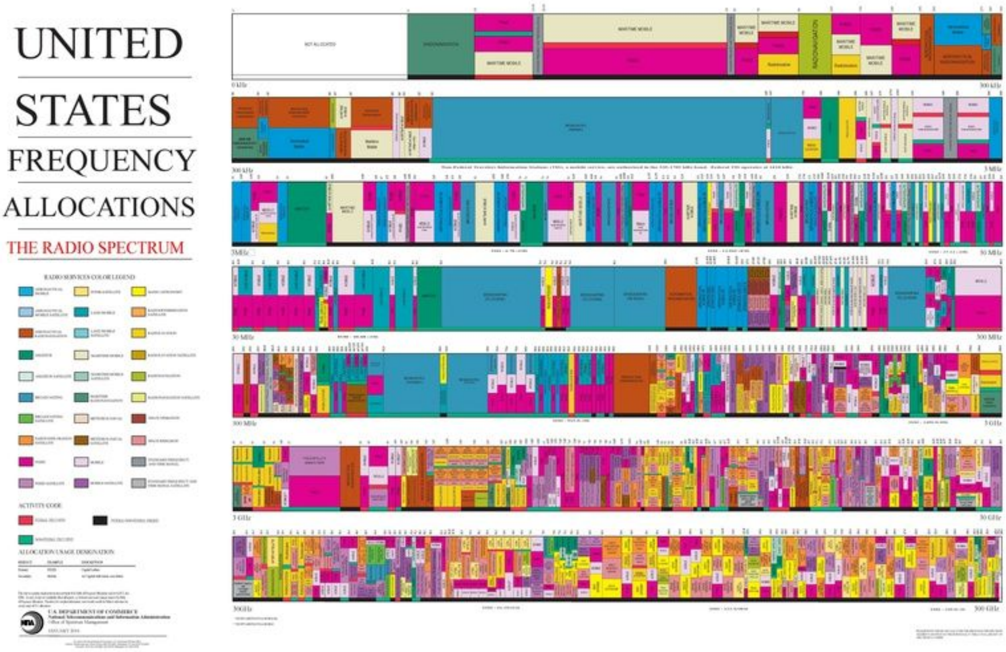
Περιεχόμενα

Εικόνες

1 Εισαγωγή

Τις τελευταίες δεκαετίες, ο κόσμος μας έχει παρακολουθήσει μια επανάσταση στην εξέλιξη των συστημάτων κινητής επικοινωνίας που ξεκίνησε από τα πρώτα δίκτυα επικοινωνίας πρώτης γενιάς προς τα δίκτυα δεύτερης, τρίτης και τέταρτης γενιάς. Αυτή η εξέλιξη υπήρξε μια σημαντική κινητήρια δύναμη πίσω από την πρόοδο και την ανάπτυξη του κόσμου μας σε ένα προηγμένο και δικτυωμένο περιβάλλον. Η κινητή επικοινωνία έχει εξελιχθεί από μια υπηρεσία που είναι διαθέσιμη και προσιτή μόνο για λίγους ανθρώπους σε μια υπηρεσία που χρησιμοποιείται από την πλειοψηφία του παγκόσμιου πληθυσμού. Εκατομμύρια άνθρωποι σε όλο τον κόσμο αλληλοσυνδέονται μέσω της κινητής επικοινωνίας. Έχει γίνει αναπόσπαστο κομμάτι της καθημερινής μας ζωής και οι εφαρμογές της είναι πανταχού παρούσες σε κάθε καθημερινή δραστηριότητα. Έχει πρωτοστατήσει σε πολλές εξελίξεις και έχει να δείξει επιτεύγματα σε διάφορους τομείς όπως η υγειονομική περίθαλψη, οι μεταφορές, η ενέργεια, η μεταποίηση, η αρχιτεκτονική, η γεωργία, η μηχανική, οι επιχειρήσεις, η εκπαίδευση, η μετεωρολογία, η ραδιοφωνία, τα μέσα μαζικής ενημέρωσης και η ψυχαγωγία. Ταχεία ανάπτυξη στην παγκόσμια αγορά συσκευών κινητής τηλεφωνίας έχει παρατηρηθεί τα τελευταία χρόνια. Επιπλέον, η ανάπτυξη αυτή γίνεται συνεχώς αισθητή και ο κόσμος θα συνεχίσει να απολαμβάνει τον πολλαπλασιασμό συσκευών κινητής επικοινωνίας, όπως smartphones, tablet, φορητών συσκευών, και φορητών υπολογιστών, μαζί με νέες και υπάρχουσες υπηρεσίες και εφαρμογές που παρέχονται από συστήματα κινητής επικοινωνίας όπως φωνητική κλήση, τηλεδιάσκεψη, online gaming, ζωντανή ροή βίντεο και πολλά άλλα. Στην ψηφιακή εποχή, οι χρήστες και οι συσκευές εξαρτώνται όλο και περισσότερο από διάφορες εφαρμογές και υπηρεσίες που περιλαμβάνουν τη δημιουργία, την πρόσβαση / επικοινωνία, την επεξεργασία και την αποθήκευση ψηφιακού περιεχομένου. Αυτές οι εξελίξεις επιταχύνθηκαν έντονα από τις ασύρματες / κινητές τεχνολογίες, οι οποίες προσέφεραν ασύγκριτες ευκαιρίες πρόσβασης / επικοινωνίας στους χρήστες. Το 5G αναμένεται να κυριαρχεί κυρίως από τις ακόλουθες κατηγορίες εφαρμογών: μαζική επικοινωνία τύπου μηχανής (mMTC), ενισχυμένη επικοινωνία κινητής ευζωνικής σύνδεσης (eMBB), εξαιρετικά αξιόπιστη επικοινωνία και σύνδεση χαμηλής καθυστέρησης (URLLC). Αυτές οι κύριες κατηγορίες θα διευκολύνουν τα σενάρια που σχετίζονται με κρίσιμες και απαιτητικές εφαρμογές για την υλοποίηση έξυπνων πόλεων, καθώς και την υλοποίηση εφαρμογών για τη βιομηχανία 4.0 και τις πτυχές αυτοματισμού. Επίσης, αναμένεται να υπάρξουν αυστηρές απαιτήσεις, προκειμένου να εξασφαλιστεί αξιόπιστη και ασφαλής εξυπηρέτηση με πολύ υψηλό ποσοστό διαθεσιμότητας.

Η άδεια εκπομπής είναι ένας τύπος άδειας ραδιοφάσματος που παρέχει στον κάτοχο της άδειας, άδεια χρήσης ενός μέρους του φάσματος ραδιοσυχνοτήτων σε μια δεδομένη γεωγραφική περιοχή για σκοπούς μετάδοσης. Οι άδειες περιλαμβάνουν γενικά περιορισμούς, οι οποίοι διαφέρουν από ζώνη σε ζώνη [1]. Το φάσμα μπορεί να διαιρεθεί ανάλογα με τη χρήση. Η μετάδοση χωρίς άδεια αναφέρεται σε νόμιμες συσκευές που επιτρέπεται να μεταδίδουν με χαμηλή ισχύ χωρίς άδεια και πειρατικούς σταθμούς, οι οποίοι παραβιάζουν το νόμο.



Εικόνα 1: Διάγραμμα κατανομών συχνότητας ΗΠΑ 2016.[2]

Η διαδικασία απόκτησης νέας άδειας μετάδοσης είναι χρονοβόρα. Υπάρχει ένας περιορισμένος χρόνος για την άδεια, μόλις αποκτηθεί και κρατάει περίπου πέντε χρόνια.

Με βάση την παραπάνω περιγραφή αδειοδότησης και της χρήσης των συχνοτήτων από τηλεοπτικούς σταθμούς, ραδιοφωνικούς σταθμούς, δορυφορικούς σταθμούς, στρατό, ναυτικό, αεροπορία και άλλους κρατικούς μηχανισμούς, μπορεί κάποιος να κατανοήσει το μεγάλο πρόβλημα στη χρήση των συχνοτήτων από τηλεπικοινωνιακούς παρόχους.

Για να καλυφθούν οι προδιαγραφές που έχουν οριστεί, οι τηλεπικοινωνιακοί πάροχοι θα χρειαστεί να δεσμεύσουν περαιτέρω πόρους σε τηλεπικοινωνιακό φάσμα. Ωστόσο, η επιτυχία των υπηρεσιών εξαρτάται σε μεγάλο βαθμό από τις κυβερνήσεις και ρυθμιστικές αρχές. Η ταχύτητα, η απόσταση και η ποιότητα των υπηρεσιών εξαρτώνται από την υποστήριξη των αρχών αυτών με έγκαιρη πρόσβαση στο σωστό ποσό και τύπο προσιτού φάσματος, υπό τις σωστές συνθήκες. Διακύμανση στο ποσό του εκχωρηθέντος φάσματος και των τιμών που πληρώνονται, σημαίνει ότι οι υπηρεσίες θα διαφέρουν μεταξύ τους, από τύπο σε τύπο, από περιοχή σε περιοχή, αλλά και ανάμεσα σε χώρες. Αυτό, με τη σειρά του άμεσα επηρεάζει την ανταγωνιστικότητα αλλά και την οικονομία.

Η εκθετική αύξηση της ζήτησης δεδομένων στα ασύρματα δίκτυα και η προσέγγιση των θεωρητικών ορίων στη χωρητικότητα των ασυρμάτων συνδέσεων, μας υποχρεώνει να βρούμε νέες λύσεις και καινοτόμα σχέδια για να διαχειριστούμε την τεράστια κυκλοφορία δεδομένων. Τα ετερογενή δίκτυα (HetNets) θα μπορέσουν να προσφέρουν μία αποτελεσματική λύση στο πρόβλημα της χωρητικότητας, χρησιμοποιώντας ταυτόχρονα φάσμα από διάφορες περιοχές ραδιοσυχνοτήτων, πρέπει όμως να λειτουργήσουν με σωστό τρόπο ώστε να έχουν το επιθυμητό αποτέλεσμα. Τα ετερογενή δίκτυα 5^{ης} γενιάς θα παρέχουν επαρκή αύξηση της χωρητικότητας καθώς θα χρησιμοποιούν πολυεπίπεδη αρχιτεκτονική αποτελούμενη από πολλαπλών επιπέδων κυψέλες (cells) (π.χ. macro cell, small cell, κτλπ), αναμετάδοση από συσκευή σε συσκευή (device to device) μέσω δικτύων πλέγματος, νέα δυναμικά συστήματα πρόσβασης φάσματος με διαφορετικούς χρήστες όπου θα κάνουν χρήση κοινών ραδιοσυχνοτήτων, ευέλικτη διαχείριση πόρων ραδιοφώνου και πολλά ακόμα. Αυτή η πολυεπίπεδη αρχιτεκτονική για την αύξηση την χωρητικότητας δημιουργεί ορισμένες προκλήσεις στα HetNets που πρέπει να μελετηθούν, να αναλυθούν και να βρεθούν κατάλληλες λύσεις στα προβλήματα που θα προκύψουν κατά τη χρήση τους. Για παράδειγμα οι παρεμβολές μεταξύ κυψελών και χρηστών παραμένει η μεγαλύτερη πρόκληση στα HetNets, προσφέροντας κατάλληλο έδαφος για ουσιαστική μελέτη και ανάπτυξη υπερσύγχρονων HetNets. Αποτελεσματικές τεχνικές και στρατηγικές θα επιτρέψουν την επιπρόσθετη και ουσιαστική αύξηση στη χωρητικότητα.

Ένας από τους προτεινόμενους τρόπους για την απόκτηση πρόσβασης υψηλής ποιότητας 5G σε μεγαλύτερο αριθμό ατόμων είναι μέσω δικτύωσης πλέγματος. Η δικτυωτή και ευέλικτη δικτύωση θα παρέχει την ευκαιρία για σύνδεση διαφορετικών συσκευών. Τα δίκτυα 5G mesh θα αλλάξουν τον τρόπο που σκεφτόμαστε τη συνδεσιμότητα και την επικοινωνία.

Χρήση νέων συχνοτήτων που είχαν δεσμευτεί στο παρελθόν για στρατιωτικούς ή άλλους σκοπούς. Η χρήση αυτή θα πραγματοποιείται με συστήματα πρόσβασης φάσματος (SAS) με χρήση του Citizens Broadband Radio Service (CBRS) αρχικά στο φάσμα των 3,5 GHz, το οποίο αντιπροσωπεύει 150 MHz φάσματος στη ζώνη 3,5 GHz, όπου μέχρι τώρα χρησιμοποιούνταν μόνο από το στρατό, ναυτικό, αεροπορία και μελλοντικά θα χρησιμοποιηθεί και φάσμα άνω των 25GHz.

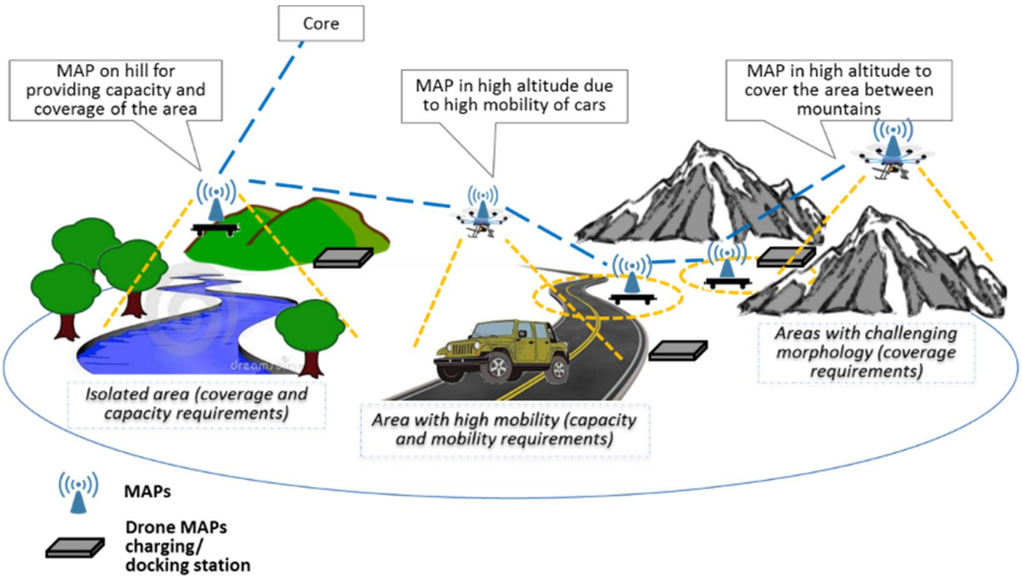
Πολλαπλοί μηχανισμοί και μέθοδοι θα πρέπει εξεταστούν για να μπορέσουμε να πετύχουμε την καλύτερη χρήση των συχνοτήτων που υπάρχουν. Πολλαπλές συνδέσεις, οικοσύστημα δικτύου πολλαπλών επιπέδων, νέες μέθοδοι ελέγχου και ενσωμάτωσης επιπέδου χρήστη και προσαρμογής πρωτοκόλλου μπορούν να προσφέρουν ευκαιρίες για επανασχεδιασμό ή βελτίωση διαφόρων λειτουργιών, (π.χ. διαχείριση παρεμβολών, έλεγχος ισχύος, έλεγχος RAN) αναθέτοντας τις υπηρεσίες και τις εφαρμογές στους κατάλληλους πόρους για την πλήρωση των απαιτήσεων με αποτελεσματικό και βιώσιμο τρόπο. Carrier Aggregation και Advanced RAN Coordination είναι μια συνδυασμένη λύση που βελτιστοποιεί την κάλυψη, τη χωρητικότητα και τον λανθάνοντα χρόνο των 5G μεσαίων και υψηλών ζωνών. Μαζί, αυτές οι λύσεις επιτρέπουν στους παρόχους υπηρεσιών να βελτιστοποιήσουν τη χρήση των στοιχείων του φάσματος τους κατά την ανάπτυξη 5G. Ένα καλύτερο δίκτυο 5G θα παρέχει σε περισσότερους συνδρομητές υψηλότερες ταχύτητες δεδομένων, επιτρέποντας ταυτόχρονα πλήθος νέων εφαρμογών χαμηλού λανθάνοντος χρόνου.

Ο τεμαχισμός δικτύου 5G μπορεί να παρέχει σε επιχειρήσεις τις εξατομικευμένες δυνατότητες που χρειάζονται για να τους επιτρέψουν να λειτουργήσουν, αλλά χρησιμοποιεί την ίδια φυσική υποδομή. Ο τεμαχισμός δικτύου θα επιτρέπει την εκτέλεση πολλαπλών, εξατομικευμένων δικτύων χρησιμοποιώντας την ίδια κοινόχρηστη υποδομή. Τα δίκτυα μπορούν να προσφέρουν διαφορετικές λειτουργίες, ειδικά προσαρμοσμένες σε άτομα ή υπηρεσίες, χωρίς να χρειάζονται ξεχωριστές δομές δικτύου για κάθε μία. Το ένα δίκτυο κινητής τηλεφωνίας ουσιαστικά «χωρίζεται» σε διαφορετικά προσαρμόσιμα μέρη που ικανοποιούν τις ανάγκες διαφορετικών καταναλωτών. Αντί οι εταιρίες, προγραμματιστές και καταναλωτές να πρέπει να προσαρμοστούν σε αυτά που προσφέρουν τα δίκτυα κινητής τηλεφωνίας, το δίκτυο κινητής τηλεφωνίας 5G θα πρέπει να είναι σε θέση να προσαρμοστεί στις ανάγκες των καταναλωτών.

Όλοι αυτοί οι μέθοδοι και ακόμα περισσότεροι, ερευνήθηκαν, αναλύθηκαν και δοκιμάστηκαν κατά τη διάρκεια της ερευνητικής δραστηριότητας δίνοντας αποτελέσματα με βάση τους τρόπους χρήσης τους. Στα επόμενα κεφάλαια παρουσιάζονται με λεπτομέρεια οι μεθοδολογίες, οι αλγόριθμοι αλλά και τα αποτελέσματα τα οποία καταγράφηκαν και παρουσιάστηκαν σε άρθρα, περιοδικά και βιβλία.

2 Δίκτυα πλέγματος

Κατά την διάρκεια της έρευνας αναπτύχθηκαν και μελετήθηκαν θέματα σχετικά με τον καθορισμό του τρόπου επιλογής μονοπατιών για τη δημιουργία ευέλικτων δικτύων που σχηματίζουν πλέγμα με σκοπό την επανασύνδεση της επικοινωνίας σε προβληματικές περιοχές. Τέτοιες περιοχές θα μπορούσαν να είναι είτε απομακρυσμένες ζώνες όπου δεν υπάρχει εγκατεστημένο δίκτυο τηλεπικοινωνιών είτε περιοχές όπου έχουν υπάρξει καταστροφές στις τηλεπικοινωνιακές δομές λόγο σεισμών, ή άλλων φυσικών καταστροφών είτε ακόμα και σε περιπτώσεις όπου χρειάζεται η πληροφορία να συλλέγεται σε ένα κόμβο και να στέλνετε από αυτόν και μόνο για μικρότερη κατανάλωση στο δίκτυο πλέγματος. Σε αυτές τις περιπτώσεις ένα δίκτυο από κινούμενους κόμβους, μικρά αυτοκινούμενα οχήματα με ρόδες για το έδαφος ή ακόμα και ιπτάμενα οχήματα (drones) θα μπορούσαν να χρησιμοποιηθούν για να λύσουν το πρόβλημα. Αυτά τα οχήματα (MAP) θα ήταν δυνατόν να κινηθούν προς τα κατάλληλα σημεία ώστε να δημιουργήσουν συνδέσεις μεταξύ τους και να μεταφέρουν πληροφορίες από ένα σημείο στο άλλο, αντικαθιστώντας τις κατεστραμμένες τηλεπικοινωνιακές δομές που υπήρχαν στην περιοχή .



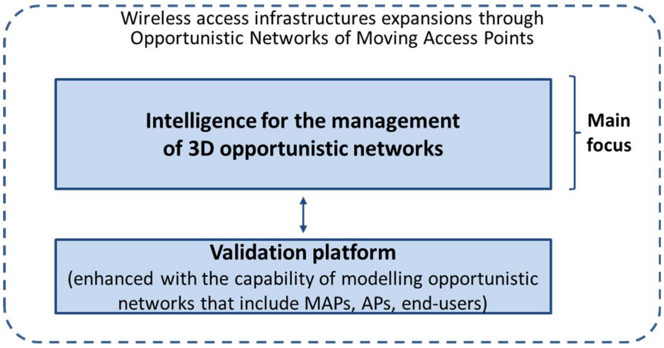
Εικόνα 2: Σενάριο όπου τα ευέλικτα δίκτυα καλύπτουν μία περιοχή όπου δεν υπήρχε τηλεπικοινωνιακή υποδομή.

Σε αυτή την έρευνα αναλύθηκε και μοντελοποιήθηκε η συνδεσιμότητα αυτών των οντοτήτων ανάλογα με τους πόρους που είχαν όπως, μπαταρία, εμβέλεια κάλυψης σήματος, δυνατότητα μετακίνησής τους σε άλλη περιοχή και πολλά ακόμα. Η ανακάλυψη των καλύτερων μονοπατιών έγινε με βάση δύο αλγορίθμους. Συγκεκριμένα χρησιμοποιήθηκαν οι αλγόριθμοι, Προσομοιωμένη Ανέλιξη (Simulated Annealing) και Μέγιστης Ροής (Maximum Flow).

Ο αλγόριθμος Simulated Annealing [3] προέρχεται από την ανάληψη της φυσικής ιδιότητας των μετάλλων. Συγκεκριμένα είναι η διαδικασία που χρησιμοποιείται στην μεταλλουργία για να μαλακώσουμε ή να σκληρύνουμε μέταλλα θερμαίνοντας τα σε υψηλή θερμοκρασία και στην συνέχεια ψύχοντάς τα σταδιακά, επιτρέποντάς έτσι το υλικό να στερεοποιηθεί σε μία κρυσταλλική κατάσταση χαμηλής ενέργειας.

Στη συνέχεια χρησιμοποιήθηκε ο αλγόριθμος Μέγιστης Ροής (Maximum Flow). Ο αλγόριθμος αυτός προσπαθεί να υπολογίσει τις διαδρομές ώστε να επιτευχθεί η μέγιστη ροή από ένα αρχικό σημείο σε ένα άλλο τελικό σημείο.

Με βάση τους παραπάνω αλγορίθμους θα πρέπει να διαχειριστούμε τους πόρους που υπάρχουν στο σύστημά μας με όσο το δυνατόν καλύτερο τρόπο. Για να γίνει αυτό χρησιμοποιήσαμε μία οντότητα όπου μπορεί και παρακολουθεί ποιος αλγόριθμος είναι καλύτερος ανά πάσα στιγμή και να επιλέξει το αντίστοιχο μονοπάτι. Αυτή είναι μία έξυπνη οντότητα που επικυρώνει κάθε φορά αν οι πόροι χρησιμοποιήθηκαν με σωστό τρόπο.



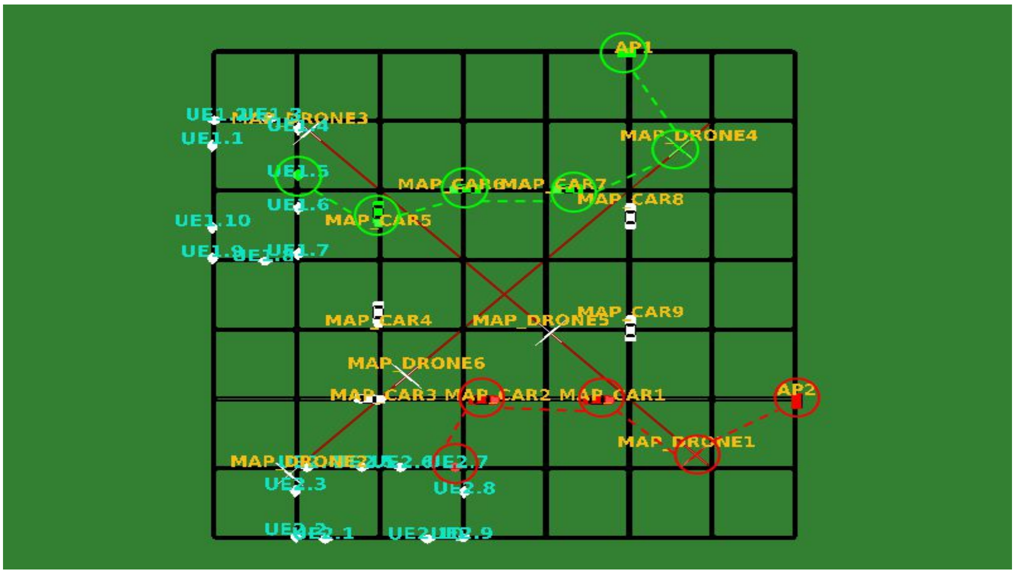
Εικόνα 3: Έξυπνη διαχειριστική οντότητα.

Για την αξιολόγηση των μοντέλων και αλγορίθμων πραγματοποιήθηκε με τη χρήση τριών προσομοιωτών του VEINS, OMNET++ και SUMO [4][5][6].

Κάποιες από τις ρυθμίσεις προσομοίωσης των αλγορίθμων είναι:

- Ο χάρτης έχει δημιουργηθεί σε εξομοιωτή SUMO 7 x 7 πλέγμα με δύο λωρίδες προς κάθε κατεύθυνση. Δύο διαγώνιοι αεροδιάδρομοι (εικονογραφημένοι με κόκκινο χρώμα) που χρησιμοποιούνται από το Drones για να παρέχουν καλύτερη κάλυψη σε καταστάσεις όπου τα επίγεια rovers δεν μπορούν.
- Μια διασταύρωση από την άλλη απέχει 100 μέτρα, δίνοντας συνολική έκταση 4900 τετραγωνικών μέτρων.
- Οι UE, MAPs και APs έχουν εύρος κάλυψης εύρους εκκίνησης ασύρματης μετάδοσης περίπου 150, 200 και 250 μέτρων αντίστοιχα (αναφέρεται ως $1^{η}$ περίπτωση) και στη συνέχεια η σειρά αυτή επεκτείνεται
- Για τα UE και MAP (εκτός από τα Drones) έχει ρυθμιστεί μέγιστη ταχύτητα 5 m/s, δηλ. Περίπου 18 km/h.
- Τα drones που πετούν πάνω από το έδαφος έχουν ταχύτητες περίπου 36 km/h (10 m/s), αλλά μειώνουν την ταχύτητά τους όταν βρίσκονται κοντά στα AP ή UEs δίνοντας μεγαλύτερο χρόνο σύνδεσης με αυτές τις οντότητες και κινούνται με υψηλότερες ταχύτητές κατά τη διάρκεια των μεσαίων τμημάτων.

Όταν τα MAP έχουν μεγαλύτερο εύρος μετάδοσης, περισσότεροι κόμβοι μπορούν να επικοινωνούν μεταξύ τους με αποτέλεσμα να υπάρχουν πολλαπλές διαδρομές. Επιπλέον, σε αντίθεση με τα στατικά δίκτυα, στην προσομοίωση μας όλοι οι κόμβοι κινούνται συνεχώς αλλάζοντας την τοπολογία του δικτύου.



Εικόνα 4: Αποτελέσματα των δύο αλγορίθμων για τα διαφορετικά σενάρια που δοκιμάστηκαν

Η Εικόνα 4 είναι στιγμιότυπο από τον προσομοιωτή SUMO που δείχνει τα αποτελέσματα από τους αλγόριθμους MF και SA. Αυτές οι διαφορετικές διαδρομές / τοπολογίες που δημιουργούνται από τη MF και την SA παράγουν τις διαφορές που βλέπουμε στα αποτελέσματα και παρουσιάζονται με λεπτομέρεια στο αντίστοιχο κεφάλαιο αυτής της διατριβής.

3 Μοντελοποίηση και ανάλυση διαχείρισης ετερογενών υποδομών

Προκειμένου να ικανοποιηθούν οι απαιτήσεις 5G, έχουν εισαχθεί κάποιες αρχικές αλλά συγκεκριμένες περιπτώσεις χρήσης που μπορούν να υποστηριχθούν από τη χρήση και την εκμετάλλευση τηλεπικοινωνιακών συστημάτων μικρής εμβέλειας:

α) Μαζική χρήση IoT

β) Ασύρματη ευρυζωνικότητα

γ) Εξαιρετικά αξιόπιστες επικοινωνίες

δ) Κινητικότητα υψηλής ταχύτητας

Για την περίπτωση των Internet of Things (IoT), πρέπει να συνδεθούν με το δίκτυο τεράστιοι αριθμοί φορητών συσκευών, έξυπνοι μετρητές, αισθητήρες κλπ. Οι συσκευές αυτές μπορούν να τοποθετηθούν σε εσωτερικές ή εξωτερικές θέσεις, συνδεδεμένες σε small cells. Οι εξαιρετικά αξιόπιστες επικοινωνίες είναι μια πρόβλεψη μελλοντικών αναγκών σε επικοινωνίες που δεν υπάρχουν στα σημερινά 4G ασύρματα κυψελοειδή συστήματα. Αυτά τα συστήματα θα χρειαστούν σχεδόν 100% uptime υπηρεσίες για να υποστηρίξουν εφαρμογές όπως ασφάλεια/ έκτακτη ανάγκη, βιομηχανικό/ στρατιωτικό έλεγχο και διαχείριση, κλπ. Στην περίπτωση της ευρυζωνικής ασύρματης σύνδεσης, δημιουργείται μια μαζική ανάπτυξη μικρών κυψελών (small cells) για την παροχή μιας ομοιόμορφης ευρυζωνικής εμπειρίας στους χρήστες που απαιτούν υψηλό ρυθμό δεδομένων και περιορισμένη καθυστέρηση για την παροχή εφαρμογών όπως streaming πολυμέσων υψηλής ευκρίνειας, παιχνίδια, βιντεοκλήσεις και υπηρεσίες cloud. Επιπλέον, οι χρήστες που κινούνται από εσωτερικούς σε εξωτερικούς χώρους (και αντίστροφα) πρέπει να υποστηριχθούν. Σε περιβάλλοντα που χαρακτηρίζονται από κινητικότητα υψηλής ταχύτητας δεν θεωρείται αποτελεσματική λύση η χρήση μικρών κυψελών όπως γίνεται στις προηγούμενες περιπτώσεις χρήσης. Η χρησιμοποίηση πολύ μεγάλου αριθμού small cells μπορεί να επιφέρει επίσης μερικά αρνητικά αποτελέσματα, για παράδειγμα, μείωση της αποδοτικότητας φάσματος. Σημειώνεται ωστόσο, ότι οι απώλειες στην απόδοση του ραδιοφάσματος σε μικρότερη μεταξύ τους απόσταση (ISDs) (50m και κάτω) λόγω υψηλών επιπέδων παρεμβολής/SINR σε πολύ μικρά κελιά, η χωρητικότητα της περιοχής (παρά τις απώλειες) είναι σημαντικά υψηλότερα σε σύγκριση με μεγαλύτερα ISDs. Ένα small cell που έχει δυνατότητες αναγνώρισης του περιβάλλοντός του μπορεί να αποφύγει παρεμβολές χρησιμοποιώντας τη λειτουργικότητα της αίσθησης του φάσματος επιλέγοντας συγκεκριμένα κανάλια που δεν αντιμετωπίζουν παρεμβολές από το περιβάλλον, προκειμένου να αποφευχθεί η πρόκληση παρεμβολών από άλλα γειτονικά κελιά.

Τα small cells μπορούν να χρησιμοποιήσουν διαφορετικές συχνότητες και διαφορετικά συστήματα επικοινωνίας. Κάτι τέτοιο θα σήμαινε πως σε κάθε περίπτωση θα υπάρχουν διαφορετικοί πόροι και διαφορετικές συνθήκες με διάφορα αρνητικά και θετικά στοιχεία. Στα δίκτυα 5^{ης} γενιάς σε ένα ετερογενές περιβάλλον ένας χρήστης θα πρέπει να συνδεθεί και να ανταλλάξει πληροφορίες ανάλογα με τις ανάγκες του όπως αναφέρθηκαν προηγουμένως αλλά και ανάλογα με τους πόρους που διαθέτει. Αν μία συγκεκριμένη συχνότητα δεν είναι διαθέσιμη σε κάποιον χρήστη αλλά είναι στο base station της περιοχής του ή το αντίθετο, πρέπει να γίνουν οι ανάλογες κινήσεις ώστε να μην χρησιμοποιηθεί αυτή. Όπως μπορούμε να κατανοήσουμε προσπαθούμε να λύσουμε για ακόμα μία φορά το πρόβλημα των ευκαιριακών συνδέσεων ειδικά αν ο χρήστης κινείται και ζητά διαφορετικές υπηρεσίες από το τηλεπικοινωνιακό σύστημα. Μία διαχειριστική οντότητα θα πρέπει να αναλάβει τον ρόλο της δρομολόγησης και της εύρεσης της καλύτερης σύνδεσης χωρίς όμως την κατασπατάληση πόρων (over provisioning) που μπορεί να είναι σημαντικοί για άλλους χρήστες σε επόμενες χρονικές στιγμές. Κάθε στιγμή θα πρέπει να γίνεται όσο τον δυνατόν καλύτερος διαμοιρασμός των πόρων από το σύστημα στους χρήστες που όμως θα πρέπει να επιτυγχάνετε η επίτευξη της ποιότητας της σύνδεσης και μεταφοράς δεδομένων αλλά και της ποιότητα των υπηρεσιών που αντιλαμβάνεται ένας χρήστης.

Κάθε small cell θα πρέπει να έχει τη δυνατότητα να χρησιμοποιεί φάσμα είτε χωρίς άδεια είτε με άδεια για να αυξήσει την χωρητικότητα, την κάλυψη και την ισορροπία της κυκλοφορίας με κεντρική κατανομή διαφορετικών ζωνών για τη μείωση του επιπέδου παρεμβολών, λόγω του ότι επιτυγχάνονται τα όρια των αδειοδοτημένων πόρων. Επίσης, η εισαγωγή τεχνικών δυναμικής επιλογής καναλιών και διαχωρισμού καναλιών σε small cells με βάση τα επίπεδα παρεμβολής θα μπορεί να βελτιώσει την ποιότητα του καναλιού και έτσι να βελτιώσει την ποιότητα της εμπειρίας (QoE). Ακόμα η εξισορρόπηση φορτίου μεταξύ μιας ομάδας γειτονικών small cells, η διαχείριση των παρεμβολών με τη χρήση μη αδειοδοτημένου φάσματος ή κατανομής καναλιών είναι πολύ σημαντικά στοιχεία. Η ενσωμάτωση στα small cells Wi-Fi και LTE εντός του cellular system βοηθά την διαχειριστική μονάδα να βελτιστοποιήσει τη χρήση του δικτύου. Έτσι παρέχεται μια περαιτέρω βελτίωση στις επιδόσεις, δημιουργώντας μια απρόσκοπτη εμπειρία πολλαπλών RAT για τους συνδρομητές τους.

Ο Πίνακας 1 παρέχει μια συνοπτική περίληψη των κυριότερων δεικτών και απαιτήσεων που συνάδουν με τις τιμές στόχου του NGMN για παρόμοια σενάρια. Αυτά τα KPIs είναι σύμφωνα με ορισμένες προδιαγραφές που έχουν εισαχθεί από άλλους οργανισμούς (π.χ. ITU).

KPI	Requirements			
	Massive IoT communications	Mobile Broadband	Ultra-reliable communications	High-Speed mobility
User Experienced Data Rate	From tens to hundreds of Kbps	DL: 300 Mbps UL: 50 Mbps	DL: several kbps UL: several kbps	DL: 50Mbps UL: 25Mbps
E2E Latency	Order of seconds or more	10 ms	1 ms upwards	10 ms
Mobility	On demand	On demand, 0-100 km/h	Mostly static	On demand (up to 500km/)
Connection Density	Up to 200,000 devices/km^2	200-2500 /km^2	Mostly low, but variable	~2000/km^2
Traffic Density	Not critical	DL: 750 Gbps/km^2 UL: 125 Gbps/km^2	Highly variable	DL:100Gbps/km^2 UL: 50Gbps/km^2

Εικόνα 5: Απαιτήσεις Συστημάτων 5^{ης} γενιάς.

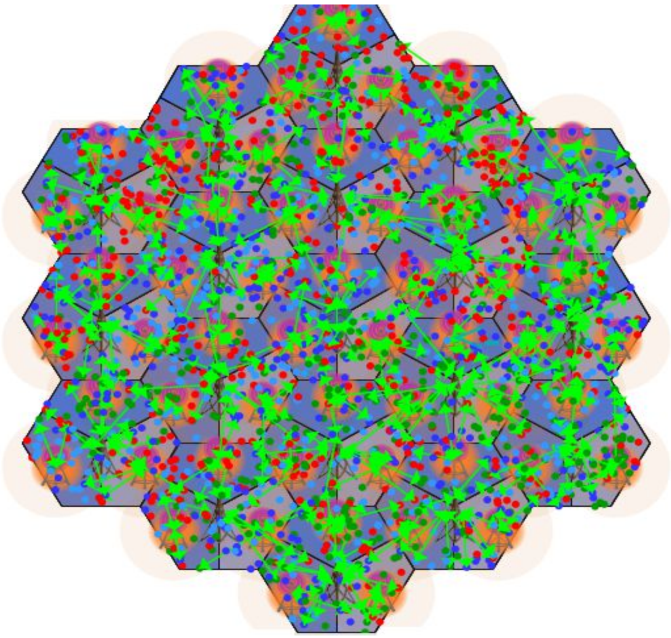
Ορισμένες από τις απαιτήσεις ενδέχεται να έχουν ήδη επιτευχθεί σε ορισμένες περιπτώσεις χρήσης, αλλά ορισμένες άλλες χρειάζονται πιο εμπεριστατωμένη έρευνα (π.χ. ρυθμός δεδομένων, υποστήριξη υπερβολικά πυκνών τηλεπικοινωνιών συστημάτων, εξαιρετικά υψηλή κινητικότητα).

Διάφοροι αλγόριθμοι θα βοηθήσουν επίσης να αποφασιστεί ο τρόπος με τον οποίο κατανέμονται και διαχειρίζονται οι πόροι σε ετερογενείς τεχνολογίες πρόσβασης με τη χρήση μη εξουσιοδοτημένων, αδειοδοτημένων συχνοτήτων τηλεποινωνιών ή διαμόρφωση πλαισίων για νέες διαδικασίες ανίχνευσης, βελτιώνοντας έτσι την απόδοση του συστήματος χρησιμοποιώντας τους ίδιους διαθέσιμους πόρους.

Μια πρώτη έκδοση της προτεινόμενης λύσης παρέχεται στο κείμενο που ακολουθεί. Συγκεκριμένα, έχουν διερευνηθεί δύο αλγόριθμοι. Ο πρώτος είναι ένας αλγόριθμος τυχαίας εκχώρησης καναλιού. Αυτός ο αλγόριθμος δεν έχει κάποια λογική για την εκχώρηση καναλιών και γι 'αυτό ονομάζεται «τυχαίος». Ο αλγόριθμος τυχαίας εκχώρησης καναλιών χρησιμοποιείται ως βάση προκειμένου να αξιολογηθεί η αποτελεσματικότητα του επόμενου αλγορίθμου που προτείνουμε και ονομάζεται "Αλγόριθμος εκχώρησης καναλιών με βάση SINR". Η διαδικασία επιλογής διαφέρει από τον αλγόριθμο τυχαίας εκχώρησης καναλιού αφού εδώ εισάγουμε ένα σημείο ελέγχου για τον έλεγχο των καλύτερων διαθέσιμων καναλιών για να επιλεγούν (αν είναι διαθέσιμα). Το καλύτερο κανάλι προσδιορίζεται σύμφωνα με το SINR και αν το SINR ενός νέου καναλιού είναι καλύτερο από το τρέχον χρησιμοποιούμενο, τότε το UE θα μεταβεί στο καλύτερο κανάλι. Γενικά, αναμένεται ότι μέσω αυτού του αλγορίθμου θα είναι δυνατή η επίτευξη καλύτερης ποιότητας (π.χ. υψηλότερη απόδοση, μικρότερη καθυστέρηση, μικρότερη διάρκεια περιόδου λειτουργίας).

Επίσης, πρέπει να αναφερθεί ότι ο αλγόριθμος προσπαθεί να αντιμετωπίσει το πρόβλημα της λιμοκτονίας, δημιουργώντας έναν κατάλογο προτεραιοτήτων των εξυπηρετούμενων UEs (ως αποτέλεσμα, κάποια προτεραιότητα θα δοθεί σε UEs που εξυπηρετούνται λιγότερες φορές σε σύγκριση με άλλα).

Η τοπολογία του δικτύου που δημιουργείται από το εργαλείο προσομοίωσης απεικονίζεται στην Εικόνα 6. Συγκεκριμένα, οι χρήστες εμφανίζονται ως μικροί. Επιπλέον, τα πράσινα βέλη απεικονίζουν τη διαδικασία μετάδοσης και την τοπολογία σύνδεσης μεταξύ UEs και BSs κάθε χρήστη σε ένα συγκεκριμένο κελί του δικτύου μας.



Εικόνα 6: Τοπολογία δικτύου που δημιουργήθηκε από το εργαλείο προσομοίωσης.

Τα αποτελέσματα από τον πειραματισμό μας δείχνουν ότι ο αλγόριθμος SINR ήταν σε θέση να επιλέξει δυναμικά το καλύτερο κανάλι με βάση την παρεμβολή της τρέχουσας θέσης και έτσι να επιτρέψει σε κάθε χρήστη να συνδεθεί με υψηλότερη ταχύτητα και να λάβει το αρχείο γρηγορότερα με χαμηλότερη απώλεια πακέτων. Αντίθετα, ο τυχαίος αλγόριθμος κατά μέσον όρο έκανε την χειρότερη επιλογή των καναλιών, επομένως οι χρήστες δεν ήταν σε θέση να μεταφορτώσουν με πλήρη ταχύτητα και με υψηλότερο πακέτο πακέτων απώλειας, δημιουργώντας έναν συνεχή βρόχο κακής επιλογής καναλιών χωρίς να μπορούν να ξεπεράσουν αυτή τη κατάσταση.

4 Διαχείριση κοινών πόρων ραδιοσυχνοτήτων σε περιβάλλον πολλών παρόχων (Resource Sharing)

Οι λειτουργίες του RRM ελέγχουν πολλά και διαφορετικά στρώματα, αλλά και στοιχεία του συστήματος δικτύων. Για παράδειγμα υποστηρίζουν τον μηχανισμό "αποδοχής / ιεράρχησης / καθοδήγησης", ο οποίος λαμβάνει αποφάσεις σχετικά με το εάν θα αποδεχθεί το σύστημα μία νέα ροή κυκλοφορίας δεδομένων, ποιο επίπεδο προτεραιότητας θα έπρεπε να έχει, και σε ποιο σταθμό βάσης (base station) ή σταθμούς βάσης να την κατευθύνουν. Τα αποτελέσματα που εξάγονται από τον μηχανισμό στέλνονται στο MAC επίπεδο ώστε να τα κατευθύνει προς το κατάλληλο RAT (Radio access technology) δηλαδή στην κατάλληλη τεχνολογία φυσικού επιπέδου (LTE, Bluetooth, WiFi, κτλπ) ώστε να διατηρηθεί μια αντιστοίχιση μεταξύ τύπων της κυκλοφορίας και των διαθέσιμων ζωνών. Για μία τέτοια διαδικασία θα μπορούσαν να θεσπιστούν κανόνες σύνδεσης, ανάλογα με τις απαιτήσεις κάθε συγκεκριμένου τύπου δεδομένων που αποστέλνονται και σε ποιες ζώνες φάσματος ή RAT να είναι διαθέσιμα.

Μπορούν να συμπεριληφθούν πρόσθετα κριτήρια για τον καλύτερο χειρισμό κίνησης εκτός από την ποιότητα του σήματος που λαμβάνει ο χρήστης (QoS) ή και σφάλματα από δεδομένα που έχουν χαθεί, όπως ο αναμενόμενος ρυθμός αποστολής δεδομένων και η αναμενόμενη κάλυψη που παρέχεται από ένα συγκεκριμένο RAT ή μπάντα ραδιοσυχνοτήτων. Η εξισορρόπηση φορτίου (Load Balance) στοχεύει στην αποτελεσματικότητά της χρήσης του περιορισμένου φάσματος για την αντιμετώπιση των ανισοτήτων, προκειμένου να βελτιωθεί η αξιοπιστία του δικτύου από τη μείωση της πιθανότητας συμφόρησης σε περιοχές υψηλού ενδιαφέροντος (hotspots) εντός κυψελοειδών δικτύων.

Εκτός από την επιλογή για το κατάλληλο φάσμα, μπορεί να επιλεγεί επίσης ο αριθμός των καναλιών που θα χρησιμοποιηθούν σε μια μπάντα εάν χρειαστεί. Ακόμα μπορεί να προτείνει τη λειτουργική διαμόρφωση για κάθε ζώνη ή ακόμα και τη διαμόρφωση του MAC επιπέδου. Άλλες διαδικασίες είναι για παράδειγμα η "συνεργασία μεταξύ RAT" ώστε να μπορέσουν να συνυπάρξουν πολλές τεχνολογίες ασύρματης δικτύωσης στην ίδια μπάντα. Για παράδειγμα, στη συχνότητα των 5 GHz, η οποία είναι μια συχνότητα που λειτουργεί χωρίς άδεια όπου η μεταφορά δεδομένων πρέπει να συνυπάρξει με το WiFi, με σωστό τρόπο και μόνο όταν αυτό είναι δυνατόν να συμβεί. Η διαδικασία επιλογής των καλύτερων διαθέσιμων πόρων μπορεί να γίνει είτε στο RRM είτε στο MAC επίπεδο, ανάλογα με το χρονοδιάγραμμα και τις διαθέσιμες πληροφορίες που υπάρχουν. Στην περίπτωση των αυτόνομων μικρών κυψελών, η επιλογή καναλιού μπορεί να εκτελείται αυτόνομα στο επίπεδο MAC, και τα αποτελέσματα μπορούν ενδεχομένως να επαληθευτούν από το RRM.

Ένα κανάλι μπορεί να έχει επίσης διάφορες πολιτικές και περιορισμούς στη χρήση του ραδιοφάσματος για διάφορες περιοχές ή χρονικές περιόδους. Το σύστημα ακόμα μπορεί να έχει συγκεκριμένους στόχους ποιότητας ή επιδόσεων και επίσης να παρακολουθεί την τρέχουσα απόδοση.

Μια ιεραρχική διαχείριση των τεχνολογιών ασύρματης δικτύωσης που μοιράζεται διανεμημένες και συγκεντρωτικές λύσεις για εξαιρετικά πυκνό πολυζωνικά δίκτυα θα μπορούσε να είναι επωφελείς προκειμένου να ξεπεραστούν οι προκλήσεις που δημιουργούνται σε αυτά τα περιβάλλοντα. Η κεντρική διαχείριση, που χρησιμοποιείται ως βάση, μπορεί να επεκταθεί και να διανεμηθεί με τη λήψη αποφάσεων διαχείρισης που σχετίζονται με την επιλογή τεχνολογιών ασύρματης δικτύωσης / φάσματος / καναλιού πιο κοντά στο επίπεδο κόμβου. Έτσι, μια κεντρική διαχείριση μπορεί να ενεργοποιήσει την κατανεμημένη διαχείριση και αντίστροφα. Οι αλγόριθμοι που εκτελούνται κατά κατανεμημένο τρόπο θα είναι σε θέση να παρέχουν καλύτερες επιδόσεις, να υπολογίζουν και να αναλύουν σημαντικά λιγότερα δεδομένα από τα συγκεντρωμένα σε μία κεντρική οντότητα. Ωστόσο, σε περίπτωση που η κατανεμημένη προσέγγιση δεν παρέχει μια ικανοποιητική λύση, μια κεντρική λύση μπορεί να ενεργοποιηθεί. Ο κύριος σκοπός της κατανεμημένης διαχείρισης είναι η παροχή μιας λύσης σε σχεδόν πραγματικό χρόνο σε αντίθεση με τους περιορισμούς της κεντρικό-ποιημένης διαχείρισης.

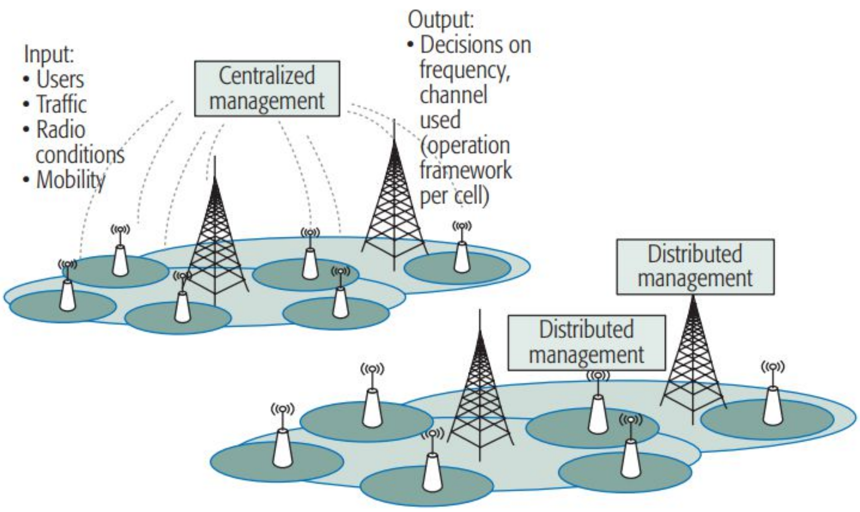
Ο προτεινόμενος αλγόριθμος βασίζεται σε τεχνικές μηχανικής μάθησης (machine learning) για τις τεχνολογίες ασύρματης δικτύωσης / φάσμα / κανάλι , κλπ που πρέπει να επιλεγούν στη ζώνη των 3,5 GHz για καλύτερη επίτευξή επιδόσεων, ειδικά σε πυκνά και με υψηλή συμφόρηση 5G περιβάλλοντα. Ο προτεινόμενος αλγόριθμος αναλαμβάνει να διαχειριστεί συχνότητες με διαφορετικές αδειοδοτήσεις (άδεια / χωρίς άδεια / μερική άδεια) για την ανάγκη εκπλήρωσης συγκεκριμένων απαιτήσεων μεταφοράς δεδομένων.

Η λύση βασίζεται στο μοντέλο του συστήματος πρόσβασης φάσματος (SAS) που λειτουργεί στη ζώνη των 3,5GHz. Το ιεραρχικό μοντέλο SAS [7] αποτελείται από ένα μοντέλο τριών επιπέδων: τους χρήστες με άδεια χρήσης (με άδεια χρήσης), τους χρήστες με προτεραιότητα πρόσβασης (PAL) (μερικώς αδειοδοτημένοι) και τους χρήστες γενικής άδειας πρόσβασης (χωρίς άδεια). Οι χρήστες PAL και GAA ελέγχονται από το σύστημα SAS και, συνεπώς, πρέπει να καταχωρίζουν και να ελέγχουν όλες τις λειτουργίες τους προκειμένου να εξασφαλίσουν ασφαλές και χωρίς παρεμβολές περιβάλλον σε χρήστες υψηλότερων επιπέδων.

Επιπλέον, μετά από χρήση ενός καναλιού των 3,5GHz, ο αλγόριθμος θα συνεχίσει να ελέγχει οποιαδήποτε πληροφορία για το κανάλι που δίνεται σε συγκεκριμένη βαθμίδα. Ιδιαίτερα για τους χρήστες PAL και GAA, μπορεί να υπάρχουν οδηγίες για χρήση καναλιών από χρήστες υψηλότερου επιπέδου, που σημαίνει ότι πρέπει να αλλάξουν το κανάλι που χρησιμοποιούν. Με την βοήθεια του SAS, ο αλγόριθμος είναι σε θέση να αποκτήσει γνώση σχετικά με τη χρήση των καναλιών και ειδικά εάν η μπάντα βρίσκεται υπό χρήση των PAL ή GAA. Το πιο σημαντικό είναι ότι ο μηχανισμός SAS δεν είναι χρονοπρογραμματιστής σε πραγματικό χρόνο, γι' αυτό και ο προτεινόμενος αλγόριθμος είναι απαραίτητος, ώστε να παράσχει την κατάλληλη, αποτελεσματική και ταχύτερη επιλογή των καναλιών χωρίς προστριβές μεταξύ των χρηστών για την ορθή λειτουργία του συστήματος. Ο προτεινόμενος αλγόριθμος θα τρέξει πρώτα με κατανεμημένο τρόπο, σε κάθε κυψέλη . Από εκεί θα συγκεντρωθούν πληροφορίες σχετικά με τη διαθεσιμότητα των τεχνολογιών ασύρματης δικτύωσης, φάσματος και καναλιών και θα εισαχθούν στο μηχανισμό μάθησης προκειμένου να εξαχθούν δεδομένα σχετικά με τους χρήστες, τα χρησιμοποιούμενα συστήματα αδειών χρήσης, τις γειτονικές κυψέλες, αναθέσεις συχνοτήτων κτλ. Μετά από μια επιτυχημένη επιλογή ή ακόμα και πρόβλεψη ενός συγκεκριμένου καναλιού, όλες οι πληροφορίες ανταλλάσσονται με το επίπεδο MAC.

Προκειμένου να αξιολογήσουμε καλύτερα τον αλγόριθμό, ένα πλήθος από σενάρια και περιπτώσεις έχουν δοκιμαστεί. Το καλύτερο κανάλι προσδιορίζεται σύμφωνα με το SINR (αναλογία σήματος προς παρεμβολή-συν-θόρυβο). Αν το SINR ενός νέου καναλιού είναι καλύτερο από αυτό που χρησιμοποιούσε μέχρι εκείνη τη στιγμή, τότε ο χρήστης θα μεταβεί στο νέο καλύτερο κανάλι.

Ο αλγόριθμος καθιστά εφικτή την χρήση του μηχανισμού διαχείρισης ραδιο-πόρων (RRM) σε κάθε σταθμό βάσης (Base station) όπου θα μπορεί να συλλέξει τα επίπεδα παρεμβολής για κάθε χρήστη που συνδέεται με ένα συγκεκριμένο κανάλι για να συμπεράνει την κατάσταση του περιβάλλοντος και να την αξιοποιήσει καταλλήλως. Γενικά, μέσω αυτού του αλγορίθμου είναι δυνατόν να επιτευχθεί καλύτερη ποιότητα (π.χ. υψηλότερη απόδοση και λιγότερη καθυστέρηση).



Εικόνα 7: Κεντρικοποιημένο και κατανεμημένο σύστημα.

5 Νοητός διαχωρισμός δικτύων (Network Slicing)

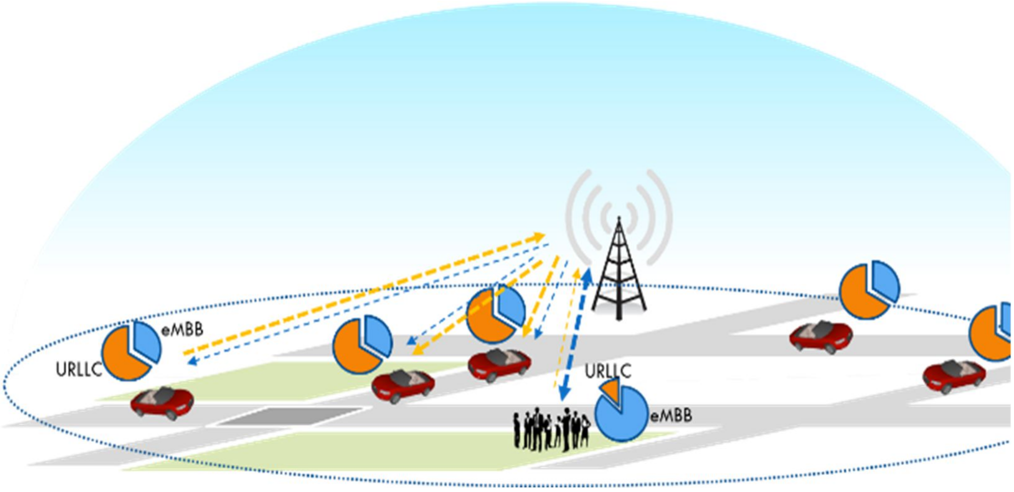
Στο πλαίσιο του κεφαλαίου αυτού θα παρουσιαστεί η μελέτη και ανάλυση απαιτήσεων για βελτιωμένη διαχείριση πόρων σε περιβάλλοντα τα οποία χαρακτηρίζονται από υψηλό βαθμό συγκέντρωσης χρηστών και δυναμικότητας, αξιοπιστίας και διαθεσιμότητας δικτύου για καλύτερη μεταφορά δεδομένων από και προς τους χρήστες. Ακολουθήθηκε καθορισμένη διαδικασία για τον ορισμό των δικτύων αυτών, τον εντοπισμό του προβλήματος, αλλά και επίλυση αυτού με την χρήση του τεμαχισμού δικτύου.

Μετά από μελέτη σε μεγάλο μέγεθος θεμάτων σχετικά με τον τεμαχισμό του δικτύου που ερευνώνται σε διάφορες δημοσιεύσεις, προτείνεται ένας αλγόριθμος για τη δημιουργία και τη λήψη αποφάσεων σχετικά με τη δυναμική κατανομή πόρων των τεμαχίων του δικτύου. Ο προτεινόμενος αλγόριθμος ανασυνθέτει και προσαρμόζει τα τεμάχια έτσι ώστε να παρέχονται τα κατάλληλα επίπεδα ποιότητας υπηρεσίας (QoS) στους κόμβους των πελατών. Επίσης, μία από τις καινοτόμες πτυχές είναι ότι ο αλγόριθμος διακρίνει τους τύπους των υπηρεσιών (URLLC, eMBB) προκειμένου να δοθεί προτεραιότητα στην κατάλληλη υπηρεσία την κρίσιμη στιγμή. Επίσης, με τον αλγόριθμο αυτό επιδιώκεται η σωστότερη και καταλληλότερη χρήση των πόρων (δηλαδή η προμήθεια πόρων από μη κρίσιμες δραστηριότητες, για την εξυπηρέτηση κρίσιμων δραστηριοτήτων εάν είναι δυνατόν).

Σε γενικές γραμμές, ένα τεμάχιο μπορεί να θεωρηθεί ως ένα λογικό δίκτυο που βασίζεται σε ένα υποσύνολο των φυσικών πόρων ενός δικτύου. Από την άποψη αυτή, μπορεί να υπάρχει τεμαχισμός του 5G RAN, καθώς και ο τεμαχισμός άλλων τμημάτων, προκειμένου να υποστηριχθεί η συνδεσιμότητα από άκρο σε άκρο. Κατά συνέπεια, ένα δίκτυο μπορεί να χωριστεί σε ένα σύνολο τεμαχίων. Κάθε τεμάχιο πρέπει να είναι επαρκές για την παροχή συγκεκριμένης υπηρεσίας.

Μια φέτα δικτύου είναι ένα εικονικό δίκτυο που δημιουργείται πάνω από ένα φυσικό δίκτυο με τέτοιο τρόπο ώστε να δίνει την ψευδαίσθηση στον χρήστη ότι το τεμάχιο λειτουργεί στο δικό του αποκλειστικό φυσικό δίκτυο. Επίσης, ένα τεμάχιο θα πρέπει να είναι ένα αυτόνομο δίκτυο με δικούς του εικονικούς πόρους, τοπολογία, ροή της κυκλοφορίας και κανόνες. Μια συσκευή μπορεί να αντιστοιχιστεί σε ένα τεμάχιο με βάση τη συνδρομή, τον τύπο της συσκευής και τις υπηρεσίες που παρέχονται από το δίκτυο ή να αφαιρεθεί από αυτό και να αντιστοιχιστεί σε ένα διαφορετικό τεμάχιο, εάν απαιτείται [8][9][10].

Για τους σκοπούς της παρούσας μελέτης εξετάστηκε ένα σενάριο που περιλαμβάνει οχήματα και χρήστες. Μέσω ενός τέτοιου σεναρίου είναι δυνατόν να δείξουμε τον αντίκτυπο των αποφάσεων τόσο στα τεμάχια με υπηρεσίες URLLC όσο και στα eMBB, λόγω του γεγονότος ότι για τις επικοινωνίες των οχημάτων οι υπηρεσίες URLLC είναι εξαιρετικά σημαντικές, ενώ για επιβάτες / πεζούς χρησιμοποιείται η χρήση των eMBB υπηρεσιών για τη μετάδοση του περιεχομένου, κλπ. Η κατανομή των πόρων αρχικά έχει προτεραιότητα για το τεμάχιο με υπηρεσία URLLC λόγω της κρίσιμης φύσης της και οι υπόλοιποι πόροι διατίθενται σε eMBB που μπορούν να προσφέρουν βίντεο υψηλής ευκρίνειας, περιήγηση κλπ.



Εικόνα 8 : Επισκόπηση σεναρίου.

Ο αλγόριθμος, ο οποίος έχει δημιουργηθεί, είναι ικανός να επιλύσει το πρόβλημα της κατανομής των πόρων σε τεμάχια. Γενικά, υποθέτοντας μια ομάδα πόρων που είναι διαθέσιμες για τα διαφορετικά τεμάχια δικτύου, γίνεται αντιστοίχιση αυτών των ομάδων πόρων στα αντίστοιχα τεμάχια (π.χ., εάν υπάρχει αίτηση για URLLC ή eMBB υπηρεσίας). Αν η ανάθεση των πόρων δεν ήταν επιτυχής ο αλγόριθμος πρέπει να τρέξει ξανά για να βρει μια άλλη πιθανή λύση (για παράδειγμα, όπως αναφέρθηκε προηγουμένως, ανακατανομή των υφιστάμενων πόρων αν είναι δυνατόν).

Σε αυτή τη μελέτη εστιάζουμε ιδιαίτερα στις υπηρεσίες URLLC και eMBB. Όλα τα στοιχεία έχουν υλοποιηθεί σε προσομοιωτή επιπέδου συστήματος με τα ακριβή στοιχεία του περιβάλλοντος, της κινητικότητας των χρηστών αλλά και των πακέτων που αποστέλλονται προκειμένου να αναπαραχθεί καλύτερα το περιβάλλον που θα χρησιμοποιηθούν στο εγγύς μέλλον οι τεχνολογίες αλλά και οι υπηρεσίες αυτές.

Η ιδέα για τον τεμαχισμό σε δίκτυο έχει προσδιοριστεί για σενάρια όπου απαιτείται ταυτόχρονη και διαφορετική χρήση από πελάτες με συγκεκριμένα SLAs. Σε αυτή την έρευνα, συνυπολογίζεται η χρήση των υπηρεσιών URLLC και eMBB, προκειμένου να αξιολογηθεί η επίπτωση της αύξησης της χρήσης μια εξ αυτών (είτε eMBB είτε URLLC) είτε ακόμα και των δύο στη συνολική απόδοση του δικτύου. Για τον λόγο αυτό, προτείνεται ένας αλγόριθμος για τη δημιουργία και τη λήψη αποφάσεων σχετικά με τη δυναμική κατανομή πόρων σε τεμάχια δικτύου, η οποία αναδιαμορφώνει και προσαρμόζει τα τεμάχια έτσι ώστε να παρέχονται τα κατάλληλα επίπεδα ποιότητας των υπηρεσιών (QoS) προς τους κόμβους των κινητών πελατών. Συνολικά, όσον αφορά τη χρήση πόρων μεταξύ του URLLC και του eMBB, φαίνεται ότι ο αντίκτυπος δεν είναι πολύ υψηλός, επομένως η κρίσιμη υπηρεσία URLLC μπορεί να εξυπηρετηθεί επαρκώς από ένα μικρό ποσοστό πόρων που έχουν αφιερωθεί στα συγκεκριμένα τεμάχια, ενώ η υπόλοιπη κίνηση αντιμετωπίζεται δυναμικά από τα τεμάχια eMBB. Από τα ευρήματα παρατηρείται ότι η καθυστέρηση της υπηρεσίας URLLC μπορεί να φτάσει κοντά στα 2ms (για μικρά πακέτα 32bytes) και γενικά δεν υπερβαίνει τα 7ms (για μεγαλύτερα πακέτα των 200 bytes) και η υπηρεσία δεν επηρεάζεται κατά πολύ, λόγω του έξυπνου αλγορίθμου που χειρίζεται τους πόρους των τεμαχίων του δικτύου.

6 Διαμοιρασμός και κατανομή πόρων νέων υπηρεσιών

Το 5G χαρακτηρίζεται από τις προκλήσεις της ταχείας ανάπτυξης των συνδέσεων κινητής τηλεφωνίας και του όγκου κυκλοφορίας [11], [12]. Για την αντιμετώπιση αυτών των προκλήσεων, πρέπει να υπάρξει αποτελεσματική εκμετάλλευση των ασύρματων τεχνολογιών, ώστε να παρέχεται μεγαλύτερη χωρητικότητα ακολουθώντας τρεις κύριες διαστάσεις: i) υπερσυμπύκνωση μέσω μικρών κυψελών, ii) εξισορρόπηση φορτίου σε όλο το διαθέσιμο φάσμα και iii) εκμετάλλευση πόρων σε διαφορετικές τεχνολογίες. Λαμβάνοντας υπόψη το συγκεκριμένο τρισδιάστατο μοντέλο, το οποίο μπορεί να ονομαστεί ως εκτεταμένη δυναμική κατανομή φάσματος (eDSA), μπορούν να διαχειριστούν από κοινού διαφορετικές ζώνες φάσματος και τεχνολογίες, ώστε να βελτιωθεί η ποιότητα εμπειρίας των χρηστών (QoE).

Ένα από τα κύρια σενάρια που πρέπει να αντιμετωπιστεί σωστά είναι η περίπτωση ετερογενών δικτύων όπου εφαρμόζεται μια τεράστια ανάπτυξη μικρών κυψελών για την παροχή μιας ομοιόμορφης ευρυζωνικής εμπειρίας στους χρήστες, λαμβάνοντας υπόψη εφαρμογές με διαφορετικές απαιτήσεις QoS, όπως ροή πολυμέσων υψηλής ανάλυσης , παιχνίδια, κλήσεις βίντεο και υπηρεσίες cloud. Μια σημαντική πρόκληση σε αυτά τα δίκτυα είναι η αποτελεσματική διαχείριση παρεμβολών συν-καναλιών (CCI) που συμβαίνει λόγω της εγγύτητας μεταξύ των SBS. Ως εκ τούτου, δεδομένου ότι τα ίδια κανάλια επαναχρησιμοποιούνται μεταξύ των SBS λόγω των σπάνιων φασματικών πόρων, το CCI αποτελεί σημαντικό περιοριστικό παράγοντα για την απόδοση του δικτύου.

Για την αντιμετώπιση αυτής της πρόκλησης, έχουν προταθεί τεχνικές δυναμικής ανάθεσης καναλιών (DCA), είτε λαμβάνοντας υπόψη μια κεντρική προσέγγιση [13] είτε μια διανεμημένη [12]. Πρέπει να σημειωθεί ότι οι συγκεντρωτικές προσεγγίσεις έχουν πολλά πλεονεκτήματα από την άποψη της απόδοσης. Παρ 'όλα αυτά, η υψηλή υπολογιστική πολυπλοκότητα τα καθιστά ακατάλληλα για την περίπτωση ενός ετερογενούς δικτύου με τεράστιο αριθμό μικρών κυψελών. Επομένως, οι κατανεμημένες τεχνικές DCA έχουν κερδίσει το ενδιαφέρον πολλών ερευνητών ως λύση που μπορεί να εφαρμοστεί σε μελλοντικά ασύρματα δίκτυα. Ωστόσο, η πλειονότητα των σχεδίων DCA στη βιβλιογραφία θεωρούν ότι τα SBS δεν κάνουν διάκριση μεταξύ αιτημάτων κυκλοφορίας από εφαρμογές εξοπλισμού χρήστη (UE), ακόμη και αν οι εφαρμογές δεν έχουν την ίδια προτεραιότητα από την άποψη του χρήστη. Λαμβάνοντας υπόψη ότι στα δίκτυα 5G, η κίνηση θα κυμαίνεται από υψηλούς ρυθμούς δεδομένων μέχρι μικρά πακέτα για επικοινωνία μηχανών, καλύπτοντας μια ποικιλία διαφορετικών εφαρμογών, υπάρχει μια αναδυόμενη ανάγκη για συστήματα DCA που προσφέρουν διαφοροποιημένο QoS σε κάθε χρήστη, αντιμετωπίζοντας τις μεταβαλλόμενες συνθήκες δικτύου και το CCI που διαφέρει από το χρόνο. Με βάση αυτήν την παρατήρηση μελετήθηκε ένας τροποποιημένος μηχανισμός διαχωρισμού διανεμημένων καναλιών που λαμβάνει υπόψη τα χαρακτηριστικά CCI και QoS των χρηστών. Το προτεινόμενο DCA (IQ-CS-DCA) που βασίζεται στον διαχωρισμό καναλιών Interference και QoS μπορεί να χρησιμοποιηθεί για να χρησιμοποιεί αποτελεσματικά τους φασματικούς πόρους και ταυτόχρονα να δίνει προτεραιότητα στους χρήστες με εφαρμογές περιορισμένης καθυστέρησης (όπως ροή βίντεο).

Τα επιλεγμένα σενάρια είναι τα εξής:

- Massive IoT (Internet of Things): αυτό το σενάριο αναφέρεται στο "low-end IoT" και καλύπτει συσκευές με σποραδική και ανεκτική καθυστέρηση κυκλοφορίας, που αποτελείται κυρίως από σύντομα πακέτα. Μεταξύ άλλων, αυτή η κατηγορία περιλαμβάνει συνήθως φορητές συσκευές, έξυπνους μετρητές, συσκευές οικιακού αυτοματισμού, υγειονομική περίθαλψη, μη κρίσιμους αισθητήρες έξυπνων πόλεων και ασύρματα δίκτυα αισθητήρων για παρακολούθηση του περιβάλλοντος.

- Εξαιρετικά αξιόπιστες επικοινωνίες: αυτό το σενάριο αναφέρεται σε ένα δίκτυο που υποστηρίζει υπηρεσίες με ακραίες απαιτήσεις για διαθεσιμότητα και αξιοπιστία. Συγκεκριμένα, προβλέπεται να υπάρχουν νέες εφαρμογές βασισμένες σε επικοινωνία M2M (από μηχανή σε μηχάνημα) και IoT με περιορισμούς σε πραγματικό χρόνο, επιτρέποντας νέες λειτουργίες για την ασφάλεια της κυκλοφορίας, την αποτελεσματικότητα της κυκλοφορίας ή τον έλεγχο κρίσιμης αποστολής για βιομηχανικές και στρατιωτικές εφαρμογές.
- Κινητικότητα υψηλής ταχύτητας: αυτή η περίπτωση χρήσης λαμβάνει υπόψη περιβάλλοντα υψηλής κινητικότητας (π.χ. τρένα υψηλής ταχύτητας, αυτοκίνητα σε αυτοκινητόδρομους κ.λπ.) όπου πρέπει να επιτευχθούν ευρυζωνικές επικοινωνίες.
- Ευρυζωνική ασύρματη σύνδεση: αυτή η περίπτωση χρήσης αποτελεί το σενάριο ενδιαφέροντος και εστιάζει σε ένα μείγμα οικιακού, επιχειρηματικού και δημόσιου περιβάλλοντος εξωτερικού και εσωτερικού χώρου που βρίσκεται σε μια πυκνοκατοικημένη αστική περιοχή (βλ. Σχήμα 1). Σε αυτήν την περίπτωση, ένας μεγάλος αριθμός μικρών κυττάρων συνυπάρχουν εντός ενός μακροκυττάρου προσφέροντας βελτιωμένη εμπειρία επικοινωνίας στους χρήστες.

Προκειμένου να ικανοποιηθούν οι απαιτήσεις 5G, που χαρακτηρίζουν τη συγκεκριμένη περίπτωση χρήσης, προτείνουμε έναν μηχανισμό DCA για την αποτελεσματική χρήση του διαθέσιμου φάσματος, με γνώμονα τον συντονισμό του CCI και των απαιτήσεων QoS των χρηστών. Η έννοια των ετερογενών δικτύων εστιάζει στη βελτίωση της φασματικής απόδοσης ανά μονάδα περιοχής χρησιμοποιώντας ένα διαφορετικό σύνολο σταθμών βάσης (BS), σε ένα συνδυασμό μακροκυττάρων και μικρών κυψελών.

Σε αυτήν την ενότητα, περιγράφουμε τον προτεινόμενο μηχανισμό IQ-CS-DCA που βασίζεται στον μηχανισμό IACS-DCA. Αρχικά, παρουσιάζουμε μια αφηρημένη διατύπωση του θεωρούμενου προβλήματος βελτιστοποίησης, ενώ στη συνέχεια δίνεται μια πιο αλγοριθμική προσέγγιση του προτεινόμενου μηχανισμού.

Ο προτεινόμενος μηχανισμός μπορεί να χωριστεί σε πέντε κύρια στάδια και κάθε φάση περιγράφεται εν συντομία.

- Φάση αρχικοποίησης: Κατά τη διάρκεια αυτής της φάσης, κάθε SBS επιλέγει τυχαία ένα κανάλι από το σύνολο των διαθέσιμων καναλιών και μεταδίδει ένα σήμα beacon σε αυτό το κανάλι.
- Φάση μέτρησης: Κάθε SBS μετρά περιοδικά την στιγμιαία ισχύ σήματος φάρου σε καθένα από τα διαθέσιμα κανάλια για μια συγκεκριμένη χρονική διάρκεια. Η ληφθείσα ισχύς μπορεί να υπολογιστεί λαμβάνοντας υπόψη τόσο την απώλεια διαδρομής όσο και τα φαινόμενα εξασθένισης για μια πιο ολοκληρωμένη ανάλυση του περιβάλλοντος ραδιοφωνικής διάδοσης.
- Δημιουργία πίνακα προτεραιότητας καναλιού: Κάθε SBS δημιουργεί τον πίνακα προτεραιότητας καναλιού βάσει των μέσων επιπέδων ισχύος CCI. Σε αυτό το βήμα, η μέση ισχύς CCI μπορεί να υπολογιστεί είτε χρησιμοποιώντας το φιλτράρισμα πρώτης τάξης παρόμοιο με το [14] είτε με τη χρήση άλλων μηχανισμών εκμάθησης / μέσου όρου που χρησιμοποιούν προηγούμενες μετρήσεις CCI και οδηγούν σε μια σταθερή εκχώρηση. Το κανάλι με το χαμηλότερο CCI εμφανίζεται πρώτο στον πίνακα προτεραιότητας και ακολουθούν τα άλλα κανάλια με φθίνουσα σειρά CCI.
- Σύνδεση UEs-SBS: Κατά τη διάρκεια αυτής της φάσης, κάθε UE συσχετίζεται με ένα SBS ανάλογα με διάφορες μετρήσεις (π.χ. την υψηλότερη ένδειξη ισχύος σήματος λήψης (RSSI), το φορτίο λόγω άλλων UE που σχετίζονται με αυτό το SBS κ.λπ.).
- Συλλογή αιτημάτων: Κάθε SBS συλλέγει τα αιτήματα καναλιού από τα UE.
- Δημιουργία πινάκων προτεραιότητας απαίτησης User QoS: Τα UE έχουν προτεραιότητα ανάλογα με την προτεραιότητα εφαρμογής τους και τα SBS διαιρούν τον πίνακα προτεραιότητας σε πολλούς πίνακες (ανάλογα με τον αριθμό των UE / εφαρμογών). Το πρώτο κανάλι κάθε πίνακα αντιστοιχεί σε

κάθε UE ανάλογα με την προτεραιότητα εφαρμογής του και την ποιότητα του καναλιού που δίνεται από το επίπεδο ισχύος CCI (καλύτερα κανάλια δίνονται σε UE με αυστηρότερες απαιτήσεις QoS).

- Εκχώρηση καναλιού: Κάθε SBS εκχωρεί τα κανάλια στους χρήστες με βάση τους πίνακες προτεραιότητας QoS.

Για την αξιολόγηση της προτεινόμενης λύσης, εφαρμόστηκαν δύο αλγόριθμοι ως πρώτο στάδιο. Για παράδειγμα, οι χρήστες IoT θα έχουν χαμηλή προτεραιότητα στον αλγόριθμό μας, οι εξαιρετικά αξιόπιστες επικοινωνίες θα έχουν την υψηλότερη προτεραιότητα, πράγμα που σημαίνει ότι θα εκχωρηθούν σε κανάλια με καλύτερη τιμή SINR. Τέλος, οι ευρυζωνικές επικοινωνίες όπου οι χρήστες θα έχουν είτε μια μεσαία προτεραιότητα σε σχέση με την ποιότητα του καναλιού που έχει εκχωρηθεί ή ακόμη και μια υψηλή προτεραιότητα, δεδομένου ότι θα μπορούσαν να ανήκουν σε μια κατηγορία χρηστών που πρέπει να έχουν χαμηλά λανθάνοντα χρόνο και υψηλή απόδοση.

Ο πρώτος αλγόριθμος που ερευνήθηκε ονομάζεται «Αλγόριθμος καναλιού τυχαίας βάσης». Αυτή η λύση δεν έχει συγκεκριμένη λογική για την εκχώρηση καναλιών και γι 'αυτό ονομάζεται "τυχαία", εκχωρεί αυθαίρετα τους χρήστες σε διαφορετικά κανάλια χωρίς καμία γνώση της τρέχουσας κατάστασης των καναλιών ή ολόκληρου του συστήματος. Χρησιμοποιείται ως βασική γραμμή για τη σύγκριση, αξιολόγηση και βελτιστοποίηση της αποτελεσματικότητας του επόμενου αλγορίθμου που προτείνουμε. Η διαδικασία επιλογής διαφέρει από τον αλγόριθμο εκχώρησης τυχαίων καναλιών. Σε αυτό το σημείο εισαγάγεται ένα κέντρο ελέγχου για την εύρεση και επιλογή των καλύτερων διαθέσιμων καναλιών (εάν υπάρχουν). Το καλύτερο κανάλι αναγνωρίζεται σύμφωνα με το SINR και εάν το SINR ενός νέου καναλιού είναι καλύτερο από το τρέχον χρησιμοποιούμενο, τότε το UE θα αλλάξει στο καλύτερο κανάλι. Αυτός ο αλγόριθμος επιτρέπει RRM με γνώμονα το περιβάλλον, καθώς κάθε σταθμός βάσης μπορεί να συλλέξει τα επίπεδα παρεμβολών για κάθε χρήστη που είναι συνδεδεμένος σε ένα συγκεκριμένο κανάλι, προκειμένου να εξαχθεί η κατάσταση του περιβάλλοντος του ραδιοφώνου και να το εκμεταλλευτεί κατάλληλα. Γενικά, αναμένεται ότι μέσω αυτού του αλγορίθμου, θα είναι δυνατή η επίτευξη καλύτερης ποιότητας (π.χ. υψηλότερη απόδοση, μικρότερη καθυστέρηση).

Για την αξιολόγηση τέτοιων εννοιών, εκτελούνται εκτεταμένες προσομοιώσεις σε επίπεδο συστήματος. Ο προσομοιωτής λαμβάνει υπόψη διάφορες παραμέτρους όπως επίπεδο κυκλοφορίας, διαθέσιμα στοιχεία υποδομής, διαθέσιμα κανάλια και αξιολογεί τις διάφορες περιπτώσεις δοκιμών.

Τα αποτελέσματα από τον πειραματισμό μας δείχνουν ότι κατά μέσο όρο, ο προτεινόμενος αλγόριθμος αποδίδει καλύτερα από τους άλλους δύο αλγόριθμους (έως και 50%), ιδίως σε υπηρεσίες υψηλής και μεσαίας προτεραιότητας, δίνοντάς τους μια ώθηση στην απόδοση. Αντίθετα, οι υπηρεσίες χαμηλής προτεραιότητας φαίνεται ότι δεν ωφελούνται όσο οι άλλες δύο. Είναι επίσης προφανές ότι ο αλγόριθμός μας αποδίδει καλύτερα σε σχεδόν κάθε δοκιμαστική περίπτωση και ειδικά σε περιπτώσεις με υψηλότερα φορτία (σε σύγκριση με προσομοιώσεις με λιγότερο φορτίο).

7 Διαχείριση πόρων ραδιοσυχνοτήτων (RRM) σε περιβάλλον πολλαπλών συνδέσεων

Σε συνέχεια της προηγούμενης έρευνάς εξετάστε τις απαιτητικές περιπτώσεις χρήσης με υψηλό φορτίο και πολύ περιορισμένη καθυστέρηση για να καλύψετε υπηρεσίες όπως η βελτιωμένη κινητή ευρυζωνικότητα (eMBB), οι μαζικές επικοινωνίες τύπου μηχανής (mMTC) και οι εξαιρετικά αξιόπιστες επικοινωνίες χαμηλής καθυστέρησης (URLLC). Μια σωστή αντιστοίχιση των φορέων συνιστωσών (CCs) σε αυτή την κατάσταση επιτρέπει την αύξηση της διακίνησης δεδομένων προς τους χρήστες έως και κατά 60% σε σύγκριση με ένα απλό αλγόριθμο λήψης ισχύος για τη διαχείριση συνδέσεων. Μια ακόμα λειτουργία που θα χρησιμοποιηθεί είναι η πολλαπλή συνδεσιμότητα [15], η οποία στοχεύει να είναι ένας από τους παράγοντες που επιτρέπουν την ευρεία χρήση των επερχόμενων δικτύων 5G. Με τη δυνατότητα πολλαπλών συνδέσεων, ο χρήστης μπορεί να συγκεντρώνει τους ραδιοφωνικούς πόρους από περισσότερους από δύο κόμβους δικτύου, επιτρέποντας έτσι την αισθητή αύξηση της ταχύτητας λήψης δεδομένων αλλά και της αξιοπιστίας.

Για την αξιολόγηση μέσω προσομοιώσεων πρέπει να εξετάσουμε ορισμένες πτυχές που σχετίζονται με τη διαμόρφωση, τα μοντέλα περιβάλλοντος, τα δίκτυα (προσομοιωμένα συστήματα), τα αναλυτικά στοιχεία, τη διαχείριση συμβάντων.

Τα αποτελέσματα από τον προσομοιωτή επιπέδου συστήματος 5G αξιολογούνται σε σχέση με ορισμένους στόχους KPI (π.χ. όσον αφορά την απόδοση ή την καθυστέρηση).

Μία περιοχή μπορεί να χαρακτηρίζεται από τον τύπο της, για παράδειγμα αστική, πυκνή ή υπερβολικά πυκνή (π.χ. καταγραφή απαιτήσεων μεγάλος συγκέντρωσης), προαστιακές ή ακόμα και αγροτικές. Διαφορετικές πυκνότητες χρηστών και κυκλοφορίας εξετάζονται ανάλογα με τον τύπο περιοχής. Τα μοντέλα αποστολής πακέτων που χρησιμοποιούνται είναι, mMTC(TR36.888 [16] και TR37.868 [17]), eMBB βάση τα μοντέλα FTP 1, 2, 3 (3GPP TR 36.814 [18]) και URLLC σύμφωνα με τα [17][19]. Επιπλέον υποστηρίζονται τα ακόλουθα μοντέλα κινητικότητας:

- Random Walk: Σύμφωνα με αυτό το μοντέλο κάθε πελάτης/χρήστης αλλάζει την ταχύτητα και την κατεύθυνσή του σε κάθε χρονικό διάστημα. Η προεπιλεγμένη τιμή του χρονικού διαστήματος είναι 1s, ενώ αυτές οι τιμές μπορούν να ρυθμιστούν εκ νέου. Για κάθε χρονικό διάστημα, η κατεύθυνση επιλέγεται από (0, 2π], ενώ η ταχύτητα ακολουθεί μια ομοιόμορφη ή Gaussian κατανομή από το [0, vmax].
- Γραμμική κίνηση: κάθε πελάτης/χρήστης επιλέγει τυχαία μια κατεύθυνση και κινείται κατά μήκος αυτού, με μια σταθερή ταχύτητα. Συνεχίζει να κινείται προς αυτή την κατεύθυνση ακόμα και αν φτάσει στα όρια των κυψελών.
- Τυχαία Κατεύθυνση: Κάθε πελάτης/χρήστης επιλέγει τυχαία μια κατεύθυνση και κινείται κατά μήκος αυτής, με μια σταθερή ταχύτητα, μέχρι να φτάσει στο όριο του κελιού. Στη συνέχεια, επιλέγει άλλη κατεύθυνση για να ταξιδέψει και κινείται με την ίδια ταχύτητα μέχρι να φτάσει ξανά στο όριο της κυψέλης.

Ένας σταθμός βάσης μπορεί να χαρακτηριστεί από διάφορες ιδιότητες. Ειδικότερα, οι τομείς έχουν πομποδέκτες (TRXs) οι οποίοι μπορούν να χαρακτηριστούν από πτυχές φάσματος. PHY / MAC abstractions και RRM μηχανισμούς. Για παράδειγμα, οι τυπικοί αλγόριθμοι RRM όπως ο round-robin, αλλά και οι αλγόριθμοι έρευνας όπως αναφέρεται στο [20] έχουν εφαρμοστεί στον προσομοιωτή επιπέδου συστήματος. Όσον αφορά τον τεμαχισμό, οι πολιτικές διαμοιρασμού και κοινής χρήσης εξετάζονται προκειμένου να διατεθούν οι κατάλληλοι πόροι στα τεμάχια σύμφωνα με τις ζητούμενες υπηρεσίες. Επιπλέον, εφαρμόζονται διάφοροι αλγόριθμοι RRM για την αξιολόγηση διαφορετικών στρατηγικών διαχείρισης πόρων, ανάλογα με τις υπό εξέταση περιπτώσεις χρήσης.

Μαζική πολλαπλή έξοδος και πολλαπλές είσοδοι (MIMO): Υποστήριξη καναλιών ελέγχου και δεδομένων για μαζικές λειτουργίες MIMO που βασίζονται σε δομή με κεντρική δέσμη, που βελτιώνουν τη φασματική απόδοση και επιτυγχάνουν υψηλότερους ρυθμούς δεδομένων, ενισχύοντας την απόδοση των καταναλωτών.

Ο τελικός χρήστης είναι συνδεδεμένος σε περισσότερους από έναν κόμβους δικτύου προκειμένου να ενεργοποιηθεί η μετάδοση πολλαπλών συνδέσμων. Αυτό μπορεί να χρησιμοποιηθεί σε συνδυασμό ή όχι με υποστήριξη διαχωρισμού πακέτων. Η αξιοπιστία μπορεί επίσης να βελτιωθεί εάν υπάρχουν πολλές συνδέσεις. Η συνάθροιση του φορέα χρησιμοποιείται για να αυξήσει το εύρος ζώνης και, συνεπώς, να αυξήσει το bitrate. Κάθε συσσωρευμένος φορέας αναφέρεται ως φορέας συνιστωσών (CC). Το CC μπορεί να έχει εύρος ζώνης 1,4 έως 20 MHz και μπορούν να συγκεντρωθούν κατ 'ανώτατο όριο πέντε μεταφορείς συνιστωσών, οπότε το μέγιστο συνολικό εύρος ζώνης είναι 100 MHz για τη μακροπρόθεσμη εξέλιξη (LTE). Για τη στήριξη του δικτύου New Radio (NR), το CC μπορεί να φτάσει τα 400MHz για ζώνες μεγαλύτερες των 6GHz σε εύρος ζώνης που σχετίζονται με NR και μέχρι 16 φορείς. Ο προσομοιωτής υποστηρίζει ρυθμιζόμενα εύρη ζώνης. Η τεχνολογία NB-IoT αναπτύσσεται "εντός ζώνης" στο φάσμα που διατίθεται για το LTE, χρησιμοποιώντας μπλοκ πόρων μέσα σε ένα κανονικό φορέα LTE. Το συντεταγμένο πολλαπλό σημείο (CoMP) χρησιμοποιείται για την αποστολή και λήψη δεδομένων προς και από έναν τελικό χρήστη από διάφορα σημεία για να διασφαλιστεί ότι επιτυγχάνεται η βέλτιστη απόδοση ακόμη και στις ακμές κυψελών.

Σε αυτή την έρευνα, μια νέα λειτουργικότητα RRM για τη διαχείριση CC ενσωματώνεται και αξιολογείται στον προσομοιωτή επιπέδου συστήματος. Αυτή η λειτουργικότητα αναφέρεται ως διαχειριστής φορέων συνιστωσών (CCM) και έχει σχεδιαστεί για τη βελτίωση των μετρήσεων απόδοσης, όπως η απόδοση του UE (για υπηρεσίες eMBB) ή η αξιοπιστία σύνδεσης (για υπηρεσίες URLLC) μέσω της διαχείρισης συνδέσμων σε ένα περιβάλλον πολλαπλών συνδέσεων. Μόλις ένα UE αποδοθεί σε ένα πρωτεύον κύτταρο (PCell) και έτσι γίνεται ο κύριος κόμβος (MN), σύμφωνα με τα κριτήρια κινητικότητας που συμμορφώνονται με το 3GPP, εκτελείται η λειτουργικότητα του CCM για αυτό το UE σε αυτόν τον κύριο κόμβο. Από τα CC που δέχεται το UE με ελάχιστη ισχύ, το CCM αναγνωρίζει το υποσύνολο των CC που πρόκειται να αποδοθούν σε αυτό το UE ως πρωτογενή δευτερεύοντα κύτταρα (PSCells) και δευτερεύοντα κύτταρα (SCells). Με το πρώτο, το CCM εκτελεί τελικά μια σύνδεση κόμβου UE-to-network. Το τελευταίο (SCells) μπορεί να συνδέεται είτε με το MN είτε με ένα δεδομένο SN για να επεκτείνει το διαθέσιμο εύρος ζώνης μεταξύ του UE και εκείνου του κόμβου δικτύου. Για να προσδιοριστεί το υποσύνολο των CC που θα εκχωρηθούν σε UE, ο CCM υπολογίζει μια βαθμολογία για κάθε διαθέσιμο CC, σύμφωνα με την πολιτική κάποιου φορέα εκμετάλλευσης. Τα κορυφαία CC αποτιμώνται στη συνέχεια σε αυτό το UE, που φιλοξενεί είτε ένα SCell (αν ανήκουν στο MN ή ένα προηγουμένως εκχωρηθέν SN) ή PSCell, εάν είναι το πρώτο CC που αντιστοιχεί σε αυτό το UE που ανήκει σε έναν κόμβο διαφορετικό από τον MN. Στο [21], η μέθοδος βαθμολόγησης εφαρμόζει ένα σύστημα βασισμένο σε κανόνες.

Σε περιβάλλοντα mMTC, μία από τις σημαντικές παραμέτρους είναι η πυκνότητα σύνδεσης των συσκευών. Σύμφωνα με το έγγραφο της ITU [20], η πυκνότητα σύνδεσης είναι ο συνολικός αριθμός συσκευών που πληρούν συγκεκριμένη ποιότητα υπηρεσίας (QoS) ανά μονάδα επιφάνειας (ανά km^2). Πρέπει να επιτευχθεί πυκνότητα σύνδεσης για περιορισμένο εύρος ζώνης ασύρματου δικτύου και αριθμό σημείων σύνδεσης. Το QoS στόχο έχει να υποστηρίξει την παράδοση ενός μηνύματος συγκεκριμένου μεγέθους μέσα σε συγκεκριμένο χρόνο και με κάποια πιθανότητα επιτυχίας. Αυτή η απαίτηση ορίζεται για τους σκοπούς της αξιολόγησης στο σενάριο χρήσης mMTC. Σύμφωνα με την ITU, η ελάχιστη απαίτηση για πυκνότητα σύνδεσης είναι 1.000.000 συσκευές ανά km^2.

Η απαίτηση ικανοποιείται εάν η πυκνότητα σύνδεσης C είναι μεγαλύτερη ή ίση με 1.000.000. Αναφέρουμε ακόμα το εύρος ζώνης συχνοτήτων προσομοίωσης που χρησιμοποιείται για την εκπλήρωση της απαίτησης. Επιπλέον, αναφέρουμε την απόδοση της σύνδεσης (μετρούμενη ως N 'διαιρεμένη με εύρος ζώνης προσομοίωσης) για την επιτευχθείσα πυκνότητα σύνδεσης. Το θεωρούμενο πρότυπο δημιουργίας πακέτων για μια τέτοια αξιολόγηση είναι, μέγεθος μηνύματος 32 bytes και 1 μήνυμα / 2 ώρες / συσκευή. Η άφιξη του πακέτου ακολουθεί τη διαδικασία άφιξης Poisson. Έχουν διεξαχθεί προσομοιώσεις σε επίπεδο συστήματος για την αξιολόγηση της πυκνότητας σύνδεσης σε περιβάλλοντα mMTC. Οι παράμετροι συχνότητας λαμβάνονται υπόψη στη προσομοίωση. Ως εκ τούτου, θεωρείται εύρος ζώνης από 180KHz έως 1.08MHz. Ο ρυθμός επιτυχίας (δηλαδή η επιτυχής μετάδοση μηνυμάτων) υπολογίζεται για να ελεγχθεί το αποδεκτό επίπεδο πυκνότητας σύνδεσης για την επίτευξη του κατωφλίου του 99% επιτυχίας (1% της απώλειας).

Σε αυτή την ενότητα πραγματοποιείται η αξιολόγηση για να δείξει τα οφέλη από τη χρήση του CCM σε περιβάλλον πολλαπλών συνδέσεων. Πρώτον, περιγράφεται μια τυπική περίπτωση βελτιστοποίησης στην οποία θα χρησιμοποιηθεί το CCM, σε ανισορροπία φορτίου δικτύου. Επιπλέον γίνονται πειράματα αναφορικά με την πυκνότητα σύνδεσης και στη συνέχεια, παρουσιάζονται τα αποτελέσματα προσομοίωσης και εξάγονται συμπεράσματα.

Τα αποτελέσματα δείχνουν ότι, δεδομένης της ανισορροπίας του φορτίου που έχουν όλα τα δίκτυα, η προτεινόμενη προσέγγιση επιτρέπει τη βελτίωση της διακίνησης των χρηστών ανεξάρτητα από τον αριθμό των αποστολών CC που δίνεται με την κοινή εκτίμηση κάθε CC, RSRQ και φορτίου. Ειδικότερα, ένα μέγιστο κέρδος απόδοσης 40% σε σχέση με την βασική γραμμή επιτυγχάνεται στην περίπτωση όπου εκχωρούνται 2 CC και επιτυγχάνεται ένα κέρδος 60% όταν εξετάζονται 3 CC στην περίπτωση για 8MB αρχείων, πράγμα που δείχνει την καλή λειτουργία του προτεινόμενου αλγόριθμου για τη διαχείριση πολλαπλών συνδέσεων. Τέλος πραγματοποιήθηκαν υπολογισμοί για τον αριθμό των συσκευών mMTC που μπορούν να υποστηριχθούν με ένα συγκεκριμένο QoS. Ένα εύρος ζώνης 1.08MHz είναι δυνατό να ικανοποιήσει αποτελεσματικά περισσότερες από 1 εκατομμύριο συσκευές ανά km^2, όπως έδειξαν τα αποτελέσματα.

8 Συμπεράσματα

Η ανάπτυξη των κινητών και ασύρματων δικτύων προορίζεται για υψηλότερους ρυθμούς δεδομένων αλλά και ποιότητα της υπηρεσίας και των προτιμήσεων των χρηστών, καθώς και της απόδοσης στον εξοπλισμό χρήστη. Για την αντιμετώπιση της εξαιρετικά ταχείας ανάπτυξης των υπηρεσιών της 5ης γενιάς δικτύων, ο συνδυασμός των αλγορίθμων, τεχνολογιών και μεθόδων είναι πολλά υποσχόμενος όπως έδειξαν και τα αποτελέσματα των ερευνών που έγιναν όλα αυτά τα χρόνια. Σε αυτή την έκθεση, προτάθηκαν ολοκληρωμένα συστήματα ελέγχου, πρωτοκόλλων και μηχανισμών. Διερευνήθηκαν πολυεπίπεδες και όπως αποδείχτηκε αποτελεσματικές λύσεις στα θεμελιώδη προβλήματα των μέχρι τώρα δικτύων για την παροχή επαρκούς απόδοσης δικτύου.

Επαληθεύτηκε η αποτελεσματικότητα των προτεινόμενων μεθόδων χρησιμοποιώντας εκτεταμένες προσομοιώσεις. Τα αριθμητικά αποτελέσματα έδειξαν βελτίωση σε σύγκριση με τα υπάρχοντα συστήματα και μεθόδους. Επιπλέον, η προσέγγιση ελέγχου δικτύου 5G με βάση τα ετερογενή δίκτυα αλλά και την μηχανική μάθηση είναι μια ενδιαφέρουσα κατεύθυνση για να αντιμετωπιστούν τα ζητήματα στο σύστημα δικτύου 5G από όλες τις οπτικές γωνίες του, είτε των εταιριών δικτύων, είτε των χρηστών είτε ακόμα και των νέων εφαρμογών που δημιουργούνται για να χρησιμοποιηθούν κατά κύριο λόγο στα δίκτυα 5ης αλλά και επόμενων γενεών.

9 Βιβλιογραφία

[1] Nuechterlein, Jonathan; Weiser, Philip J. (2005). Digital Crossroads. Cambridge, Massachusetts: The MIT Press. p. 235.

[2] https://en.wikipedia.org/wiki/File:United_States_Frequency_Allocations_Chart_2016_-

_The_Radio_Spectrum.pdf
[3] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by Simulated Annealing". Science, Volume 220, Number 4598, May 1983.
[4] C. Sommer, R. German, and F. Dressler, "Bidirectionally coupled network and road traffic simulation for improved IVC analysis," IEEE Transactions on Mobile Computing, vol. 10, no. 1, 2011.
[5] A.Varga, "The OMNET++ discrete event simulation system", European simulation multiconference (ESM 2001).
[6] M.Behrisch, L.Bieker, J.Erdmann, D.Krajzewicz "SUMO – Simulation of Urban Mobility", SIMUL 2011 : The Third International Conference on Advances in System Simulation.
[7] Federal Communications Commission, "3.5 GHz Band / Citizens Broadband Radio Service", available online at: https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-47A1.pdf
[8] 3GPP TS 38.300, "NR; NR and NG-RAN Overall Description; Stage 2 (Release 15)"
[9] 3GPP TS 22.261, "Service requirements for the 5G system; Stage 1 (Release 16)"
[10] 3GPP TS 22.891, "Feasibility Study on New Services and Markets Technology Enablers; Stage 1 (Release 14)"
[11] J. G. Andrews et al. "What will 5G be?." Selected Areas in Communications, IEEE Journal on ,32.6, pp. 1065-1082, 2014.
[12] P. Demestichas et al. "5G on the horizon: key challenges for the radio-access network." Vehicular Techn. Magazine, IEEE 8.3, 47-53, 2013.
[13] J. G. Andrews et al. "What will 5G be?." Selected Areas in Communications, IEEE Journal on ,32.6, pp. 1065-1082, 2014.
[14] R. Matsukawa, T. Obara, and F. Adachi. "A dynamic channel assignment scheme for distributed antenna networks." Vehicular Technology Conference (VTC Spring), 75th. IEEE, 2012.
[15] 3GPP, TS 37.340, "NR; Multi-connectivity; Overall description; Stage-2" Rel-15, V15.0.0, Dec. 2017.
[16] 3GPP TR 36.888 "Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE"
[17] 3GPP TR 37.868 "RAN Improvements for Machine-type Communications"
[18] 3GPP TR 36.814 "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects"
[19] 3GPP TS 23.501 "URLLC traffic model and QoS parameter", S2-178901, SA WG2 Meeting #124
[20] ITU, Revision 1 to Document 5/57-E, 17 October 2017
[21] N. H. Mahmood, D. Laselva, D. Palacios, M. Emara, M. C. Filippou, D. Min Kim, I. de-la-Bandera, "Multi-channel access solutions for 5G New Radio" in IEEE Wireless Communications and Networking Conference (WCNC), 2019, Marrakech. Accepted for publication.

1 Introduction

1.1 Motivation and scope

1.1.1 Motivation

In recent decades, our world has witnessed a revolution in the evolution of mobile communication systems that started from the first-generation communication networks to the second, third, fourth and now at the fifth-generation networks. This development has been a major driving force behind the progress and development of our world in an advanced and networked environment. Mobile communication has evolved from a service that is available and accessible to just a few people to a service used by the majority of the world's population. Millions of people around the world are connected through mobile communication. It has become an integral part of our daily lives and its applications are ubiquitous in every daily activity. It has pioneered many developments and has demonstrated achievements in various fields such as healthcare, transport, energy, manufacturing, architecture, agriculture, engineering, business, education, meteorology, radio, media and entertainment. Rapid growth in the global mobile device market has been observed in recent years. In addition, this growth is constantly being felt and people will continue to enjoy the proliferation of mobile devices such as smartphones, tablets, laptops, and laptops, along with new and existing services and applications provided by mobile communication systems such as voice calling, video conferencing, online gaming, live video streaming and more. In the digital age, users and devices are increasingly dependent on a variety of applications and services that include creating, accessing / communicating, editing and storing digital content. These developments have been accelerated rapidly by wireless / mobile technologies, which have offered incomparable access / communication opportunities to users. 5G is dominated by the following categories of applications: machine type mass communication (mMTC), enhanced mobile broadband (eMBB) communication, highly reliable communication and low latency (URLLC). These main categories will facilitate scenarios related to critical and demanding applications for implementing smart cities, as well as implementing applications for industry 4.0 and automation aspects. Strict requirements are also expected to ensure reliable and secure service with a very high availability rate.

1.1.2 Scope

The scope of this dissertation is to investigate the potentials of managing networks and services based on artificial intelligence in heterogeneous broadband environments for 5th generation and beyond. This is being achieved by thorough examination of key enablers for creation, management and evaluation of mesh networks as well as modeling and analysis of management in heterogeneous networks for achieving enhanced performance. In addition, aspects of radio frequency resource management in a multi-tenant environment; connection density; network slicing and resource management, selection and allocation are carefully analyzed and evaluated for concluding in the best practices for serving demanding applications in 5G environments and beyond.

1.2 Enablers and proposed solutions

1.2.1 Radio frequency licenses

A broadcasting license is a type of radio spectrum license that provides the licensee with a license to use part of the radio frequency spectrum in a given geographical area for transmission purposes. Permits generally include restrictions, which vary from zone to zone [1]. The range can be divided according to use. According to a graph from the National Telecommunications and Information Administration (NTIA), frequency distributions may be represented by different types of services that vary in size. There are many options when applying for a broadcast license, the FCC determines the range of spectrum to be allocated to licensees in a given area, depending on what is required for that service [1]. Frequencies used by licensees are determined by frequency allocation, which in the United States is determined by the FCC in a classification table. The FCC is authorized to regulate spectrum access for private and government use. However, the Department of Commerce's National Telecommunications and Information Administration has a range for use by the federal government (including the military). In some cases (e.g. Citizens Broadband radio), the public may use spectrum without permission. Commercial users (such as TV, AM / FM radio, and some types of two-way communications) will receive an FCC mapping to a portion of the spectrum, which may be a single frequency or a frequency band. When issuing broadcasting licenses, the FCC relies on "comparative hearings", according to which the most qualified user will be able to use the spectrum to better serve the public interest. The researchers pointed out that this process benefits the established. Violation of the terms of a license (due to a technical error or illegal content) may result in fines or revocation of the license. Licenses have also been compromised due to misrepresentation of the holder or failure to maintain a public record (in the US and Canada). Unauthorized transmission refers to legal devices that are allowed to transmit low-power unlicensed and pirated stations that violate the law.

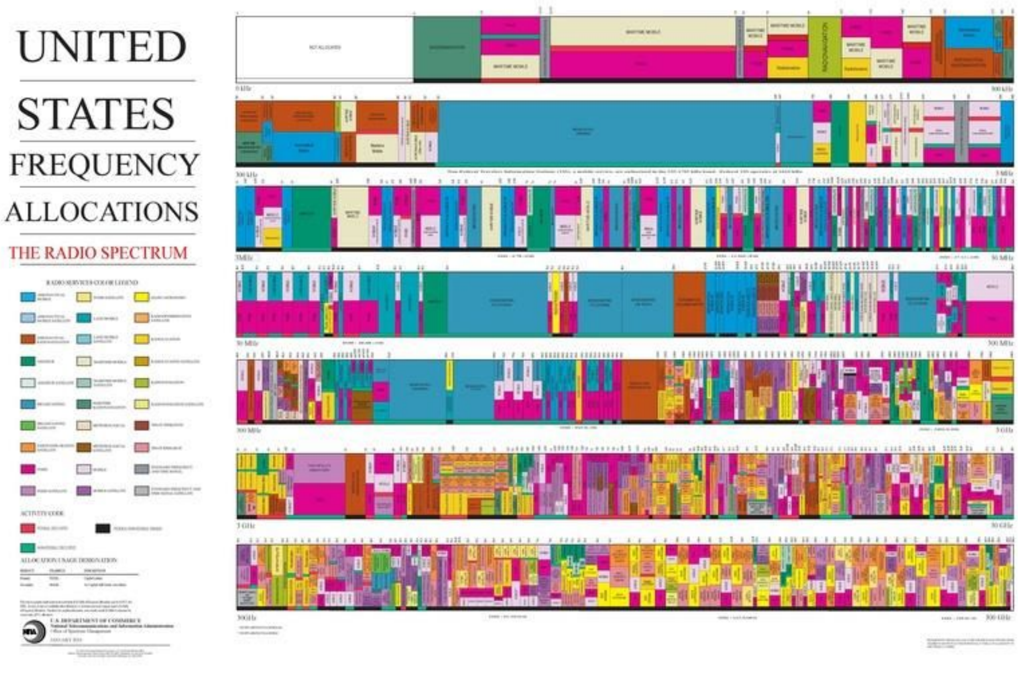


Figure 1-1: US Frequency Distribution Chart 2016. [2]

The process of obtaining a new broadcasting license is time consuming. A transmission engineer first determines an available frequency, which may not be available in an area. If there is a frequency, a technical study shall be submitted to the Broadcasting Authority on request to demonstrate that the licensee will not cause RF interference on existing stations. There is a limited time for the license, once obtained. According to the United States Government Printing Office in 1997, the time would have been around 8 years, however, it has been reduced to five years or less (depending on whether the FCC requires further evaluation) [3]. The license is issued for the first time, with the license receiving approval when it is certified that the license has been executed (after verification to ensure that all parameters are within permissible tolerances). Once an installation has been set up and put into operation, it may be allowed to operate under the program test authority until the license is issued (or rejected). When a station is close to an international border, a license may also need to be approved by the foreign broadcaster for frequency tuning. This is done even if the limit is outside the intended broadcast range of a station, as broadcasting sometimes causes stations to listen outside their service area. Existing stations apply for licenses and license modifications when making changes to their facilities (such as relocating to another site, changing the height of the radio antenna, changing the directional antenna pattern, adding or converting to digital broadcasting). Other situations (such as a change in license city) are covered in U.S. decision-making processes, which may be a prerequisite for moving a station a considerable distance (leaving its original community outside the new coverage area). Temporary situations are covered by the Special Provisional Authority (STA) to operate with a difference from the license or license or a Limited Service License (RSL) to operate for a specified period of reduced validity. While these are FCC and Ofcom terms, respectively, other countries have similar regulations. In the US, court cases can be extended when mutually exclusive requests are received. The FCC opens application window periods of approximately one week. Some applications have been pending for years. Others end up in administrative law courts or arbitration, sometimes with one applicant seeking redemption for another.

1.2.2 Radio Frequencies and Heterogeneous Networks

Based on the above description of licensing and use of frequencies by television stations, radio stations, satellite stations, army, navy, aviation and other government agencies, one can understand the big problem in the use of frequencies by telecommunications providers.

To meet the set standards, telecommunications providers will need to commit further resources to a telecommunications spectrum. However, the success of the services depends to a large extent on governments and regulators. The speed, availability and quality of services depend on the support of these authorities with timely access to the right amount and type of affordable range, under the right conditions. Variation in the amount of the allocated range and the prices paid, means that the services will differ from one type to another, from region to region, but also between countries. This, in turn, directly affects competitiveness and the economy.

The exponential growth of data demand in wireless networks and the approach of theoretical limits to the capacity of wireless connections, forces us to find new solutions and innovative plans to manage the huge data traffic. HetNets will be able to offer an effective solution to the problem of capacitance, using a range of different radio frequency bands at the same time, but they must work properly to have the desired effect. The 5th generation heterogeneous networks will provide a sufficient increase in capacity as they use a multilevel architecture consisting of multilevel cells (e.g. macro cell, small cell, etc.), device to device retransmission through mesh networks, new dynamic spectrum access systems with different users where they will use common radio frequencies, flexible management of radio resources and much more. This multi-layered architecture for increasing capacity creates some challenges in HetNets that need to be studied to analyze and find appropriate solutions to problems that will arise during their use. For example, interference between cells and users remains the biggest challenge in HetNets, providing a fertile ground for the effective study and development of state-of-the-art HetNets. Effective techniques and strategies will allow for additional and substantial increase in capacity by optimizing the minimum available resources.

1.2.3 Methods of better network management

One of the suggested ways to gain high quality 5G access to a larger number of people is through mesh networking. Mesh networking is currently available for other simple types of wireless signals. Mesh networks become crucial for edge points to be connected both to each other and to the central hub in more reliable ways, with a lower risk of losing connectivity or bandwidth. Individual routers called nodes connects with one of the nodes acting as "gateways" to receive the signal. The gateway node then distributes the signal to the other nodes, which act as small satellites to bounce the signal to each other. This allows the signal to be transmitted over a much wider area, but it also has many other benefits:

- You do not need to rely on multiple hubs or stations as the gateway extends from individual hubs.
- Network self-treatment.
- Mesh networks are designed so that if one of the nodes collapses, the network automatically redirects the signal resulting in reduced downtime and easy detection of necessary repairs.
- No cables.
- Coverage of large areas.

Mesh can generate much wider signal areas than a single tower and provide low latencies that could be used to power smart cities more efficiently.

- Nodes make the signal stronger.

Networking and flexible networking will provide the opportunity to connect different devices. 5G mesh networks will change the way we think about connectivity and communication.

Use of new frequencies previously reserved for military or other purposes. This use will be made with spectrum access systems (SAS) using the Citizens Broadband Radio Service (CBRS) initially in the 3.5 GHz band, which represents 150 megahertz of bandwidth in the 3.5 GHz band. A network must use SAS and dynamically manage spectrum usage through an Environmental Detection Network (ESC) to avoid interference with existing users e.g. Navy radar. This system allows users with different rights to initially use the frequency of 3.5GHz, where until now it was used only by the army, navy, air force and in the future to use a band over 25GHz. Users have 3 levels of rights,

- licensed
- lightly licensed
- unlicensed

offering cheaper and more flexible use of the spectrum, only when needed, to give extra resources to the system when needed. This mechanism can be used in combination with other methods to offer even greater impetus to the communications of the 5th and future generations.

Multiple mechanisms and methods will be examined to achieve the best use of the existing frequencies. Multiple connections, multi-level network ecosystem, new user-level control and integration methods, and protocol customization can provide opportunities to redesign or improve various functions, (e.g. interference management, power control, RAN control) by outsourcing services and applications to adequate resources to meet the requirements in an efficient and sustainable manner. Carrier Aggregation and Advanced RAN Coordination is a combined solution that optimizes the coverage, capacity and latency of 5G medium and high bands. Together, these solutions allow service providers to optimize the use of their spectrum components when developing 5G. A better 5G network will provide more subscribers with higher data rates, while allowing the host of new low latency applications. Coupling 5G low-bandwidth with 5G mid-bandwidth can improve mid-bandwidth coverage, expanding the range and increasing the population that can be supported by the mid-bandwidth.

There are many industries that do not take full advantage of mobile networks due to the different requirements that they have. A company may need to offer very low communication times to its machines to be able to rely on a mobile network. Another may need high speeds for their communication or large bandwidth. Creating a separate physical structure that specializes in each area is not practical, so current mobile network systems use a "one size fits all" approach. However, 5G network slicing may provide these businesses with the personalization they need to enable them to operate, but it uses the same physical infrastructure. Network slicing allows multiple, personalized networks to run using the same shared infrastructure. Networks can offer different functions, tailored to individuals or services, without the need for separate network structures for each. One mobile network is essentially "sliced" into different customizable parts that meet the needs of different consumers. Instead of companies, developers and consumers having to adapt to what mobile networks offer, the 5G mobile network should be able to adapt to consumers' needs. The customer experience will be like the network they use being personalized to them, but the infrastructure that provides these features will also offer different features for other customers. The network slice will give current business models a better way to connect, but it also promotes the development of future technologies that are not feasible in the current system. Remote machine operation, smart metering and augmented reality are some areas that can be further developed using 5G and network slicing options.

All the methods presented here and much more, were researched, analyzed and tested during the research activity, giving results based on each use case. The following chapters present in detail the methodologies, algorithms and results that were recorded and presented in articles, magazines and books.

1.2.4 Dissertations' Contribution

The dissertation deals with management of networks and services based on artificial intelligence in heterogeneous broadband environments for 5th generation and beyond. As a result, its main contribution can be categorized at the following topics:

- Mesh networks: The aim of this topic is to design, develop and validate mechanisms for deciding on the creation and set-up of flexible and dynamic networks based on moving access points (MAPs) for mobile client nodes. Mobile client nodes can be smart/connected cars in civilian use cases, or any kind of operational vehicles and/or robots/drones in military cases. Mesh networks will be formed in an ad hoc manner and will be dynamically reconfigured through the suitable self-organization and adjustment of the position and coordination of MAPs so as to provide appropriate Quality of Service (QoS) levels towards mobile client nodes. Such applications can be exploited in a military context in order to establish communications in remote areas and battlefields where fixed communication infrastructure is not available or destroyed. Mesh can create more reliable connections in the same way as other wireless networks and in the future use unlicensed or mmWave spectrum. This topic is elaborated in Chapter 2 of this dissertation.
- Modeling and analysis of management in heterogeneous infrastructure: This topic elaborates on the main 5G requirements and presents the status and challenges in hardware and software development. It focuses on the status and challenges in 5G wireless communications by focusing on physical layer, MAC and radio resource management (RRM). The chapter investigates the benefits of machine learning in 5G network management. It describes the use of various machine learning mechanisms for the service classification problem and also investigates the performance of different algorithms. This topic is elaborated in Chapter 3 of this dissertation.
- Management of shared resources in a multi-provider environment: Today, lack of dynamic control across wireless network resources is leading to unbalanced spectrum loads and a perceived capacity bottleneck. As a result, the objective of this topic is to present a new framework for MAC and RRM layers for supporting extended Dynamic Spectrum Access (eDSA) and requirements of the next-

generation networks by addressing traffic allocation over heterogeneous wireless technologies, and better load balancing across available spectrum bands, capacity boosting through aggregation of available resources whilst ensuring fair coexistence. This topic is elaborated in Chapter 4 of this dissertation.

- Network Slicing: The purpose of this work is to design, develop and validate mechanisms for creating and deciding on the dynamic resource allocation of network slices. Our proposed algorithm reconfigures and adjusts the slices so as to provide appropriate Quality of Service (QoS) levels towards mobile client nodes. We demonstrate how our proposal provides better quality of communication and satisfaction for the individual user categories based on specific traffic requirements. This topic is elaborated in Chapter 5 of this dissertation.
- Sharing and allocation of resources for new 5G services (URLLC, eMBB, mMTC) : In this work, we propose an algorithm for enabling dynamic channel assignment in the 5G era that receives information about the interference and QoS levels and dynamically assigns the best channel. This algorithm is compared to state-of-the-art channel assignment algorithm. Results show an increase of performance, e.g., in terms of throughput and air interface latency. Finally, potential challenges and way forward are also discussed. This topic is elaborated in Chapter 6 of this dissertation.
- Radio Frequency Resource Management (RRM) in a multi-connection environment: This work presents essential aspects of 5G system level simulator for enabling advanced component validations and optimizations. System level simulations in the 5G and beyond era, consider demanding use cases with high load and very limited latency in order to cover services such as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC) and ultra-reliable low-latency communications (URLLC). As such, appropriate configuration, environment and network models need to be defined in order to proceed to performance evaluation. A framework for multi-connectivity management has been integrated and assessed in a load-imbalanced scenario. In addition, connection density is considered as is plays an important role in these new environments. The usage of narrowband technologies is encouraged, especially for small and frequent transmissions. As a result, the provided evaluations consider these assumptions in order to show the number of devices that can be supported with a specific QoS. This topic is elaborated in Chapter 7 of this dissertation.

1.3 Dissertation's Structure

The dissertation is structured in chapters and each chapter provides a detailed description and results of the conducted research activities for achieving the overall goal of managing networks and services based on artificial intelligence in heterogeneous broadband environments for 5th generation and beyond.

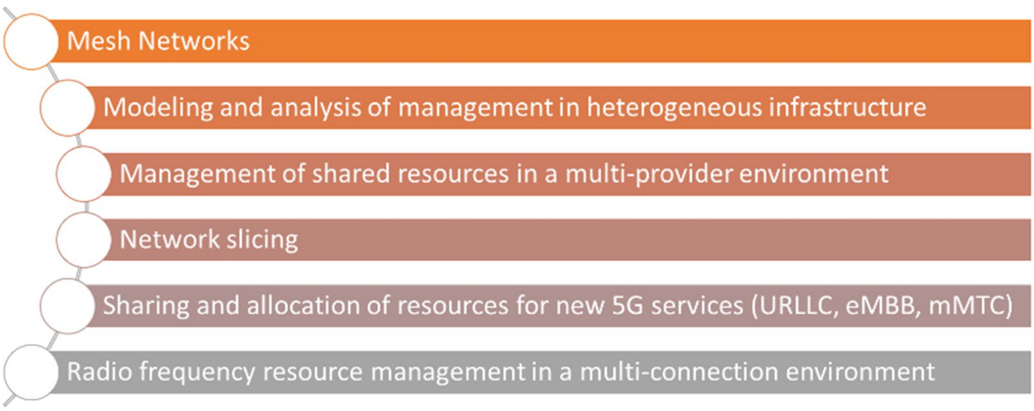


Figure 1-2: Dissertation's Structure.

Chapter 1 provides the main introduction and motivation of our work and sets the requirements of the necessary research for managing networks and services based on artificial intelligence.

Chapter 2 investigates and evaluates mesh networks for achieving enhanced performance, even when connectivity is challenging. Research of the proposed solution resulted in the following publication:

- Andreas Georgakopoulos, Ioannis-Prodromos Belikaidis, Kostas Tsagkaris, Vera Stavroulaki, Panagiotis Demestichas, "Wireless Access Infrastructure Expansions Through Opportunistic Networks of Moving Access Points", in Proc. 2016 European Conference on Networks and Communications (EuCNC), Athens, 27-30 June, 2016, pp. 163-167.

Chapter 3 focuses on the modeling and analysis of management for heterogeneous infrastructure and the research of the proposed solution resulted in the following publication:

- Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos, Stavroula Vassaki, Orestis-Andreas Liakopoulos, Vassilis Foteinos, Panagiotis Vlacheas, Panagiotis Demestichas, "Emerging technologies in software, hardware and management aspects towards the 5G era: Trends and Challenges", 5G Networks: Fundamental Requirements, Enabling Technologies, and Operations Management, Oct. 2018.

Chapter 4 elaborates on radio frequency resource management in a multi-provider environment with emphasis on hierarchical radio resource management scheme and the research of the proposed solution resulted in the following publications:

- Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos, Valerio Frascolla, Panagiotis Demestichas, "Management of 3.5-GHz Spectrum in 5G Dense Networks: A Hierarchical Radio Resource Management Scheme", IEEE Vehicular Technology Magazine.
- Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Panagiotis Demestichas, Benoit Miscopein, Marcin Filo, Seiamak Vahid, Bismark Okyere, Michael Fitch, "Multi-RAT Dynamic Spectrum Access for 5G Heterogeneous Networks: The SPEED-5G Approach", IEEE Wireless Communications, vol. 24, no. 5, pp. 14-22, October 2017.

Chapter 5 discusses the notion of network slicing and how slicing can lead to better service provisioning in demanding environments by blending different traffic types (e.g. URLLC, eMBB etc.). Research of the proposed solution resulted in the following publication:

- Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Kostas Tsagkaris, Zwi Altman, Sana Ben Jemaa, Michalis Michaloliakos, Panagiotis Demestichas, Nikolas Mitrou, "5G Radio Access Network Slicing: System-Level Evaluation and Management", IEEE Vehicular Technology Magazine, vol. 14, no. 4, pp. 49-55, December 2019.

Chapter 6 provides useful insights on sharing and allocation of resources with emphasis on dynamic channel assignment for the new 5G services such as URLLC, eMBB and mMTC. Research of the proposed solution resulted in the following publication:

- Ioannis-Prodromos Belikaidis, Stavroula Vassaki, Andreas Georgakopoulos, Aristotelis Margaris, Federico Miatton, Uwe Herzog, Kostas Tsagkaris, Panagiotis Demestichas, "Context-aware Radio Resource Management Below 6 GHz for Enabling Dynamic Channel Assignment in the 5G era", EURASIP Journal on Wireless Communications and Networking, December 2017, 2017:162.

Chapter 7 elaborates on RRM issues in a multi-connection environment with emphasis on 5G component carrier management and is being evaluated with system level simulations. In addition, the requirements of ultra-high connection density of IoT devices in mMTC environments are evaluated. Research of the proposed solution resulted in the following publications and parts of this work was included also in an ITU study regarding evaluation of 5G.

- Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos,, MIsabel de la Bandera-Cascales, David Palacios, Raquel Barco, Panagiotis Demestichas, “5G Component Carrier Management Evaluation by Means of System Level Simulations”, in Proc. 2019 European Conference on Networks and Communications (EuCNC), Valencia, 18-21 June, 2019.
- Andreas Georgakopoulos, Evangelos Kosmatos, Ioannis-Prodromos Belikaidis, Martin Kurras, Lars Thiele, Panagiotis Demestichas, “Enabling Advanced 5G Component Validations and Optimizations by Means of System Level Simulations Platform, Abstractions, Models, Results and Further Challenges”, in Proc. 2018 European Conference on Networks and Communications (EuCNC), Ljubljana, 18-21 June, 2018.
- ITU Document 5D, FINAL EVALUATION REPORT FROM THE 5G INFRASTRUCTURE ASSOCIATION ON IMT-2020 PROPOSALS IMT-2020, Feb. 2020, https://www.itu.int/md/R19-WP5D-C-0053/en
(on behalf of the Independent Evaluation Group 5G Infrastructure Association, edited by Dr. Werner Mohr, Nokia Deutschland).

1.4 References

[1] Nuechterlein, Jonathan; Weiser, Philip J. (2005). Digital Crossroads. Cambridge, Massachusetts: The MIT Press. p. 235.

[2] https://en.wikipedia.org/wiki/File:United_States_Frequency_Allocations_Chart_2016_-_The_Radio_Spectrum.pdf

[3] The Museum of Broadcast Communications - Encyclopedia of Television". www.museum.tv. Retrieved 19 March 2018.

2 Mesh networks

The main aim of this chapter is to design, develop and validate mechanisms for flexible topology and specifically mesh networks based on moving access points. Mobile access points can be smart/connected cars or any kind of vehicles and/or robots/drones. Mesh networks can effectively support a more reliable, dynamic connection, with network topography automatically forming to create the most efficient route from point-to-point to provide appropriate Quality of Service (QoS) levels towards mobile client nodes. These networks can be exploited in any context in order to establish communications and ensure that data and applications can be securely delivered and controlled across a wide range of environments. In cases of rural areas, remote areas, areas that were affected by earthquakes and even battlefields where fixed communication infrastructure is not available or destroyed.

2.1 Introduction

In the world of wireless networks, devices and services, the expectations of users are shifting towards greater, constantly available connectivity. Mobile-connected devices are expected to exceed thousands of billions within the next years, including diverse devices ranging from smartphones to connected vehicles. In this context, significant increases in network capacity are required. However, infrastructure is expensive to build and maintain. In addition, mobile communication networks (either in a commercial or in a military context) are faced with challenging situations such as moving hotspots, areas with difficult morphology where it is difficult to set up infrastructure or areas where infrastructure has become unavailable due to natural disasters or hostile actions. Also, simple centralized connections does not offer the reliability or flexibility needed to support the wide variety of use cases. These continuously increasing requirements and challenges motivate the quest for further efficiency in resource provisioning. Efficiency can be coupled with targets like: (i) the higher utilization of resources; (ii) extra coverage; (iii)latency; (iv) reduction of transmission powers and energy consumption; (v) reduction of operational expenditures (OPEX) and capital expenditures (CAPEX).

2.2 Main motivation

The motivation for this research is to address these challenges with the use of flexible networks of Moving Access Points. Mesh system networks can be temporary, coordinated extensions of the infrastructure. They can be dynamically created, in places and at the time they are needed and can comprise network elements of the infrastructure, Moving Access Points (MAPs) and mobile client nodes/devices potentially organized in an infrastructure-less manner. These Networks and MAPs represent efficient means for offering communication services with reduced CAPEX, due to the absence of permanent infrastructure, and increased resource utilization. MAPs are capable of autonomously moving and establishing an opportunistic radio network in short time, with limited centralized management [1],[2].

2.3 Main concept

The concept of this research is to design, develop and validate mechanisms for deciding on the creation and set-up of flexible networks of MAP entities for mobile client nodes. Mobile client nodes can be smart/connected cars in civilian use cases, or any kind of operational vehicles and/or robots/drones in military cases. Flexible networks can be formed in an ad hoc manner and will be dynamically reconfigured through the suitable self-organisation and adjustment of the position and coordination of MAPs so as to provide appropriate Quality of Service (QoS) levels towards mobile client nodes. The decision making process will be supported by mechanisms for Knowledge building based on artificial intelligence techniques.

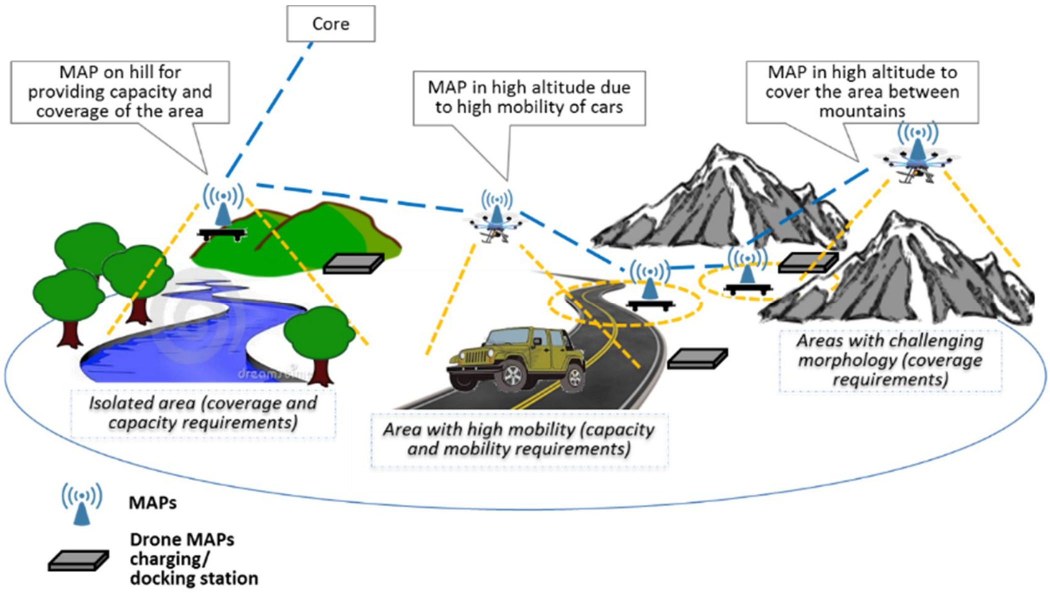


Figure 2-1: Illustration of main concept.

This work considers diverse types of MAPs, altitude in the deployment of MAPs and more details in terms of mobility aspects of client nodes.

2.4 Problem Statement

This section provides the problem statement, specifically, a remote area where deployed, wireless infrastructure is not widely available will be considered. The area shall comprise network entities, i.e., the set of moving access points (MAPs) denoted as 'M' and the set of APs denoted as 'A' that offer access to remote users denoted as 'U'. In addition, it will be assumed that users require services at specific QoS levels (e.g., in terms of bitrate, latency, etc.) denoted as 'q', while a capacity 'cap' (e.g. in terms of number of users that can be served) and a transmission range based on standardized propagation models will be assigned to the MAPs and the fixed access points (APs).

The work focuses on the solution of a complex optimization problem that can be generally expressed as follows:

"Given information and knowledge on:

- The context that has to be handled, in terms of (i) a set of mobile clients that need coverage, (ii) mobility and traffic profiles of the client nodes, (iii) radio quality, (iv) options for connecting to wide area networks, (v) the locations of docking/charging stations for drone MAPs, (vi) the current locations of the MAP elements, a (potentially large) set of candidate final positions to which the MAP entities can move. MAPs can assume position characterized by latitude, longitude and altitude, i.e. of the form(x, y, z);
- The capabilities of the mobile client nodes and the MAP entities in terms of (i) communication networking (e.g., RATs and spectrum that can be operated, capacity and coverage that can be provided etc.), (ii) physical movement, (e.g., possible speed of the element, path-types and obstacles that it can overcome, etc.); (iii) the type of the MAP (e.g. drone/flying MAP);

- Potential policies that need to be followed by the mobile client nodes, the MAP entities and their network;

Find the optimal:

- Positions to which each MAP entity should move;
- Configuration of the radio network of the MAP entities;
- Allocation of nodes to MAPs
- Selection of nearby docking/charging stations for drone MAPs

So as to ensure connectivity of the appropriate QoS to mobile nodes."

MAPs can be considered as RAT-agnostic, since the proposed solution is not bound to a specific RAT e.g., 3GPP-based or WiFi-based etc. That means that the concept will work with any RAT, within the limitations of the particular technology. More specifically, this work considers more, diverse types of MAPs, also altitude in the deployment of MAPs and will go into more details in terms of mobility aspects of client nodes. Solutions are validated based on applications in a wide set of very indicative test cases and through simulation. Previous work of various research groups have also tackled similar issues by proposing various solutions such as [4][5]. In order to achieve the main objective, we conducted work for addressing the following technical challenges:

- Acquisition and exchange of contextual information (e.g., of (i) a set of mobile clients that need coverage, (ii) mobility and traffic profiles of the client nodes, (iii) radio quality, (iv) options for connecting to wide area networks, (v) the locations of docking/charging stations for drone MAPs, (vi) the current locations of the MAP elements, a (potentially large) set of candidate final positions to which the MAP entities can move).
- Development and evaluation mainly through simulations of optimization strategies on the joint radio network optimization and position selection, for enabling the creation of opportunistic networks of moving access points in a highly dynamic environment.

2.5 Mathematical Formulation

In this section the main problem is mathematically formulated. Specifically, let *M* be the set of the MAP entities, *A* will be the set of the APs, while *U* will denote the set of users (with mobile clients) and *D* will be the set of docking stations e.g., for charging drones. In addition, *L* will denote the set of the locations at which the network elements (MAPs, APs) can be placed, while $e(i)$ will depict the element that is located at $i \in L$. Also, each MAP $m \in M$ has a capacity cap_m. Capacity can reflect for instance the number of users that are served by a specific MAP. In order to avoid congestion issues (e.g., a lot of users are served at the same time by one only MAP), we can set low values of capacity to MAPs.

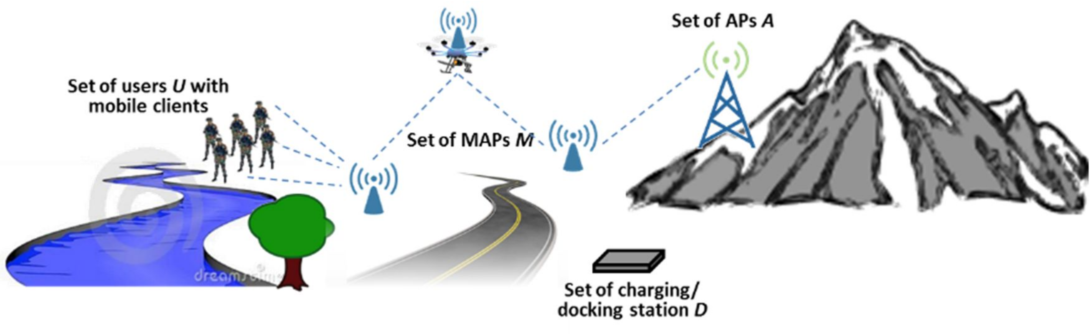


Figure 2-2: Identified mathematical sets.

Moreover, the following decision variables are considered:

$$X_{mi} = \begin{cases} 1, if \text{ MAP } m \in M \text{ is located in } i \in L \\ 0, otherwise \end{cases}$$

$$Y_{mn} = \begin{cases} 1, if \text{ MAP } m \in M \text{ is connected with MAP } n \in M \\ 0, otherwise \end{cases}$$

$$Z_{um} = \begin{cases} 1, if \text{ user } u \in U \text{ is connected with MAP } m \in M \\ 0, otherwise \end{cases}$$

$$Q_{ma} = \begin{cases} 1, if \text{ MAP } m \in M \text{ is connected with AP } a \in A \\ 0, otherwise \end{cases}$$

$$D_{md} = \begin{cases} 1, if \text{ MAP/drone } m \in M \text{ is not connected with docking station } d \in D \\ 0, otherwise \end{cases}$$

Furthermore, the connection of two entities (e.g. when two MAPs are connected, when a MAP is connected to an AP, or when a MAP serves a user) results in a communication cost *ct*. In general, this cost depends on the frequency used for the communication and the distance of the entities. As a result *ct* values can be high for entities that are far away between each other, in order to let the algorithm determine closer entities (if available). Also, the movement of drone/MAP to a docking station for charging results in a moving cost *mc*. In general, this cost depends on the distance of the drone/MAP from the current position to the docking station.

Accordingly, the overall optimization problem can be formulated as follows:

Minimize

$$OF = \sum_{m \in M} \sum_{i \in L} (X_{mi} \cdot ct(m,i)) + \sum_{m \in M} \sum_{\substack{n \in M \\ m \neq n}} (Y_{mn} \cdot ct(m,n)) + \sum_{m \in M} \sum_{u \in U} (Z_{um} \cdot ct(m,u)) + \sum_{m \in M} \sum_{a \in A} (Q_{ma} \cdot ct(m,a)) + \sum_{m \in M} \sum_{d \in D} (D_{md} \cdot mc(m,d))$$

Constraints:

$$\sum_{i \in L} X_{mi} = 1, \ \forall m \in M$$

$$\sum_{m \in M} X_{mi} \leq 1, \ \forall i \in L$$

$$\sum_{n \in M} Y_{mn} \geq 1, \ \forall m \in M$$

$$\sum_{m \in M} Q_{ma} \geq 1, \ \forall a \in A$$

$$\sum_{m \in M} Z_{um} = 1, \ \forall u \in U$$

$$\sum_{m \in M} D_{md} = 1, \forall d \in D$$

$$\phi_m \leq cap_m, \forall m \in M$$

Where

$$\phi_m = \sum_{u \in U} Z_{um} + \sum_{n \in M} \left(Y_{mn} \cdot \phi_n \right), \forall m \in M$$

The objective function *OF* monitors the location of all MAPs and calculates the total communication costs that are related from these locations, as well as the finding of the nearest docking station for charging (of drone MAPs). The first term of the function illustrates the communication cost related to the MAPs location. The second term depicts the communication cost due to the connections among MAPs. The third term denotes the communication cost because of the connections among users and MAPs. The fourth term depicts the communication cost of the connections of the MAPs with the APs. The fifth term depicts the cost related to the docking station where the drones should go for charging.

Regarding the constraints, the first constraint denotes that every MAP can be placed at one only location. The second constraint denotes that at each location one MAP at most can be placed. The third constraint depicts that every MAP should be connected with at least another MAP (in order to realize the opportunistic network). The fourth constraint denotes that all APs should be connected with at least one MAP and the fifth constraint denotes the fact that each user can be served by one only MAP at a time. The sixth constraint denotes that each docking station can serve one drone/MAP at a time. Also, φ_m represents the MAP's load which cannot exceed its capacity cap_m.

2.6 Specifications

2.6.1 Simulated Annealing

Simulated annealing (SA) algorithm is a well-known way to reach to an optimal solution (e.g., with respect to the MAPs positions as this study suggests). The SA algorithm takes into account simulation steps and in its step it replaces the current solution by a random neighbouring solution $s_i \in S$, chosen with a probability that depends on the difference between the corresponding *OF* values and on a global parameter temperature *T* which is gradually decreasing during the process (cooling) with a cooling rate *c*. In this manner, the algorithm searches the solution space in a random way and at the same time it avoids becoming stuck at local minima [6]. The input of the algorithm can include the following aspects:

Input:

- Sets of MAPs and their positions with respect to users, other MAPs, final APs and docking stations (if drone MAPs are considered)
- An initial solution $s_0 \in S$
- The initial temperature T_0
- The temperature's cooling rate *c*
- The maximum number of iterations i_{max}

The process of the algorithm is described in Figure 2-3. It starts by getting the input and initialization of the solutions and the temperature. SA starts with a relatively high value of temperature *T*. The algorithm proceeds by attempting a certain number of neighboring moves at each temperature, while the temperature parameter is gradually dropped. When the lower allowed temperature is reached the algorithm ends. The output of the algorithm will lead to a sub-optimal solution $s_{final} \in S$. Apart from the SA algorithm (which has been widely studied e.g., in [7], [8] and [9], it can be mentioned that suitable distributed heuristic (e.g., bio-inspired solutions) can follow as well.

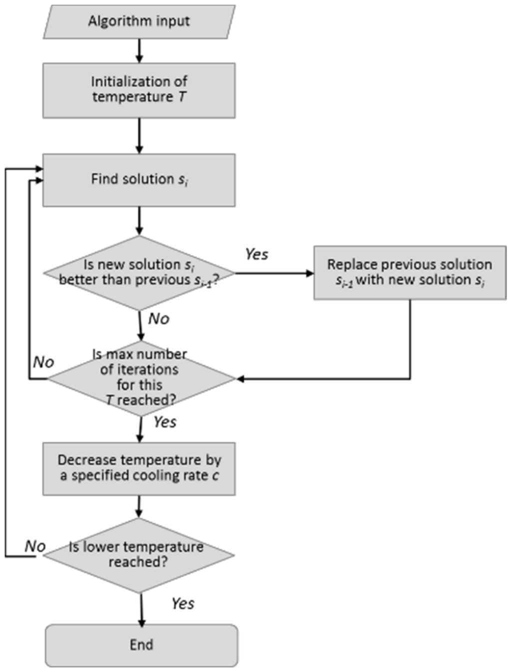


Figure 2-3: Flowchart of simulated annealing.

2.6.2 Maximum flow algorithm

As soon as the positions of MAPs are decided e.g., according to the SA or other algorithms, then the flow of traffic between MAPs can be optimized (in order to resolve any potential congestion issues e.g., when many users are trying to connect to only one specific MAP). An algorithm based on the Ford-Fulkerson maximum flow algorithm [10] [11] [12] can be used. The main notion of the algorithm is that each user $u \in U$ tries to connect to an access point $a \in A$, by making use of MAPs $m \in M$. Each user should find a set of paths to some of the indicated APs in order to gain access to a network. Each path has a set of MAPs/nodes, the capacity of these MAPs cap_m, and the communication cost ct of the links. Capacity can reflect for instance the number of users that are served by a specific MAP. In order to avoid congestion issues (e.g., a lot of users are served at the same time by one only MAP), we can set low values of capacity to MAPs. Also the communication cost depends on the frequency used for the communication and the distance of the entities. As a result ct values can be high for entities that are far away between each other, in order to let the algorithm determine closer entities (if available).

Input:

The input of the algorithm consists of the following:

- Sets of MAPs and APs that are considered in the solution;
- Set of users that want to connect to MAPs and APs;
- Paths from sources (users) to destinations (APs) through MAPs. Each path can originate from a ‘virtual’ source where all users are connected and ends to a ‘virtual’ destination where all APs are connected;

- Capacity of each MAP and AP (to know how many users can be supported).

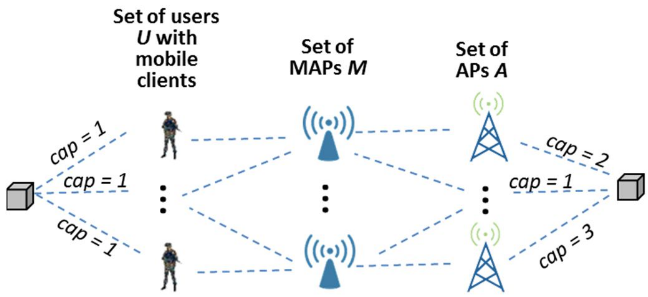


Figure 2-4: Main sets considered by the algorithm.

The process of the algorithm is described in Figure 2-5. It starts by getting the input and assuming that a path consists of many links which are distinguished by starting, ending points and capacity (which can be associated also to the available radio access technology-RAT). For instance, 3G links for sure would have lower capacity value compared to 4G etc. Once, the input is retrieved, the algorithm picks one of the discovered paths according to the breadth-first search method which yields the shortest path. From the selected path, it finds the link with the smallest capacity. Then, it sets the flow of the path equal to the smallest link capacity and updates the residual capacities of the rest of the links. The algorithm continues until there is no unchecked path from source to destination. Once all paths are checked, a set of MAPs from source to destination is created. As a result, the problem of achieving wireless access infrastructures expansions through opportunistic networks of MAPs is solved.

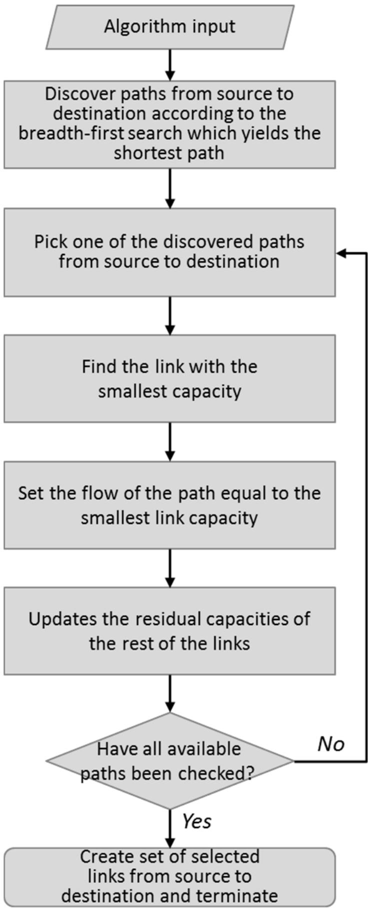


Figure 2-5: Flowchart of the algorithm.

In order to evaluate the performance of the proposed algorithm, several simulations were performed. Also, the simulation environment that was used has been described in detail in order to show the potentials of the tool and the solution for the realization of networks consisting of MAPs. Aspects like position, speed and height of MAPs have been taken into account in order to run various scenarios with different parameters and check their respective performance.

2.7 Development

2.7.1 Simulation Environment

For the evaluation of our model, we use one of the broadly used simulation tools in academy which is a very powerful open source network simulator OMNeT++ [2].

In order to allow the most accurate modeling of MAPs (Rovers and Drones) and UEs movements a hybrid simulation framework is required which is composed of the network simulator OMNeT++, a road traffic simulator SUMO [3] which is well-established in the domain of traffic engineering and the appropriate framework that combines those two simulators, called VEINS [1].

SUMO (Simulation of Urban Mobility) is an open source microscopic traffic simulator licensed under General Public License (GNU) and developed by Institute of Transportation Systems at the German Aerospace Center using C++ standard. It allows users to create a road network of their preferences containing buildings and streets or to import a road network from different format and convert it into a

SUMO network. Also each vehicle can be modeled explicitly, in order to move individually through the network and has their own route updating the position of each vehicle every time step, which gives SUMO the feature of time-discrete vehicle movement. This traffic simulator also provides an OpenGL graphical user interface. Traffic simulation in SUMO can be conducted in two ways as described below and the overview of the simulation process is given in Figure 2-6.

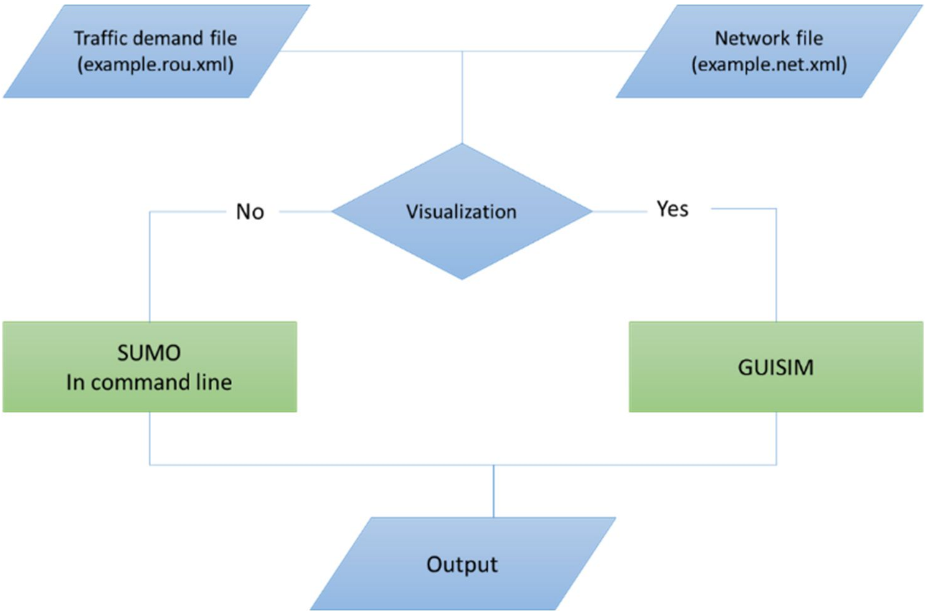


Figure 2-6: Flowchart of traffic simulation process of SUMO.

OMNeT++ is an extensible, modular, component-based C++ simulation library and framework, primarily for building network simulators. "Network" is meant in a broader sense that includes wired and wireless communication networks, on-chip networks, queuing networks, and so on. Domain-specific functionality such as support for sensor networks, wireless ad-hoc networks, Internet protocols, performance modeling, photonic networks, etc., is provided by model frameworks, developed as independent projects. OMNeT++ offers an Eclipse-based IDE, a graphical runtime environment, and a host of other tools. There are extensions for real-time simulation, network emulation, database integration, System C integration, and several other functions.

VEINS is an open source framework for vehicular network simulations consisting of SUMO and OMNeT++ simulators to tender a complete suite for Inter Vehicle communication (IVC). It was designed by Transportation and Traffic Science community. Veins is part of MiXiM framework of OMNeT++ adding support for IEEE 802.11p and IEEE 1609 family - WAVE technology. Thus, this framework handles Wave Short Messages (WSM) and provides beaconing WAVE services, access categories for QoS (Quality of Service) and multi channels operations. The bidirectional communication between OMNeT++ network simulator and SUMO road network simulation is shown in Figure 2-7.

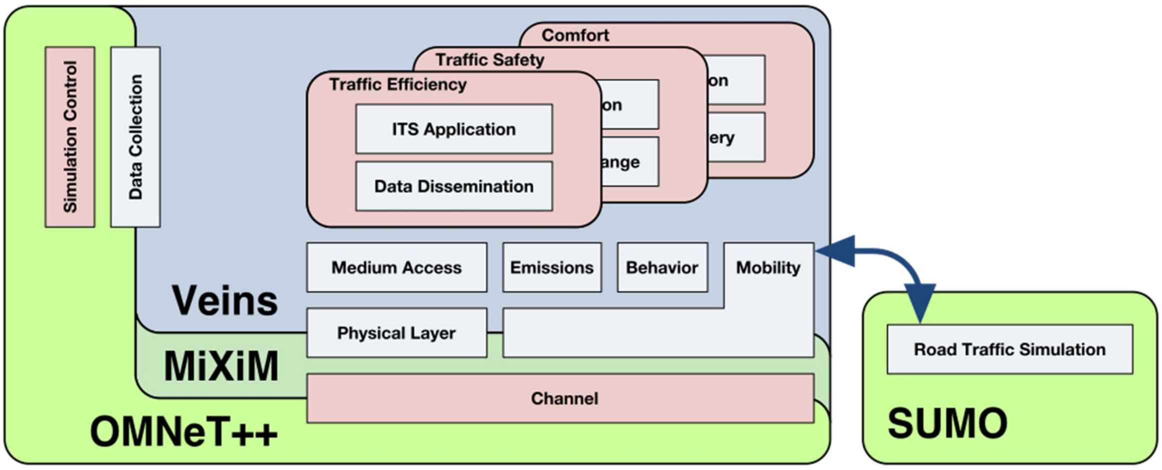


Figure 2-7: Flowchart of communications between SUMO-VEINS-OMNeT++.

2.7.2 Simulation settings

- The simulation will start the computations of the Simulated Annealing (SA) algorithm when the roads are fully populated and run for s seconds.
- Beacon intervals are set to 1 sec, for example 5 messages will be transmitted and received in about 5 seconds.
- The SA algorithm will run every 5 seconds and run for 200 iterations, i.e. 1000 seconds in simulation time.
- Virtual Source and Sink are used in order to better simulate our algorithm. Thus virtual Source is linked with the UEs and virtual Sink is linked with APs respectively.
- For the map (Figure 2-8) we have created from scratch, in sumo simulator a 7 by 7 grid with two lanes in every direction. Two diagonal Aerial "roads" (illustrated with red color) used by Drones in order to provide better coverage in situations that rovers cannot.
- One junction from another is 100 meters apart, giving a total of 490000 square meter area.
- UEs, MAPs and Aps have a starting wireless transmission coverage range of about 150, 200 and 250 meters respectively (mentioned as 1st case) and then this range expands.
- All the entities in our simulation can exchange information regarding their ids, speed, position and direction of movement and even their battery life. Hence every node have the knowledge of the surrounding area and can create a list of its one hop neighbors.
- From the exchange messages an adjacency matrix of all the network is created.
- We compute the distance between every two entities that are used for the Objective Function (OF) computation. The objective function has been elaborated and presented in detail in the second deliverable (D2) of this study.
- For the UEs and MAPs (except Drones) we have set a maximum speed of 5m/s, i.e. about 18Km/h.
- Drones which fly above ground have speeds at about 36Km/h (10m/s), but they lower their speed when near APs or UEs giving them more connection time with those entities and move with their higher speeds during the middle sections of their aerial lanes (Figure 2-8).

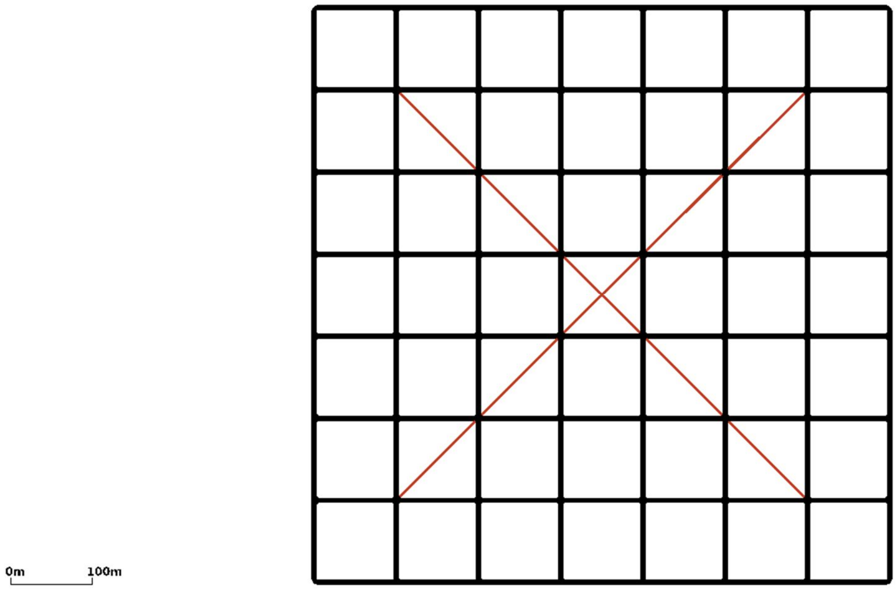


Figure 2-8: Grid created in SUMO for the simulation.

Figure 2-9 illustrates the defined topology with all the entities and their instantaneous positions. In our experimentation topology participate two groups of UEs, two groups of MAPs; the Rovers presented as MAP_CAR in this screenshot of the SUMO simulator and the Drones as MAP_DRONES and also two stationary APs at the positions 1 and 2 in our grid. The UEs can move inside the light blue area and the MAPs only at the yellow area while the MAP Drones can fly anywhere in the red X road providing better coverage inside our map.

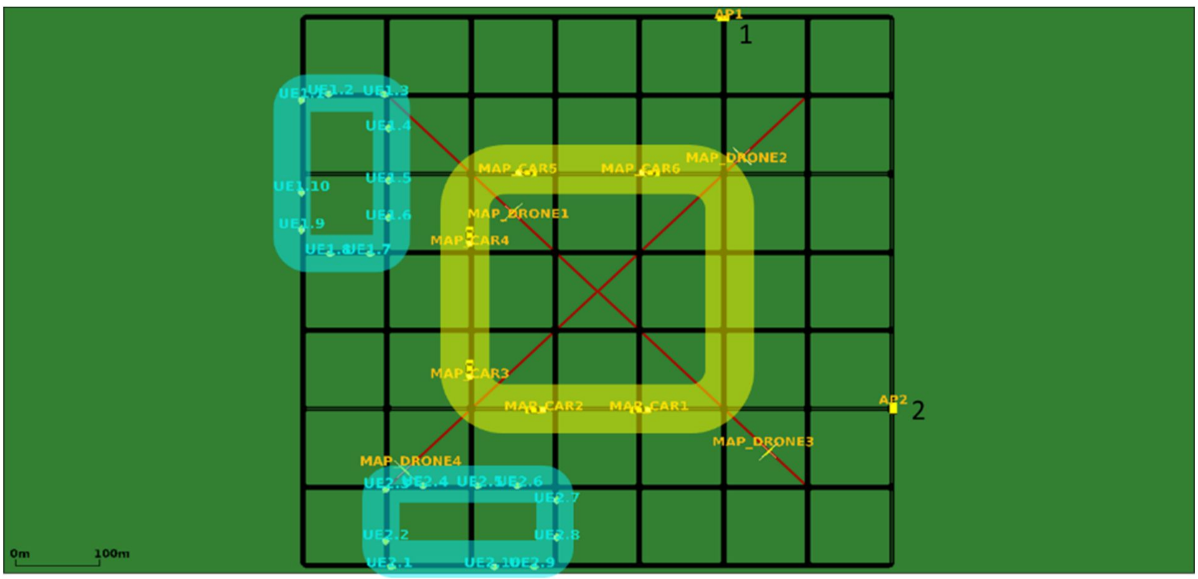


Figure 2-9: Screenshot from SUMO with the simulation topology with the different entities.

Figure 2-10 illustrates a screenshot from the OMNet++ simulator. Specifically, the blue arrows correspond to indicative paths of the beacons that are send from one node (MAP) to its vicinity, demonstrating the coverage area of MAPs in our simulation.

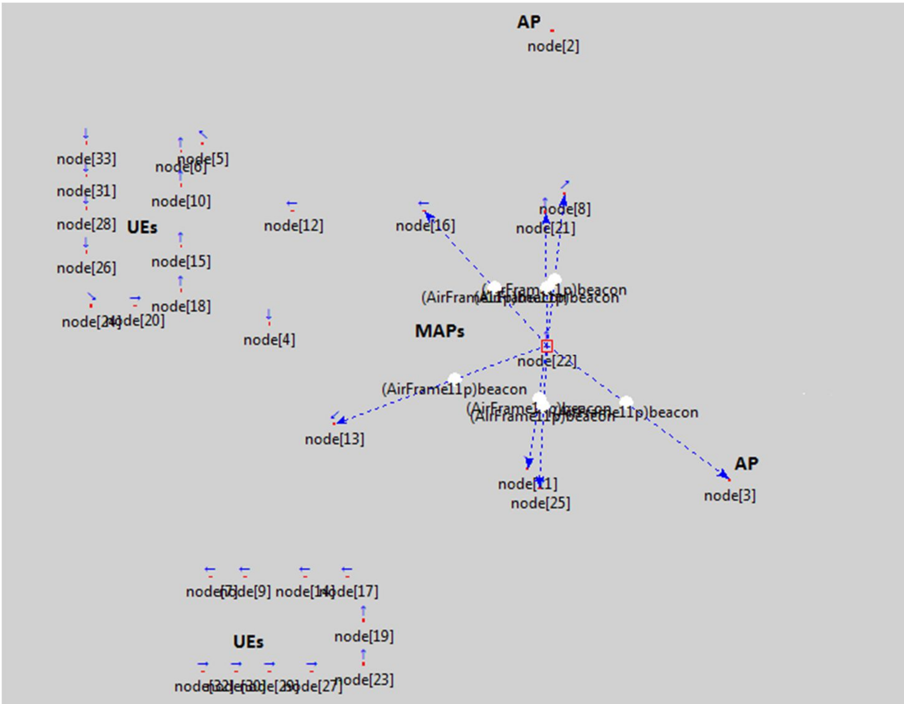


Figure 2-10: Screenshot from the OMNeT++ simulator illustrating communication with beacons.

These frameworks were selected due to inclusion of a comprehensive suite of models to make vehicular network simulations as realistic as possible, without sacrificing speed and also more real-life environments which can be introduced in future work.

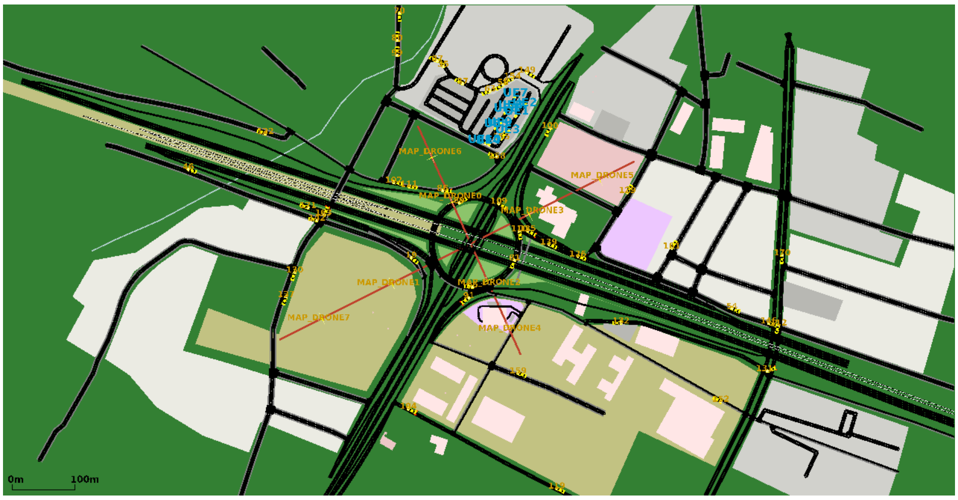


Figure 2-11: Screenshot of the map and entities topology from SUMO simulator.

2.8 Evaluation

In this section, comparative results for the evaluation of the solution will be provided.

2.8.1 Methodology

In order to evaluate the MF & SA algorithm, we have created some Scenarios in our simulation with the general values of the parameters presented in Table 2-1 .

Table 2-1: General Simulation parameters

Parameter	Value
Terrain size	700m x 700m
Grid size	7 x 7
Number of APs	2
Number of UEs	20

2.8.2 Results

Table 2-2 presents all the test cases we have used in our simulation already presented in previous deliverables (Report 3 and Report 4) but also we have included 3 more experimentation cases (10,11 and 12).

Our simulation now has a density that ranges from 10 to 16 MAPs with 32 up to 38 nodes in total participating.

Table 2-2: Simulation parameters.

Test Case	Drone Number	Car Number	UE Range (m)	MAP Range (m)	AP Range (m)
1	4	6	150	200	250
2	4	6	170	220	270
3	4	6	200	250	300
4	6	6	150	200	250
5	6	6	170	220	270
6	6	6	200	250	300
7	6	8	150	200	250
8	6	8	170	220	270
9	6	8	200	250	300
10	6	10	150	200	250
11	6	10	170	220	270

12	6	10	200	250	300

Figure 2-12 illustrates the best OF values solutions for the SA and MF algorithm with respect to the number of iterations past until these values where found. **On average, MF is faster compared to SA by around 32%.**

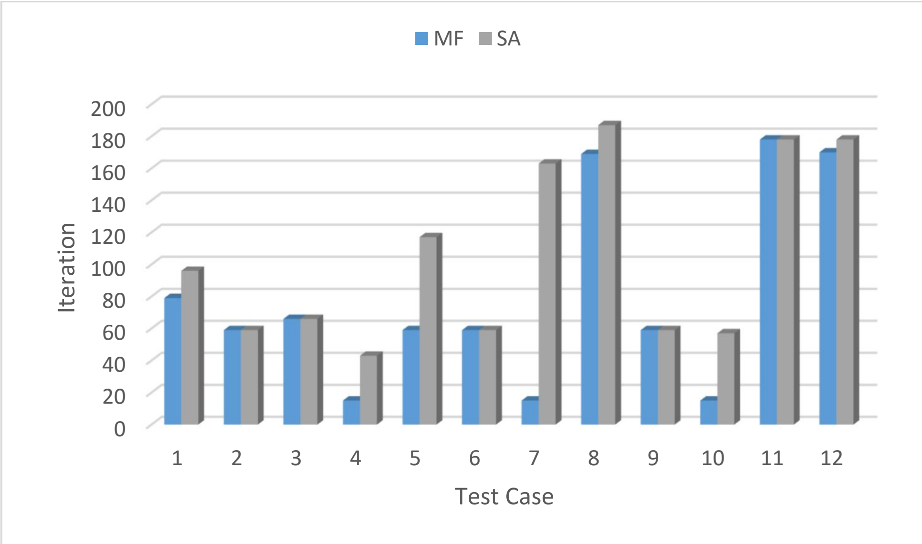


Figure 2-12: MF and SA for Best Objective Function value (average results).

Figure 2-13 *presents the best OF values for every test case. As we can see these two algorithms are computing almost the same results for these various cases. Cases number 1, 4, 7 and 10 have the highest best values but all the others values are quite at the same level. We can understand that for transmission areas that range from 170 to 200 meters for the UEs and from 200 to 250 meters for the MAPs (via* Table 2-2*) the signal quality between these entities is at its best.*

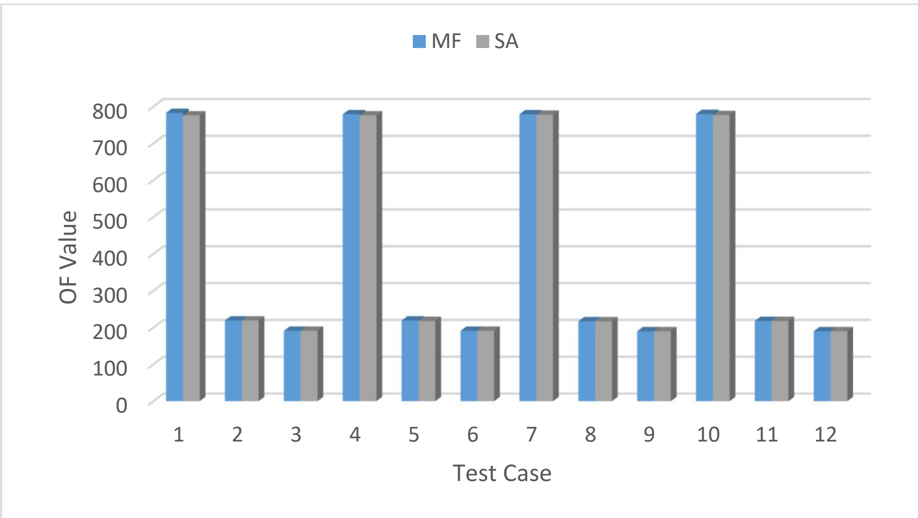


Figure 2-13: MF and SA Best values for Objective Function for all test cases (best results).

The bar chart form Figure 2-14 also provide us with some insights of the computation of the MF and SA algorithms. SA algorithm produce only one path with each execution and tries to optimize the network connection (better signal quality) by choosing the best nodes as already explained in 2.6.1 . On the other hand MF algorithm computes the maximum value of paths based on the cap values (explained in 2.6.2) and for this reason it can produce up to 3 paths in average for some test cases. Thus MF algorithm can provide a more distributed result in order to solve the problem via dividing the communication from one UE to one or more APs through multiple paths created by various MAPs nodes. Also by increasing the transmission area of the entities for only 20 meters can result to a higher number of paths computed from the MF algorithm. For example test case 7 is about 1.7 paths (average number) and the next case is at almost 2.25 paths and goes to a little lower than 3 at test case 9.

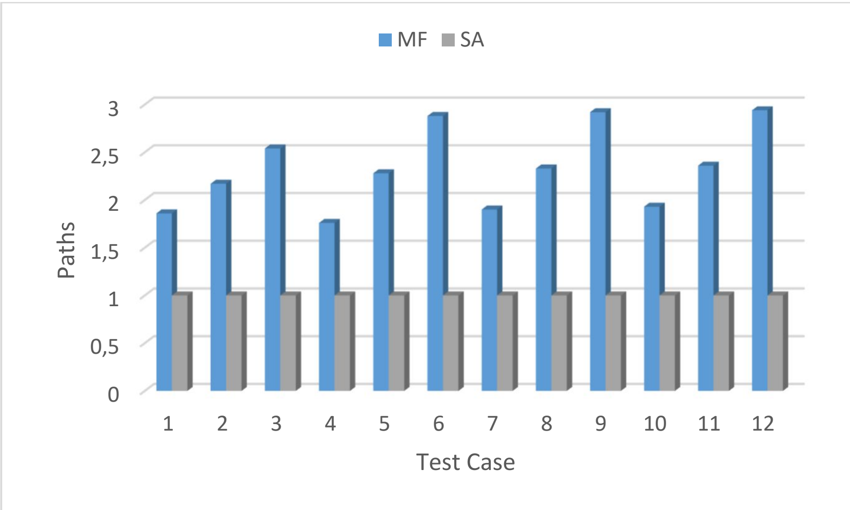


Figure 2-14: Number of Paths/Flows for MF and SA computations.

Figure 2-15 *clearly shows that the SA algorithms is better that MF algorithms when it comes to overall objective function values even if* Figure 2-13 *illustrate the same results for the best OF values of these two algorithms. Also the chart compares the OF values for the different changes of the transmission ranges, e.g. the case 5th is only the half value of the 4th case and again the 6th case is almost the half from the 4th case. Overall it highlights that there is no significant change in the average OF value even if the density of the network increase every 3 test cases.*

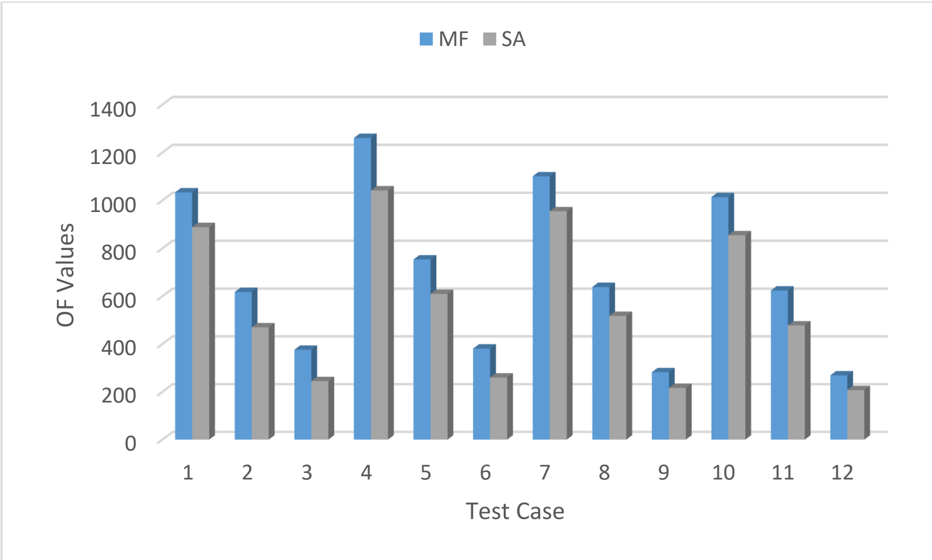


Figure 2-15: Average OF Values for MF and SA computations on each test case.

Figure 2-16, Figure 2-17 and Figure 2-18 are screenshots from the SUMO simulator that show the results from the MF and SA algorithms. The green color indicate the path created from the SA algorithm, red for the MF algorithm and with blue are the nodes that participate to both the paths produced from the algorithms. These different paths/topologies created by MF and SA generate the differences that we see in all the previous figures.

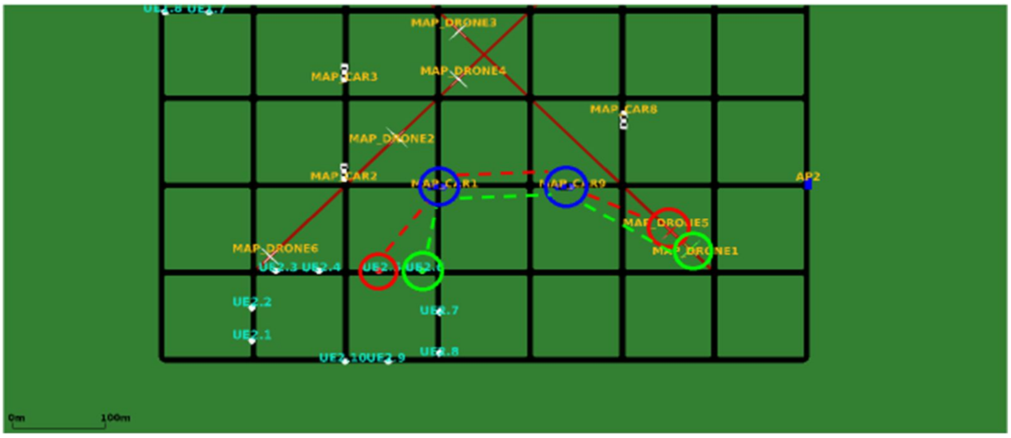


Figure 2-16: Screenshot of SUMO simulator for MF & SA algorithms.

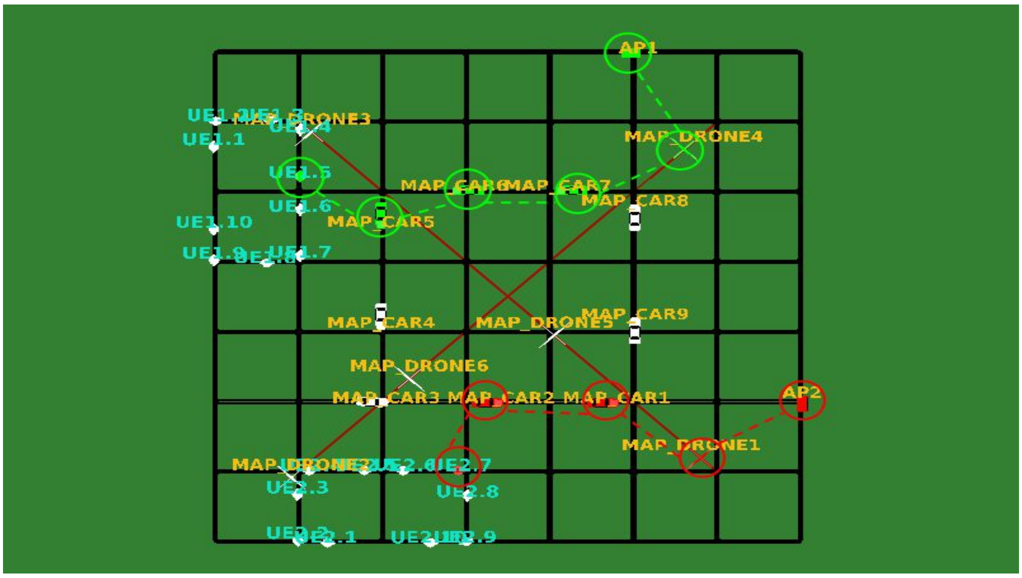


Figure 2-17: Screenshot of SUMO simulator for MF & SA algorithms.

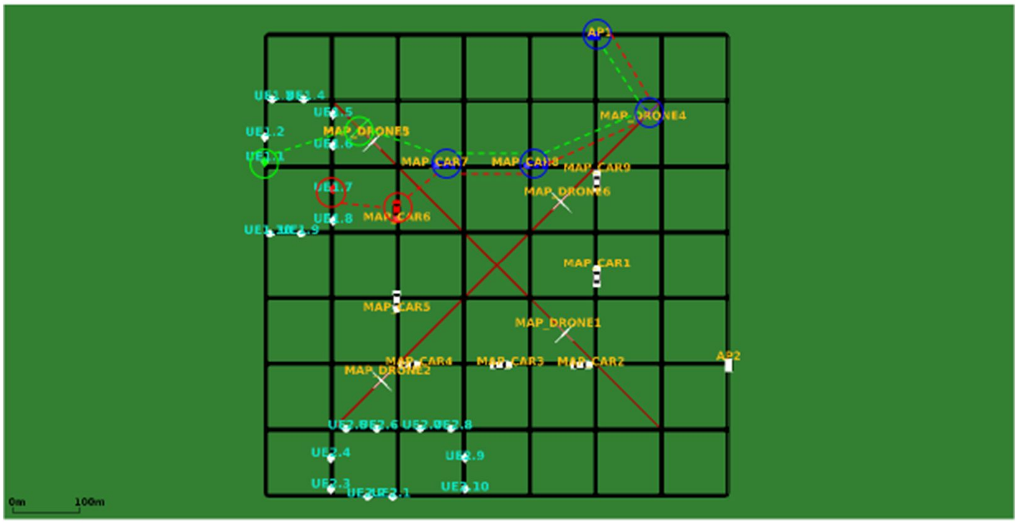


Figure 2-18: Screenshot of SUMO simulator for MF & SA algorithms.

2.9 Conclusions

The work considered radio access points which are capable of moving and establishing a radio network connection with other entities (MAPs, APs, and UEs) for efficiently serving users in contexts under stringent conditions. Specifically, the research solved the complex problem of finding the best candidates from the MAPs in the vicinity based on their position, by calculating the movement and communication costs, in order to provide connectivity to users in areas that have no infrastructure. The complex optimization problem was mathematically formulated and solved through two algorithms. The Simulated Annealing meta-heuristic and the Maximum Flow Algorithm. Results from testing the algorithms into different cases with different transmission areas and MAP densities showcased that the solutions provided by the Maximum Flow algorithm were around 32% better (in terms of finding the best Objective

Function Value faster) than the Simulated Annealing algorithm. Also the MF algorithm produce more than one paths in average while SA only one at the time.

2.10 References

[1] Z. Xu and S. Sukkarieh, “Decentralised control of robot teams with discrete and continuous decision variables”, 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 4780-4785, Shanghai, China, May 2011.

[2] Rezaee, H.; Abdollahi, F., "A Decentralized Cooperative Control Scheme With Obstacle Avoidance for a Team of Mobile Robots," in Industrial Electronics, IEEE Transactions on , vol.61, no.1, pp.347-354,Jan.2014

[3] C. Sommer, R. German, and F. Dressler, “Bidirectionally coupled network and road traffic simulation for improved IVC analysis,” IEEE Transactions on Mobile Computing, vol. 10, no. 1, 2011.

[4] A.Varga, “The OMNET++ discrete event simulation system”, European simulation multiconference (ESM 2001).

[5] M.Behrisch, L.Bieker, J.Erdmann, D.Krajzewicz “SUMO – Simulation of Urban Mobility”, SIMUL 2011 : The Third International Conference on Advances in System Simulation

[6] D. Karvounas, K. Tsagkaris and P. Demestichas, “Nature-inspired optimization of Moving Access Point-based radio networks”, Annals of Telecommunications, vol. 67, no. 9-10, pp. 423-436, 2015

[7] “Strategies for Creating Networks of Moving Cognitive Access Points” (MCAPs), project under “Naval International Cooperative Opportunities in S&T Program” (NICOP) of the “Office of Naval Research” (ONR) of the “Naval Research Enterprise” (NRE).

[8] R.W. Eglese, “Simulated Annealing: A tool for Operational Research”, European Journal of Operational Research 46, pp. 271-281, 1990

[9] S. Kirkpatrick, C. Gelatt and M. Vecchi, “Optimization by Simulated Annealing”, Science, New Series, vol. 220, no. 4598, pp. 671-680, 1983

[10] Rutenbar, R.A., "Simulated annealing algorithms: an overview," Circuits and Devices Magazine, IEEE , vol.5, no.1, pp.19,26, Jan. 1989

[11] Fleischer, M., "Simulated annealing: past, present, and future," Simulation Conference Proceedings, 1995. Winter , vol., no., pp.155,161, 3-6 Dec 1995

[12] Guoqiang Deng; Min Tang; Guangxi Chen, "An improved algorithm for maximum flow problem," Communications, Circuits and Systems, 2009. ICCCAS 2009. International Conference on , vol., no., pp.591,594, 23-25 July 2009

[13] Mahlous, A.R.; Fretwell, R.J.; Chaourar, B., "MFMP: Max Flow Multipath Routing Algorithm," Computer Modeling and Simulation, 2008. EMS '08. Second UKSIM European Symposium on , vol., no., pp.482,487, 8-10 Sept. 2008

[14] Ford, L.-R., Fulkerson, D.-R., “Maximal flow through a network”, Canadian Journal of Mathematics, no. 8, pp. 399–404, 1956

3 Modelling and analysis of management in heterogeneous infrastructure

The exponential growth of mobile traffic drastically increase network complexity, management, computing and many more functions at the infrastructure level. Different traffic models, data transmission schemes, spectrum, protocols and even hardware needs should be analyzed in order to design future infrastructures that meet all requirements. This chapter aims to analyze the requirements, present the challenges and provide solutions to problems that may appear in the near future.

3.1 Introduction

As the number of smartphones and demand for higher data-rate connections keep explosively growing, the technology has to pursue this trend in order to be able to provide the suitable communication schemes. It is calculated that the total global mobile wireless user devices will be over 10 billion, with the mobile data traffic growing more than 200 times over compared to 2010 numbers. The new 5G communication systems are promising a complete network structure with unlimited access to information, providing the requested service demands to users far beyond what 4G offers by supporting innovative new wireless technologies and network architecture to meet the extremely high performance requirement. These features include support for new types and massive number of devices, for very high mobile traffic volumes, universal access for users, very high frequency reuse and spectrum reuse in wireless technologies, automated provisioning, configuration and management of a wide range of new network services, ultra-reliable, ultra-low latency, ultra-densification and even more. 5G networks would be a heterogeneous networks (HetNets), meaning that different networks will be integrated all together to a unified system, enabling aggregation of multiple existing radio access technologies (RATs) such as LTE-A, WIFI, D2D and even lightly-licenced. Delivering all 5G requirements in order to support the new features and services, a substantial change on the network architecture is inevitable. Normally in the past such a process would need of deployment of specialized devices build for a specific application and with fixed functionalities. Thus any development and transformation to follow the constantly increasing and heterogeneous market requirements demands a huge investment to change/deploy hardware. Nowadays, various technologies and architectures have been utilized in order to solve this problem and provide a faster introduction and adaptation of new technologies to the communications systems. One of these elements that offers reprogrammability of the network elements in order to solve new problems or to establish new more suitable functions, is Software Defined Networking (SDN). This architecture provides dynamic, manageable, cost-effective and adaptable, making it ideal for the high-bandwidth, dynamic nature of today's applications. SDN decouples the dependence of implementing instructions are provided by multiple, vendor-specific devices and protocols, making it bounded to an exact hardware. Networking, computing and storage resources would be integrated into one programmable, unified and flexible infrastructure, that will be more cost effective and with higher scalability at minimum cost. This unification will allow for an optimized and more dynamic utilization of various distributed resources, and the pooling of fixed, mobile and broadcast services. This could be realized through the introduction of radio resource management to the system, in favor of managing all the available radio access technologies (RATs). In that aspect, 5G is designed to be a viable, robust and scalable technology. Another positive result with the introduction of 5G communication systems will be the drastic energy consumption reduction and energy harvesting that will help the industry to have an astounding usage growth. Since network services will rely progressively on software, the creation and growth will be further encouraged. In addition, the 5G infrastructures will provide network solutions and involve vertical markets such as automotive, energy, food and agriculture, city and buildings management, government, healthcare, manufacturing, public transportation.

The rest of the chapter is structured as follows: section 2 elaborates on the main 5G requirements, while section 3 presents the status & challenges in hardware and software development. Section 4 elaborates on the status & challenges in 5G wireless communications by focusing on physical layer, MAC and RRM. Finally Section 5 investigates the benefits of machine learning in 5G network management.

3.2 5G Requirements

In the digital era, users and devices are becoming more dependent on various applications and services that involve the creation, access/ communication, processing, and storage of digital content. These developments have been tremendously accelerated by wireless/mobile technologies, which have offered unparalleled access/ communication opportunities to users. 5G is expected to be dominated mainly by the following application classes including massive Machine Type Communications (mMTC), enhanced Mobile Broadband (eMBB), ultra-reliable and low latency communications (URLLC). These main classes will facilitate scenarios related to critical and demanding applications for the realization of smart cities, as well as the realization of applications for Industry 4.0 and automation aspects. Also, self-driving aspects are expected to pose strict requirements especially on the latency in order to ensure reliable and secure service with very high rate of availability. Figure 3-1 illustrates the 5G main application areas as mentioned before.

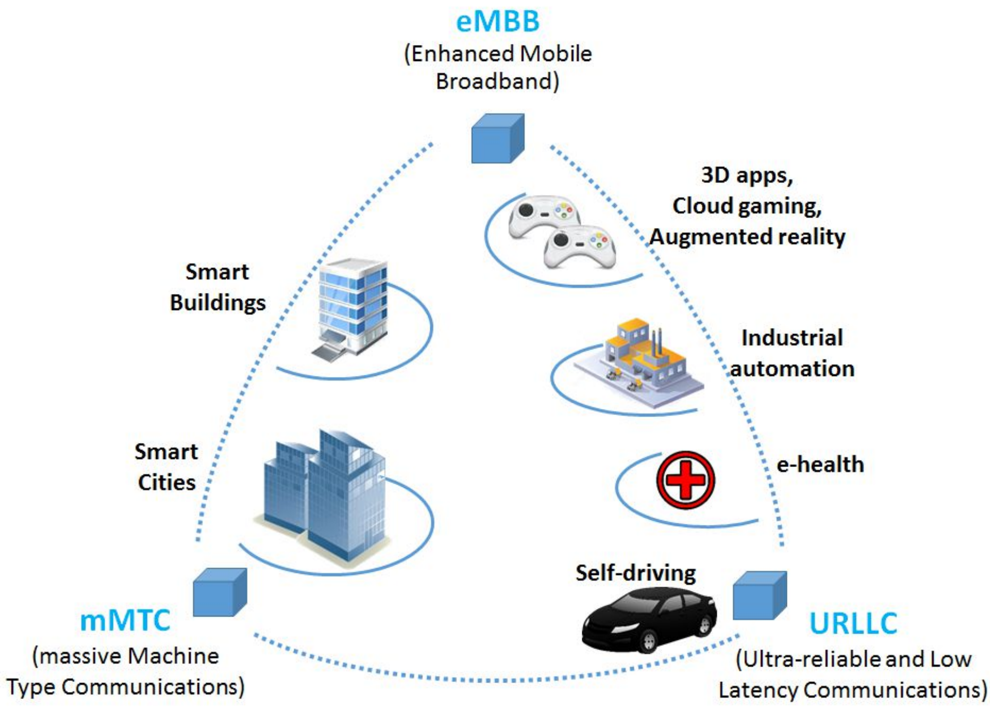


Figure 3-1: 5G main application areas.

For the successful realization of the aforementioned application areas, there are certain technological and networking trends which can lead towards the 5G direction. In terms of technology trends there are:

- Network slicing
    - Dynamic network slicing in 5G enables the design, deployment, customization, and optimization of different network slices running on a common network infrastructure. It leverages innovations in cloud mobile access and core [1].
- Cloud and Fog/Mobile Edge Computing
    - Cloud aspects can involve the allocation of resources to physical components in content servers etc. activation of the appropriate volume/type of functional components and the determination of the interconnections and links between the physical elements. In this respect, they will manage in an aggregate manner all types of resources, namely, communication, computing and storage, by taking advantage of significant processing powers/storage facilities associated with cloud platforms and virtualization through abstractions of resources and service components (which are "pooled" and universally accessible/sharable), in an on-demand, elastic and scalable manner.
- Separation of Control and User plane
    - Separation on control and user plane can potentially lead to more efficient usage of resources and energy efficiency as well. For example, the Greentouch initiative proposed that user plane (data) can be served mainly by small cells while a limited number of macro cells can serve as a signaling umbrella in order to handle control plane aspects.
- Virtualization of networking functions
    - Virtualization may include solutions based on software-defined networking (SDN) and network function virtualization principles (NFV) may be solutions for accelerating the application/service deployment times and flexibility (in general) in the network. For instance, through the standardized interfaces of the SDN model there can be instructions on how to handle new applications/services. Likewise, through NFV there can be an easier implementation of applications/services and networking intelligence (activation in cloud, and instructions towards forwarding-elements). The overall challenge is to evaluate the potential of these concepts, in terms of impact in the application/service deployment times, QoS/QoE.

In terms of wireless network trends Figure 3-2 there are:

- Massive MIMO and utilization of mmWave
    - An important direction is related to the usage of massive Multiple Input - Multiple Output (MIMO) which can significantly enhance the spectral and energy efficiency of the wireless network. Moreover, the extra capacity that is needed by 5G networks for facilitating massive IoT, mobile broadband communications etc. can be provided through the utilization of extra spectrum by exploiting bands above 6GHz. MmWave frequencies can be used for outdoor point-to-point backhaul links or for supporting indoor high-speed wireless applications (e.g., high-resolution multimedia streaming) [2]. Moreover, as the millimeter waves have a short wavelength, it becomes possible to pack a large number of antenna elements into a small area, which consequently helps realize massive MIMO at both the base stations and user devices [2].
- Novel Multiple Access Schemes
    - Novel multiple access schemes such as non-orthogonal multiple access (NOMA) is one of the techniques being considered which uses cancellation techniques in order to remove the more powerful signal. Of course well-known techniques such as orthogonal frequency division multiple access (OFDMA) can be exploited as well.
- Ultra-dense infrastructures

- Another direction that will continue to progress is the constant decrease in the cell size, at the expense of a corresponding constant increase in the number of cells that will be deployed. Cells of different sizes, characterized as macrocells, microcells, picocells or femtocells will continue to be deployed. Specifically, hundreds of small cells are deployed per macro cell. The challenge with ultra-dense networks is to deploy and operate the appropriate set of cells, so as to carry data traffic, without severely increasing the signaling traffic (increases with the number of cells), by minimizing the impact of mobility and radio conditions, and by achieving cost and energy efficiency.
- New waveform and advanced coding
    - New, candidate waveforms for 5G would be needed in order to serve specific service requirements, for example higher or lower bandwidths, sporadic traffic, ultra-low latency, higher data rates compared to legacy technologies. Also robustness towards distortion effects such as interference, RF impairments, etc. is important to be supported by new waveforms.

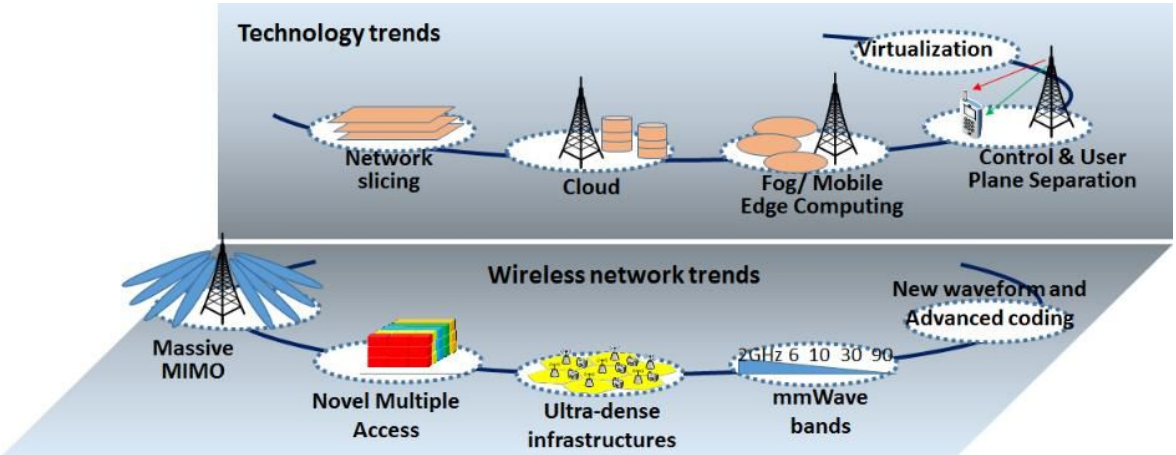


Figure 3-2: 5G Technology and wireless network trends.

3.3 Status & challenges in hardware and software development

The increasing network demands require new network services, higher performance, increased bandwidth, lower energy consumption and increased resilience. These demands imply a higher number of network devices and stations, increasing the cost and the energy consumption. The network centralization and functions virtualization become more significant as they enable better distribution of the available resources, less hardware utilization and an easier to upgrade network as current devices and architectures are meeting their limits. New implementation techniques must be introduced in order to further increase reusability, flexibility along with performance and energy consumption at the same time. According to this approach, the network functions can be moved to software as much as possible, without affecting networks latency. The functions virtualization can be achieved at any level, using a partitioning technique between software and hardware functions, which takes into account the available resources. The current network systems introduce a static and customized functions virtualization. This manual and static partitioning may lead to high performance but it is not reusable and not reconfigurable thus limiting network upgrades and resource allocation. Also, the processing power cannot be shared among nodes offering limited efficiency and spectrum capacity. Full virtualization is not always available as the devices of the underlying network might not be able of such a task, or virtualization might be limited according to available physical and computational resources. A cognitive and dynamic HW/SW partitioning can provide reconfigurable and flexible HW/SW partitioning to both device and network element architectures in 5G technologies, considering high performance and energy consumption reduction, according to the specified performance scenarios. The HW/SW partitioning is applicable for either inside a network stack layer and/or between multiple network stack layers, as shown in Figure 3-3.

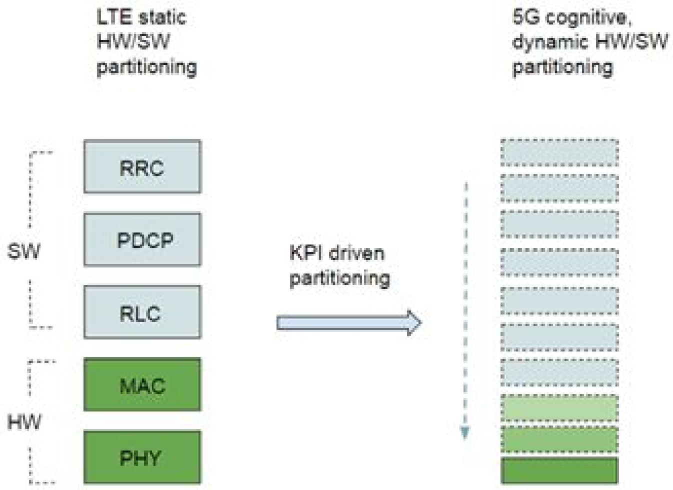


Figure 3-3: Cognitive, dynamic HW/SW partitioning in network stack layers.

3.3.1 Problem statement

The cognitive, dynamic partitioning takes into account a set of given network functions, the KPIs that have to be optimized and the KPI constraints relative to the available resources. The technique result provides the HW or SW implementation decision for each given function, according to the given policies. These policies consist of the KPIs and their constraints. The result of the implementation has to change according to the policies alteration. The cognitive dynamic HW/SW partitioning 's task is to provide the best HW/SW partitioning of the 5G network stack functions, considering the given KPI's per scenario and the available HW/SW resources. The partitioning algorithm's result can be parsed to the management programs for further decision making on the implementation part. The partitioning solution has to communicate with other programs to be aware of the available resources. Moreover the partitioning must be aware of the performance scenario and the KPI constraints; thus the partitioning would interact with management programs and monitoring agents as seen in Figure 3-4.

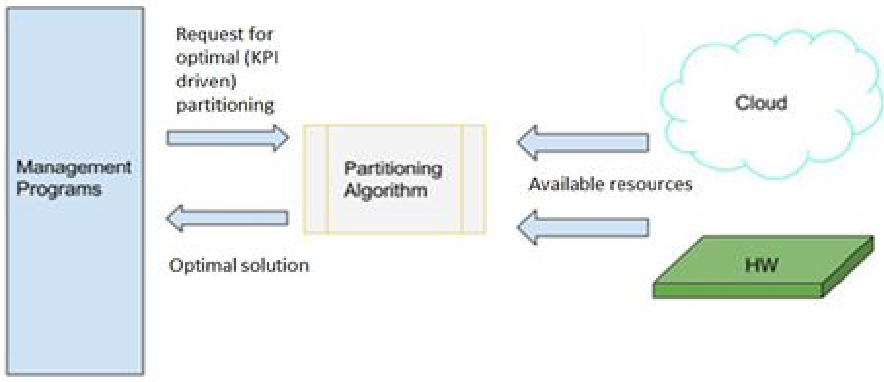


Figure 3-4: Partitioning algorithm functionality and communication.

3.3.2 Solution

The cognitive dynamic HW/SW partitioning algorithm considers a multiple and diverse set of objectives, relative to 5G requirements. To this purpose the current solution contains an evolutionary multi-objective algorithm, aiming to solve a specified optimization problem. This implies that the optimization problem formulation is one of the most significant parts of the solution, as it has to accurately, address the requirements, the KPI's and the objectives of a 5G network. The multi-objective algorithm receives as input the functional graph of the defined functions, along with the system's and functions' KPI's/Constraints and optimization goals. Then the number of possible solutions is populated according to the optimization problem formulation. After having the set of solutions, a multi-objective algorithm will search for the best solution, according to the given KPI's, constraints and optimization goals.

3.3.3 Functions definition (LTE, 3GPP based PHY functions)

The cognitive dynamic HW/SW partitioning algorithm is able to decide on device and network element functions, either inside a network stack layer and/or between multiple layers. Most of the current research and implementations are using functions from the LTE MAC layer or its reconfiguration while trying to introduce softwarization and reconfiguration of the LTE PHY layer functions. Some implementations are also able to simulate full LTE networks. The most notable among them are the LENA ns-3 [3] LTE simulator from CTTC which provides full software implementation of virtual LTE networks, the OpenAirInterface [4], OpenLTE [5] and srsLTE [6]. In order to provide a first realization of the HW/SW partitioning challenges, the selected subset of functions used for the evaluation of the HW/SW partitioning is derived from the OpenLTE physical layer software implementation in Octave source code. The functions that compose the Octave code are identified, manually implemented in Verilog code and characterized according to the KPI's. Specifically, the functions used for the evaluation of the algorithm up to now, are the following:

- eNodeB LTE PHY [7] layer, downlink, OpenLTE, Octave (**SW functions**). The Octave source code that describes an LTE based frequency domain downlink transmitter implementation in SW. This code contains the functions that are specified from the existing LTE, 3GPP standards
- eNodeB LTE PHY layer function, broadcast channel:
    - Cyclic redundancy check (CRC), Verilog code (**HW function**)
    - Convolutional encoding, Verilog code (**HW function**)
    - Rate matching, Verilog code (**HW function**)
- eNodeB LTE PHY layer function, physical downlink control channel

    - Cyclic redundancy check (CRC), Verilog code (**HW function**)
    - Convolutional encoding, Verilog code (**HW function**)
    - Rate matching, Verilog code (**HW function**)
- eNodeB LTE PHY layer function, physical downlink control channel
    - pseudo-random sequence generation, Verilog code (**HW function**)

The seven latter functions where manually implemented in Verilog code and where evaluated in order to be bit accurate in conjunction to the corresponding SW functions. These functions are the first to be implemented in HW as they introduce a significant amount of delay in the SW execution time. Furthermore while the implementation of broadcast channel and physical control channel, in Verilog, is similar their HW mapping results to different power consumption, that makes them perfect candidates to exercise the capabilities of the multi-objective algorithm.

3.3.4 Parameters (KPIs)/ Constraints definition

The partitioning algorithmic solution considers the most critical KPIs for the 5G networks regarding partitioning and virtualization, so far. The considered KPIs are described in the following lines.

- Execution time
    - The measured execution time of the LTE network functions when implemented in SW
    - The measured execution time of the LTE network functions when implemented in HW
- Energy consumption of Verilog modules derived from power analysis using appropriate FPGA IDE
- Measured SW memory utilization (e.g. RAM), of LTE functions when implemented in SW
- Communication time, considering measured time of data transferring, regarding send and receive communication functions, between HW and SW implemented, LTE functions
- Reusability referring to the available HW resources, inversely relative to HW functions utilization

3.3.5 Functional graph (Dataflow Graph) provision

The communication scheme between the utilized LTE functions forms a dataflow graph, including nodes which represent the functions implemented in HW or SW and edges which represent the link/interconnection between two functions. The communication overhead is also applied as weight to each corresponding link. The GUI accompanying the algorithmic solution provides the user with the ability to add or remove nodes/functions and edges. The partitioning algorithm receives this information and forms an array structure that includes the mapping of the interconnected components. The array's items provide the multi-objective algorithm with the ability to find the optimal decision that also considers the communication overhead. An example is provided in the following Figure 5-1.

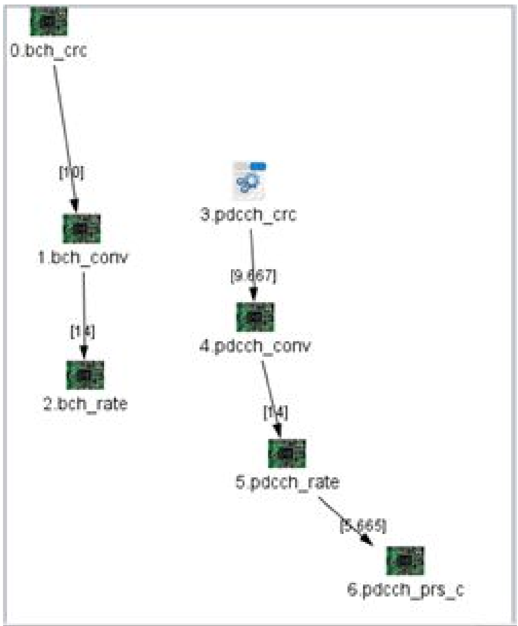


Figure 3-5: Data flow graph example.

The Figure 5-1 is generated by the user interface that controls the partitioning algorithm. The solution includes a predefined set of functions and interconnections, referring to the already implemented functions that form the graph in the Figure 5-1. The user interface provide the user with the ability to add or remove functions, interconnections and their KPIs in order to create a graph that better suits his/her needs.

3.3.6 Optimization Problem formulation

The core process of the optimization procedure is the optimization problem formulation as it guides the whole optimization process towards the goals of the problem. This process receives as input the implemented HW/SW functions along with their KPIs and formulates a Multi-Objective Optimization Problem that decides on HW or SW implementation of the specific functions so as to optimize the objectives set by the end user. The following text summarizes the optimization problem formulation:

KPI-driven binary optimization:

- N is the number of considered functions
- M is the number of considered KPIs
- x=[x1,...xN] is the decision variables vector, with xn, n = 1...N, representing HW implementation (xn=1) or SW implementation (xn=0) of function n
- f(x) = [f1(x) ,...fM(x)] is the vector of objective functions to be minimized, each objective function fm, m = 1...M, representing a corresponding KPI optimization
- fm(x)= a1mx1+....aNmxN, where, anm is the difference of the KPI m value between HW implementation and SW implementation of function n
- Each objective function fm, m = 1...M, can be normalized or prioritized by weights according to operator policies about the importance of KPI m
- Each objective function fm, m = 1...M, can have a corresponding constraint, e.g. if m is the full model execution time-latency and it should be less than 80μs, this mean fm<=80

The objective problem formulation can be applied to multi-objective optimization algorithms (heuristic, Pareto-optimal, progressive), which search for the optimal set of solutions using the results of the objective functions. The next section provides a description of the multi-objective algorithmic solution that is utilized towards this direction.

3.3.7 Evolutionary Multi-objective Algorithmic solution

This section provides a description of the multi-objective algorithmic solution that is utilizing the optimization problem formulation. The multi-objective algorithm derives a subset of solutions from the set of all possible solutions according to the provided, multiple objective functions, their KPIs, their constraints and the optimization goal. The initial set of all possible solutions is provided from the objective functions that are considered in the optimization problem formulation. The functions and KPIs/Constraints are provided by the user, utilizing the solution's user interface. The selected multi-objective algorithm for the HW/SW partitioning, is the non-dominated sorting genetic algorithm NSGA II [8]. The algorithm begins its search with an **initial population** of individuals-solutions usually created at random within a specified lower and upper bound on each variable. Once the population in initialized the population is sorted based on non-domination into each front. Once the **non-dominated sort** is complete the **crowding distance** (the Euclidian distance between each individual's solution) is assigned. After the population members are evaluated, the **selection** operator chooses better solutions with a larger probability to fill an intermediate mating pool. For this purpose the tournament selection procedure takes place in which two solutions can be picked at random from the evaluated population and the better of the two can be picked. The **crossover** operator is to pick two or more solutions (parents) randomly from the mating pool and create one or more solutions by exchanging information among the parent solutions. Each child solution, created by the crossover operator, is then mutated with a **mutation** probability so that on an average one variable gets mutated per solution. In the context of real-parameter optimization, a simple Gaussian probability distribution with a predefined variance can be used with its mean at the child variable value. This operator allows an EO to search locally around a solution and is independent on the location of other solutions in the population. The **elitism operator** combines the old population with the newly created population and chooses to keep better solutions from the combined population. Such an operation makes sure that an algorithm has a monotonically non-degrading performance. Finally, the user of an EO needs to choose **termination criteria**. Often, a predetermined number of generations is used as a termination criterion. In most cases, this algorithm is able to find much better spread of solutions and better convergence near the true Pareto-optimal front compared to other multi-objective algorithms, on a diverse set of difficult test problems.

The partitioning algorithmic solution provides cognition in terms of optimal decision based on information about functions KPIs, derived from management programs and monitoring agents, dynamicity referring to the ability to dynamically move functions from HW to SW implementation and vice versa according to the specified policies, considering also the communication overhead. Current results show improved overall performance (execution time, latency/communication overhead) by 70% and power consumption reduction by 50%. The following Figure 3-6 illustrates the improvements in overall execution time of the addressed and tested physical layer functions, when HW/SW partitioning is performed targeting high performance (left chart) and normal performance scenarios (right chart). In a high performance scenario more functions will be executed in HW while in normal performance more functions will move to SW in order to reduce power consumption.

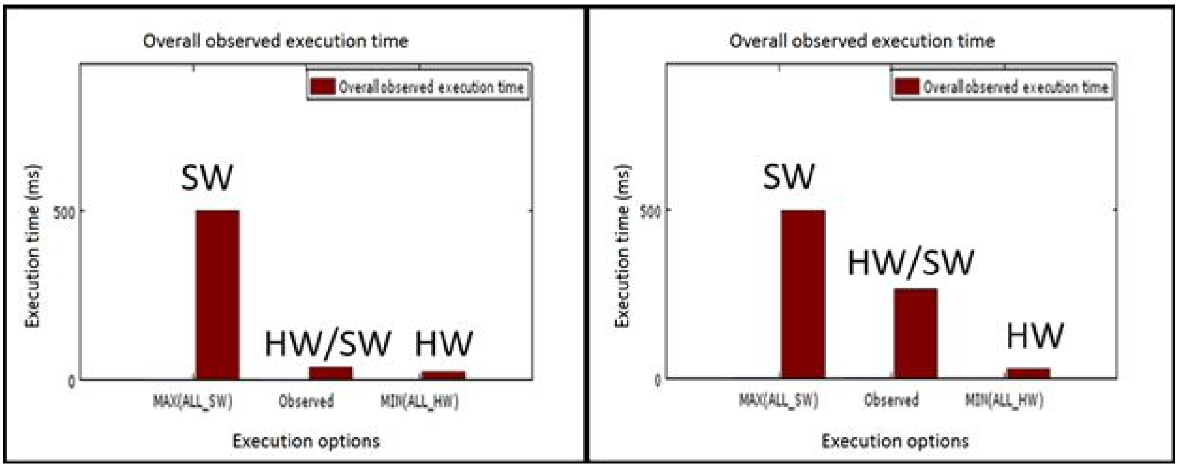


Figure 3-6: i) SW implementation, ii) HW/SW optimal partitioning, iii) HW implementation results in execution time (ms) for high performance and normal performance scenarios using Octave.

The first set of results imply that the partitioning algorithm can achieve high performance regarding algorithmic solution execution time, below thresholds for LTE current handover time and user experience maintenance and algorithmic solution complexity. In addition, the cognitive, dynamic HW/SW partitioning algorithm provides reduction of the power consumption, limited memory overhead enabling the interaction between algorithmic solution with management programs and monitoring agents. The proposed HW/SW partitioning will be further refined and upgraded considering improvements on performance, memory overhead and portability of the current algorithmic solution. Furthermore, there is ongoing investigation for integration with the CTTC's LENA ns-3 simulator in order to have a full stack LTE network implementation in order to better clarify the benefits and challenges of HW/SW partitioning. Moreover, since the algorithmic solution's design enables KPIs' and functions' extensions future implementations and KPIs will be investigated in order to meet a diverse set of objectives such as user data rate and even capacity.

3.4 Status & challenges in 5G wireless communications

This section provides useful insights on the advancements in physical layer, MAC and RRM. Details on the aforementioned advancements are provided in the subsections that follow.

3.4.1 Novel physical layer aspects

One of the key aspects that will evolve with 5G has to do with the novel design of a unified, flexible air interface, its components and procedures, so as to effectively deal with the issue of 5G requirements through such an adaptation. Therefore, it is important to develop a new spectrum agnostic 5G air interface for carrier frequencies below 6 GHz. This is motivated by the fact that today's licensed bands for cellular usage are all below 6 GHz, and the World Radio Conference (WRC) in 2015 which also focused on below 6 GHz spectrum among other aspects. Furthermore, even if higher frequency spectrum bands are made available for 5G operation in the future, having effective means for utilizing 5G below 6 GHz is still of relevance due to the more favorable radio propagation properties. A unified, flexible 5G air interface would have the following key characteristics:

- Flexibility to support the broad class of services with their associated KPIs;
- Scalability to support the high number of devices;

- Versatility to support the diverse device types and traffic/transmission characteristics;
- Efficiency to support the requirements on energy consumption and resource utilization; and
- Future-proofness to support easy integration of new features.

The new air interface will meet the requirements on the 5G main KPIs (e.g., for increased throughput, reduced latency etc.) with increased flexibility, reliability, future-proofness as well as cost and energy efficiency. Notably, devices will be designed to support one common air interface for all services and more devices will be produced applying similar/common chip sets thus achieving economy of scale for vendors. Also, the wireless system is expected to be more scalable and thus better suited to follow load variations between the services both temporal and in different locations.

3.4.2 Novel frame design based on service requirements

As previously mentioned, among the main objectives of a new, unified 5G air interface should be its flexibility to be adapted to the diverse requirements imposed by the nowadays heterogeneous service demands. This diversity of requirements creates a very challenging environment in terms of service specific KPIs and channel characteristics as it should be flexible enough to satisfy these needs while in parallel optimize the resource utilization and minimize the overhead introduced by the multi-service support functionalities/mechanisms. The 5G services are foreseen to include mobile broadband (MBB) services, supporting high data rates and high coverage, massive machine communication (MMC) services, supporting small packet sizes and infrequent transmissions, mission critical communication (MCC) services, with strict delay bounds and reliability factors, and vehicular-to-anything communication (V2X) services, supporting both bs-to-device and device-to-device transmissions.

The current status of a stiff frame structure with e.g. a fixed transmission time interval (TTI) value, either cannot satisfy the extremely strict requirements of specific services (e.g., delay requirements of MCC services) or is resulting to underutilization and waste of resources due to inefficient resource management. Therefore, a flexible frame structure supporting the coexistence of different transmission time intervals (TTIs) is more than necessary in order to accomplish these diverse service requirements. The introduction of a flexible TTI supporting different TTI durations will both accomplish ultra-low latency capabilities (e.g., in case of MCC services) by facilitating short TTI durations and high spectral efficiency gains (e.g. in case of MBB services) by utilizing long TTI durations. The selection of TTI scaling, which is the set of available TTI values, is of high importance, because it directly affects the effectiveness of the proposed flexible frame structure solution.

We proposed two methods for TTI scaling:

- 2^N scaling: Definition of a set of TTI durations with double duration each one
- Scaling based on service classification: Definition of a minimum set of TTI durations (mapped to a set of services)

In case of 2^N scaling, the minimum and maximum TTI length is defined and then a set of TTI lengths is generated based on the 2^N approach. For the 5G services mentioned above, reasonable values of TTI duration can be between 0.125ms and 4ms, therefore the generated TTI values belong to the set (0.125ms, 0.25ms, 0.5ms, 1ms, 2ms and 4ms). This approach is graphically depicted in Figure 3-7.

According to the second method, a set of predefined services (e.g. MBB, MMC, MCC, V2X) are analyzed based on their requirements in order for each service to estimate the TTI value that best reflects its characteristics, satisfy its requirements in terms of KPI values (e.g. average delay, max delay, throughput) and minimize the system resource overhead. Then a set of TTI lengths are generated which are mapped to the aforementioned services or group of services. In Figure 3-7 an indicative set of TTI lengths is depicted based on an initial analysis and service classification. Each TTI length is mapped to one or several services.

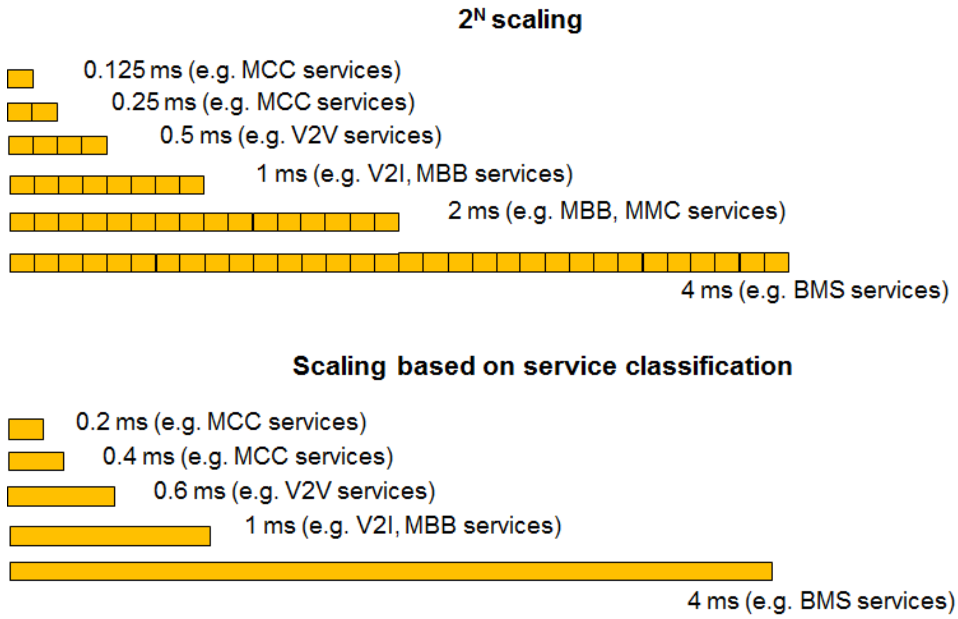


Figure 3-7: Flexible TTI - TTI Scaling.

Regarding the comparison between the two scaling methods, the second method is better fitted to the service requirements as it is based on analysis of the special characteristics and optimum TTI selection for each service type. However, in case of absent of resource partitions in the spectrum, because of the different multiply factors between TTI values, this method cannot succeed high multiplexing gain between different services in the frequency domain as resource gaps may emerge for a set of selected TTI values. The first method may not be optimal fitted to the services, although it eliminates the gaps of the spectrum (in case of no recourse partitions), while it occupies the minimum length in the packet header (for 8 different TTI values, 3 bits are required).

3.4.3 Support of different numerologies

The following parameters are made available for configuration in order to create different numerologies: TTI length, subcarrier spacing, number of subcarriers and tiling type (horizontal or vertical related to time/frequency axes). Table 3-1 illustrates the set of numerologies used during the preliminary system level evaluation.

Table 3-1: Numerologies used in simulation scenarios.

Numerologies			
RB Type	TTI (ms)	Subcarrier Spacing (kHz)	Number of Symbols per TTI

RB1 (LTE like)	1	15	14
RB2	0.5	15	7
RB3	0.25	30	7

A certain simulation set can be used by a proprietary system level simulation tool. The set evaluates a selected one-stage access protocol [9] in combination with a different numerology in order to estimate the combined gains from the use of both technical components. The results are evaluated against the LTE-A environment (ARP protocol and LTE-A frame structure). The general parameters of simulations are based on the parameters defined mainly in 3GPP case 1 reference scenario. The use case specific parameters for the set are presented in Table 3-2.

Table 3-2: Simulation scenario parameters.

Parameters	Value
Reference Scenario	3GPP case 1
Network Topology	19 3-sectorised base stations (57 cells)
Inter-site distance	500 m
Bandwidth	FDD 40MHz - 20MHz (uplink)
Request generation	Poisson
UE data traffic size	100 bytes / data report
PRACH allocation	1 PRACH allocation / TTI
Backoff window	Discrete uniform distribution (2,20) TTIs
Max connection attempts	4
Preambles	64

In the aforementioned simulation set, the performance of the combination of one-stage protocol with numerologies with smaller TTI sizes are evaluated. Figure 3-8 depicts the performance in terms of latency (primary KPI) for four alternative configurations. The first configuration is the baseline LTE-A scenario which includes the ARP protocols and the LTE-A frame structure with TTI length of 1ms. In the other three configurations, the one-stage protocol is adopted in combinations with three different numerologies with TTI lengths of 1ms, 0.5ms and 0.25ms.

Figure 3-8 shows that the LTE-A (ARP, TTI=1ms, PRACH per 1ms) has latency values between 5ms and 7.5ms, which is a performance far for the KPI target of 1ms. The other configurations show improved latency values highly affected by the request rate. In detail, the numerology of TTI=1ms have relatively low latency values only for low request volumes, while the numerologies of TTI=0.5ms and TTI=0.25ms have relatively low latency values for all the examined request volumes. Latency values below the KPI target are observed only for the numerologies of TTI=0.5ms and TTI=0.25 ms and for low request rates.

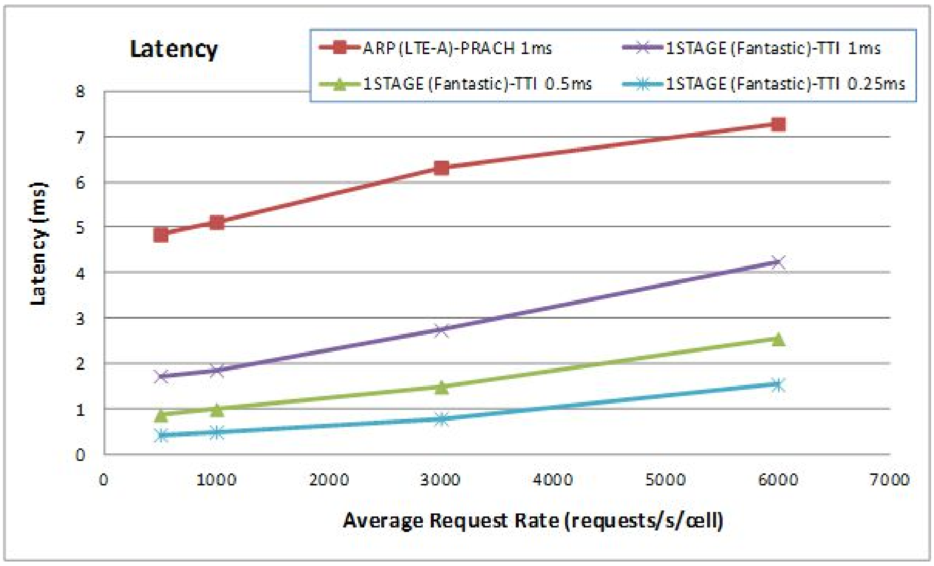


Figure 3-8: KPI 2 Latency (primary KPI) – second set of simulation results.

The aforementioned latency results depict the latency from the UE to the BS (uplink), while no processing delay was taken into consideration.

3.4.4 Enhanced Radio Resource Management (RRM) and MAC adaptation for 5G

In 5G networks, is important to proceed to the investigation and development of technologies that address the well-known challenges of predicted growth in mobile connections and traffic volume by successfully addressing the lack of dynamic control across wireless network resources which is leading to unbalanced spectrum loads and a perceived capacity bottleneck. Resource management with three degrees of freedom can be taken into consideration: (i) densification, (ii) rationalized traffic allocation over heterogeneous wireless technologies, and (iii) better load balancing across available spectrum bands in licensed, lightly-licensed and unlicensed spectrum portions. Moreover, the MAC has to be adapted in order to be able to support the deployment of 5G various services which call for increased reliability, reduced latency and higher throughput in licensed, lightly-licensed and unlicensed bands.

Traffic steering provides operators with the necessary functionality in order to let them optimize resource utilization, QoS/QoE and power consumption of cells and UEs by directing the traffic to the RAT or layer that is the most appropriate/suitable for a certain type of service. Steering of certain traffic flows depending on their type and availability of resources will enable devices to obtain guidance on how to optimally access content in indoor and outdoor environments. Various factors could be taken into account including signal strength, interference levels, availability of RATs and channels, requirements of certain services (e.g., mission critical, massive access etc.).

The Figure 3-9 shows the actions which are considered for dynamic steering by taking into account certain prioritization of traffic flows (for the selection of RATs), load balancing (for the selection of cells) and channel assignment (for the selection of most appropriate/ suitable channels). Specifically, as depicted in the Figure 3-9, reception of low quality from UEs trigger the initialization of traffic steering. RAT and channel with high load and low quality is identified in order to seek for solutions.

Furthermore, a prioritization algorithm receives reports on QoS and QoE from all UEs. In the case of triggering a low QoE/QoS then the RRM mechanism must begin some procedures to satisfy the UE's requirements. Taking into account the requested conditions and demands from each UE can then steer a UE to its more appropriate RAT. Also, a load balancing algorithm, focuses explicitly on achieving a good load balance between cells of the same RAT. Thus, based on "sensing" mechanism the system is monitoring the loads and collect the measurements from all cells of a particular RAT, an overloaded cell is identified and then under loaded cells that are in the vicinity are identified. An active and eligible UE can then be moved from the overloaded cell to the under loaded adjacent cell that meets its requirements in order to gradually arrive at the preferred load balance. In addition to the network load, the user experience (QoE) is monitored also throughout the "sensing" mechanism, and if needed an inter-RAT handover procedure is invoked to move the UE to a different RAT. Finally at the time of connection establishment, the decision-mechanism with the utilization of these algorithms and mechanisms, is pursued to select the right RAT, cell and channel to be used by the user throughout the connection.

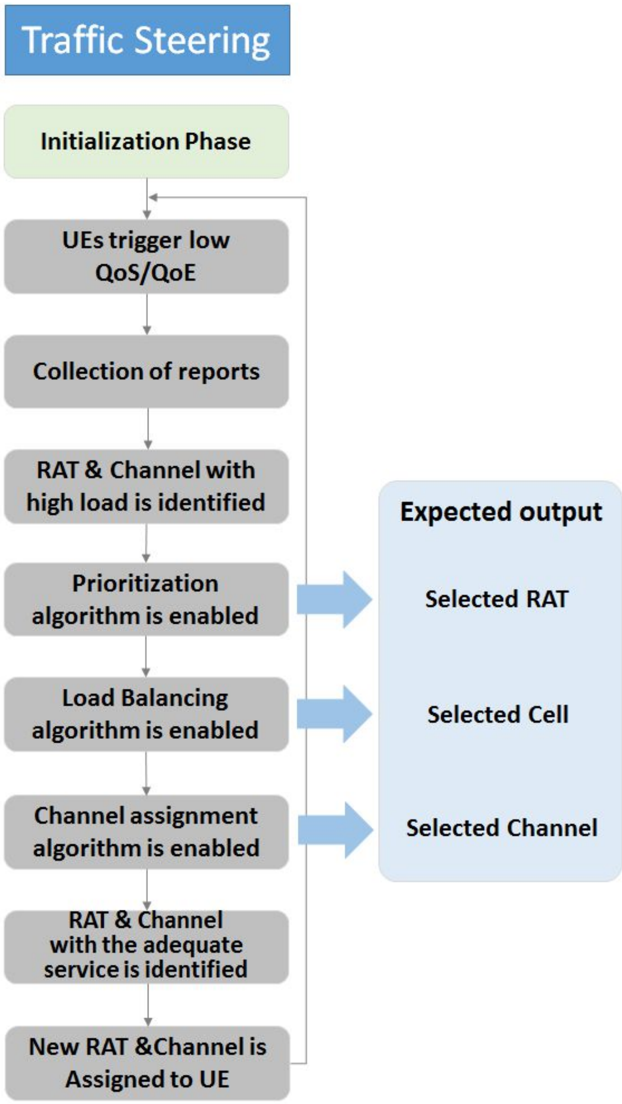


Figure 3-9: Traffic steering.

The Figure 3-10 illustrates the integral parts of the Dynamic Channel Assignment (DCA); Prioritization; Load balancing which have been previously discussed. Such concepts can be evaluated through system level simulations. Therefore, it is critically important to develop a system level simulator in a way to make a flexible, accurate and efficient simulation of this complex heterogeneous networks with multiple RAT, cells and channels for 5G networks. Specifically, it should be possible to support heterogeneous networks consisting of LTE macro cells (licensed bands), small cells that utilize lightly-licensed bands and Wi-Fi with unlicensed bands, with BSs and APs densely deployed to provide users with seamless connectivity and demanded services. And finally the radio resource management mechanism that assigns the most appropriate RAT, Cell and Channel based on the requirements of each UE and at the same time keeping the system as Load balanced as much as allowed. The particular simulator is a proprietary system-level simulation tool which is fully developed in Java with various capabilities and has been calibrated according to the 3GPP specifications. It takes into account various parameters such as traffic level, available infrastructure elements, available channels and evaluates the various test cases.

The calibration state of the proprietary simulator has been evaluated against the reference results of the 3GPP LTE calibration campaign [10]. As a result, the Cumulative Distribution Function (CDF) of coupling loss and downlink SINR have been examined in order to calibrate the tool with leading operators and vendors such as Nokia, Ericsson, DoCoMo, Huawei, Telecom Italia etc.

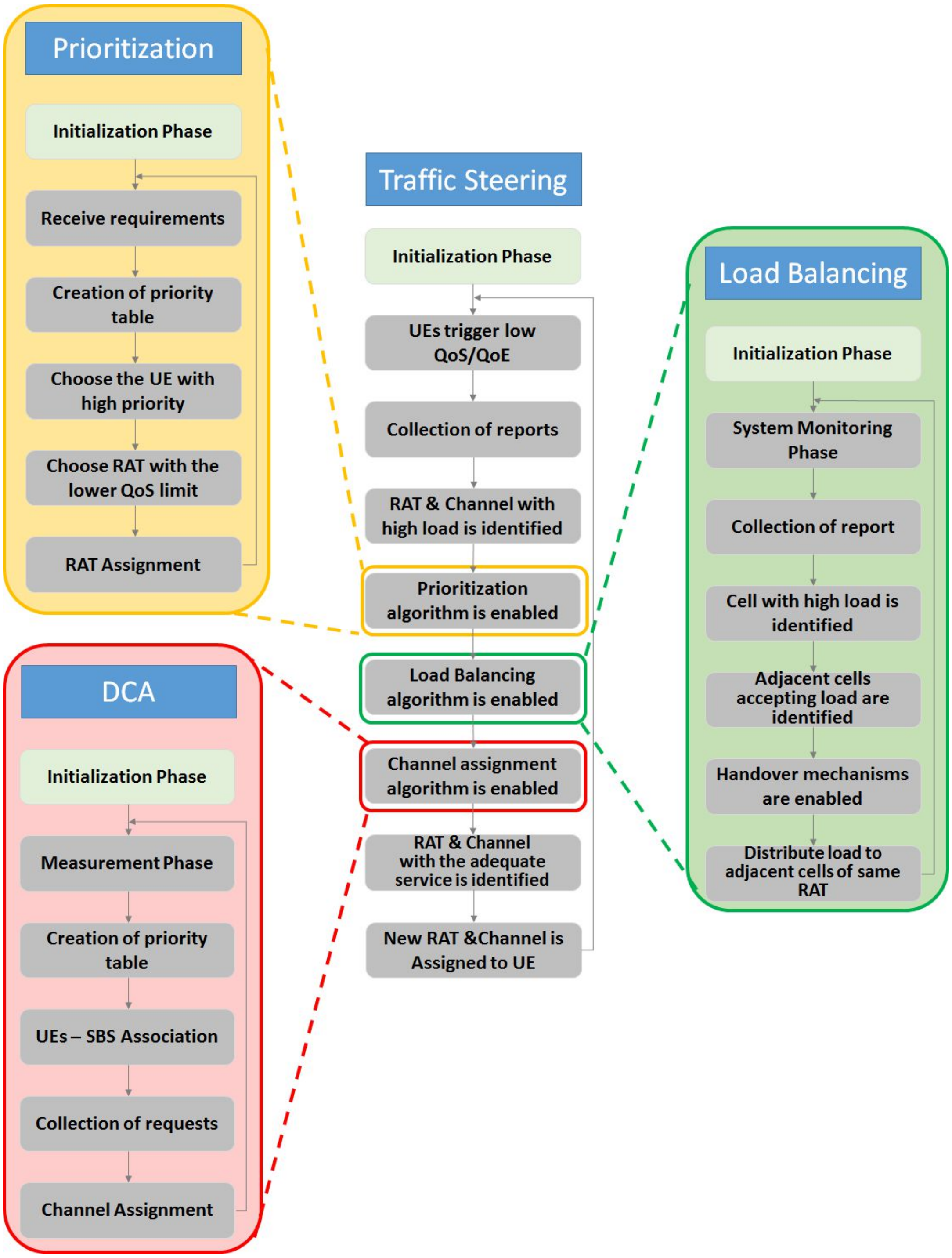


Figure 3-10: Dynamic Channel Assignment (DCA); Prioritization; Load balancing as integral parts of traffic steering.

The simulator's playground is illustrated in Figure 3-11, where multiple macrocells and small cells were deployed in order to simulate a cellular environment as an example. The simulator is capable of creating various scenarios for sparse to dense and even ultra-dense deployment of various cells and technologies.

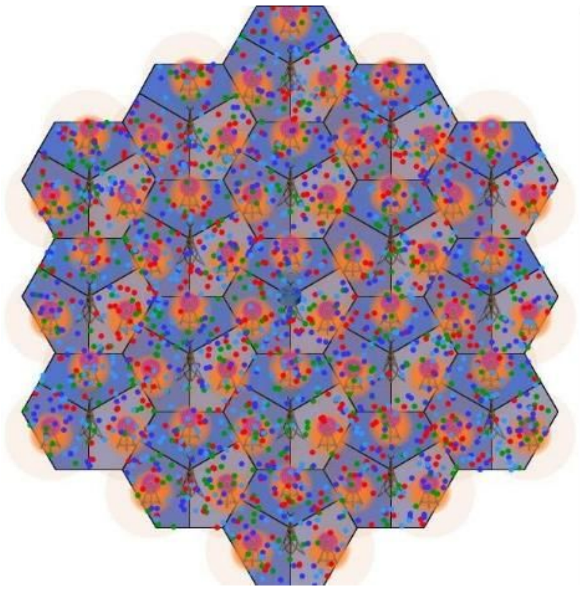


Figure 3-11: Example of system level simulator playground.

3.5 5G network management aspects enhanced with machine learning

3.5.1 Machine learning for service classification in 5G networks

A challenge for future wireless communication networks is the satisfaction of the diverse requirements coming from heterogeneous services. In 5G networks, the coexistence of different services like Mobile Broadband (MBB), Massive Machine type Communications (MMC) and Mission Critical Communications (MCC) having various requirements in terms both of capacity and QoS will constitute a key prerequisite. Hence, one of the main issues that should be addressed by the 5G management system is the simultaneous provisioning of these services satisfying the corresponding requirements so as to optimize the network in order to be resource and energy efficient. A first step towards this direction is to be able to identify each service type in order to prioritize the services and be able to allocate efficiently the network resources.

Knowledge of QoS requirements per service flow could be provided by the higher layers as e.g. assumed in the HSPA and LTE, where sets of QoS parameters are available for RRM functionalities such as admission control and packet scheduling decisions. As an example from LTE, each data flow (bearer) is associated with a QoS profile consisting of the following downlink related parameters:

- Allocation retention priority (ARP)
- Guaranteed bit rate (GBR)
- QoS class identifier (QCI)

In particular, the QCI includes parameters like the layer-2 packet delay budget and packet loss rate. However, for the cases where detailed QoS parameters are not made available from the higher layers, the use of novel service classification techniques should be considered, in which the base stations monitor the traffic flows to extract more detailed service classification information and identify the service type providing this information input to packet scheduling algorithms, and other RRM functionalities.

The support of fast and reliable traffic characterization is a necessary step in order to understand the network resource usage and to provide differentiated and high QoS/QoE through prioritization targeting in increasing resource-usage and energy efficiency. In addition, the service classification process can interact with new services and procedures provided in 5G networks to support flexibility and adaptability to traffic variability.

In this subsection, the use of various machine learning mechanisms for the service classification problem is described and the performance of different algorithms is investigated. The considered classification methods reside in the area of statistical-based classification techniques and they are realized by exploiting several flow-level measurements (e.g. such as traffic volume, packet length, inter-packet arrival time and so forth) to characterize the traffic of different services. Then, to perform the actual classification, supervised machine learning techniques are applied to these measurements. It should be noted that in contrary to other methods of traffic classification, like payload-based classification, which need to analyze the packet payload or need to use deep packet inspection technologies, statistical-based classification techniques are usually very lightweight, as they do not access packet payload and can also leverage information from flow-level monitors.

3.5.2 State of the art of machine learning mechanisms for traffic classification

In the literature, there are a lot of studies that focus on application and service discrimination based on traffic classification learning techniques as presented in detail [11] and [12]. Various machine learning mechanisms are usually employed belonging to either unsupervised or supervised machine learning as illustrated in Figure 3-12. In the first case, clustering algorithms like K-Means, DBSCAN and Autoclass [13] are investigated. The objective of these mechanisms is to group flows that have similar patterns into a set of disjoint clusters. The major advantage of these schemes is that they do not require a training phase like the supervised ones but they automatically discover the classes via the identification of specific patterns in the dataset. However, the resulting clusters do not certainly map 1:1 to services as usually the number of clusters is greater than the number of service types and even in the case of 1:1: mapping, the clusters still need to be labeled in order to be mapped to the corresponding services.

Regarding the supervised machine learning techniques, which is also the approach that is analyzed in this subsection, there are various classification schemes that have been proposed for the traffic classification problem like Naïve Bayes, Decision trees, Random forests and others [14]. Authors in [15] present the Bayesian classification techniques which use the Naive Bayes approach. During the training phase, flow parameters are used to train the classifier and create a group of services. Then, when new flows arrive, they are subjected to probabilistic class assignment, by calculating their probabilities of class membership and assigned to that class to which maximum probability is attained. In addition, statistical fingerprint-based classification techniques as presented in [16] classify traffic based on a set of pre-selected parameters (e.g. packet size, inter-arrival time). During the training phase, a dataset of flows from each service are used in order to analyze the dataset and create the service fingerprint. This fingerprint is usually a PDF (Probabilistic Density Function) vector used to identify the service. During the classification phase the algorithm checks the behavior of a flow against the available set of PDF vectors. Also, Support Vector Machine (SVM) techniques, first proposed in [17], are binary supervised classification algorithms which transform a nonlinear classification problem in a linear one, by means of what is called a “kernel trick”. Furthermore, Artificial Neural Networks (NNs) consist of a collection of processing elements that are highly interconnected and transform a set of inputs to a set of desired outputs that is inspired by the way biological nervous systems works. In [18] authors proposed a NN, in which a multilayer perception classification network is used for assigning probabilities to flows. A set of flow features are used as input to the first layer of network, while the output classify flow into a set of traffic classes by calculating the probability density function of class membership. Also, Decision Tree algorithms, which are mentioned also by [19], represent a completely orthogonal approach to the classification problem, using a tree structure to map the observation input to a classification outcome. In these supervised classification algorithms, the data set is learnt and modeled, therefore, whenever a new data item is given for classification, it will be classified accordingly learned from the previous dataset.

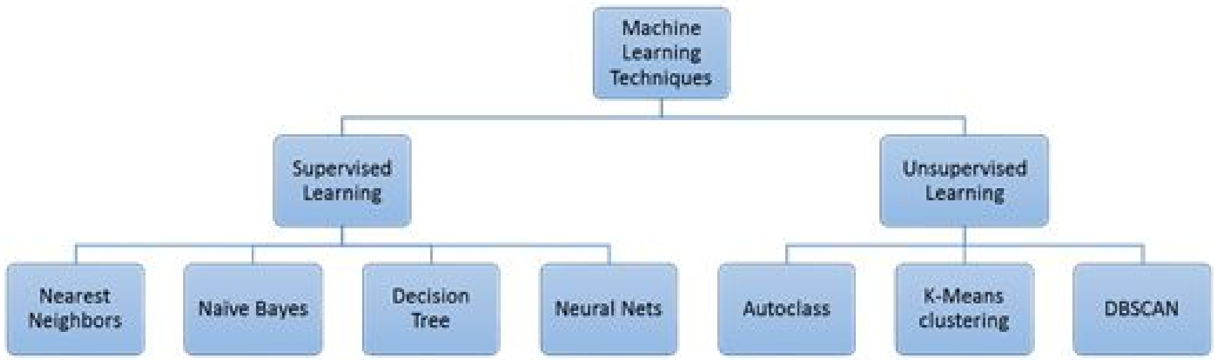


Figure 3-12: Overview of machine learning techniques.

3.5.3 Classification approach & evaluation metrics

In this subsection, the problem of service classification is investigated employing a set of different classification mechanisms that belong to the supervised machine learning category. Before presenting the performance evaluation of each mechanism, the algorithmic procedure that has been followed is described and each step is explained in detail.

Figure 3-13 presents the algorithmic procedure that is followed for the classification mechanism. As can be seen, the first step refers to the collection of a number of traces from different services. For the considered simulation scenario, three service types are considered referring to MCC, MMC and MBB communication while other services (like broadcast/multicast services etc) will also be considered in the future. The different traces of each service have been generated using specific traffic models. More specifically, the generation of different types of MCC /MMC traffic was following the traffic models presented in 802.16p [20], while for the generation of MBB traffic video streaming traffic (YouTube) which follows the traffic models presented in [21] is assumed.

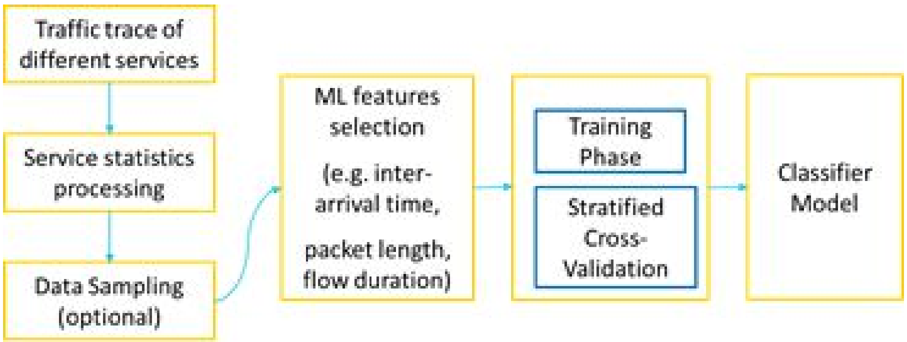


Figure 3-13: Proposed mechanism for the service classification process.

The second step refers to the statistical processing of these traces in order to separate them in flows. In particular, a flow is considered a series of packets transmissions that have the same source and destination and for which the inter-arrival time is below a specific threshold. After this processing, a number of features for each flow is generated including: inter-arrival time statistics (mean value, standard deviation), packet size statistics (total value, mean/max/min value, standard deviation) and other flow characteristics like total number of packets, source, destination as well as flow direction. Subsequently, some features engineering tasks are performed before proceeding to the main classification mechanism. These tasks include the selection of the most representative features, the transformation of categorical features into numerical values, the normalization of features' values and other tasks that guarantee a high data quality (e.g. replace missing values etc.). Then, the implementation of machine learning mechanism follows including two main phases: the training phase and the cross-validation phase. It should be noted that stratification is applied in this case in order to randomly sample the flows' dataset in such a way that each service type is properly represented in both training and testing datasets.

For the simulation scenario, a splitting of 70%-30% for training and testing sets has been considered. Obviously, for the training set, the label 'service type' of each flow is considered as known whereas for the testing set, this label is considered as unknown and each flow is labeled using the classifier model. The outcome of the proposed mechanism is a classifier model that can be employed in unknown flows in order to recognize them and label them in an accurate way. To evaluate the performance of the classification mechanisms, various metrics have been defined and can be used in the train/test sets to select the most adequate mechanism for the specific problem. To illustrate the relationship between the different evaluation metrics, a very useful tool that provides a holistic view of each algorithm's performance is the confusion matrix. The confusion matrix is actually a two dimension matrix, in which the horizontal axis represents the predicted class (outcome of the algorithm) whereas the vertical axis represents the true class. In Figure 3-14, the confusion matrix for the considered classification problem is presented, where FP, TP, FN, TN stand for False Positives, True Positives, False Negatives and True Negatives, correspondingly and they are defined as follows:

- FP: the percentage of other services' flows that are incorrectly classified as MMC service
- TP: the percentage of MMC service's flows that are correctly classified as MMC service
- FN: the percentage of MMC service's flows that are incorrectly classified as other services
- TN: the percentage of other services' flows that are correctly classified as other services.

Classification / Result / Service	MMC service	Other Services
MMC Service	TP	FN
Other Services	FP	TN

Figure 3-14: Confusion Matrix of the service classification problem.

Some of the most common evaluation metrics used for classification problems are the accuracy metric, the precision, the recall and the F1-score. More specifically:

- Accuracy is defined as the percentage of correct predictions to the total number of predictions and is given by $\frac{(TP+TN)}{(TP+FP+TN+FN)}$
- Precision is defined as the percentage of the instances that were correctly predicted as belonging in a class among all the instances that were classified as belonging in this class and is given by $\frac{TP}{(TP+FP)}$
- Recall is defined as the percentage of the instances of a specific class that were correctly classified as belonging to this class and is given by $\frac{TP}{(TP+FN)}$
- F1 Score is defined as the harmonic mean of the precision and recall and is given by $\frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$.
- To be able to choose the best mechanism for a classification problem, the investigation of a single metric, like accuracy, is not always enough as the misclassification of a specific class instances may be more important than the correct classification of others. For this reason, other evaluation metrics have also to be applied to make the most appropriate choice depending on the problem's characteristics.

3.5.4 Evaluation performance of classification mechanisms

In the considered simulation scenario, the performance of a set of different machine learning mechanisms has been investigated including base classifiers such as Naïve Bayes classifier, Support Vector Machines, Tree Classifier, K Nearest Neighbor Classifier, Logistic Regression as well as ensemble based classifiers like Random Forest Classifier. The goal of ensemble methods is to combine the predictions of several base estimators built with a given learning algorithm in order to improve generalizability and robustness over a single classifier. Usually, two families of ensemble methods are distinguished: the averaging methods (e.g. Random Forests [22]) in which several classifiers are developed independently and then the average of their predictions is used and the boosting methods (e.g. AdaBoost- Adaptive Boosting) where base classifiers are built sequentially and one tries to reduce the bias of the combined estimator. To be able to compare the various machine learning mechanisms, in the following table, the accuracy metric of each algorithm is presented, where a Dump classifier that classifies all the flows as type 0 (MMC service) is also considered resulting in 0.512 accuracy. From this table, it can be seen that Decision tree and the Random Forest algorithms lead to the highest accuracy values, outperforming the other machine learning algorithms.

Table 3-3: Accuracy score for each classification mechanism.

Classification Mechanism	Accuracy
Naive Bayes	0.808
Support Vector Machine	0.662
Decision Tree	0.976
K Nearest Neighbor Classifier	0.952
Logistic Regression	0.685
Random Forest Classifier	0.988

However, to provide a more complete view of each classifier's performance, the corresponding confusion matrices are illustrated in Figure 3-15. The horizontal axis of this matrix represents the predicted class whereas the vertical axis represents the true class. It should be noted that Class 0, Class 1 and Class 2 refer to MMC, MCC and MBB service types, correspondingly. In the considered scenario, considering that it is desired to eliminate the possibility that a MCC service is misclassified as another service type, the optimal model should have high values of recall whereas high accuracy values for the case of MMC and MBB services are required. The results of confusion matrix show that the Decision Tree and the Random Forest algorithms result in extremely good results as they misclassify only a few flows, resulting also in high values of recall and precision as can be seen in Figure 3-16. Therefore, these two classification mechanisms can be selected chosen for further consideration for the problem of service classification.

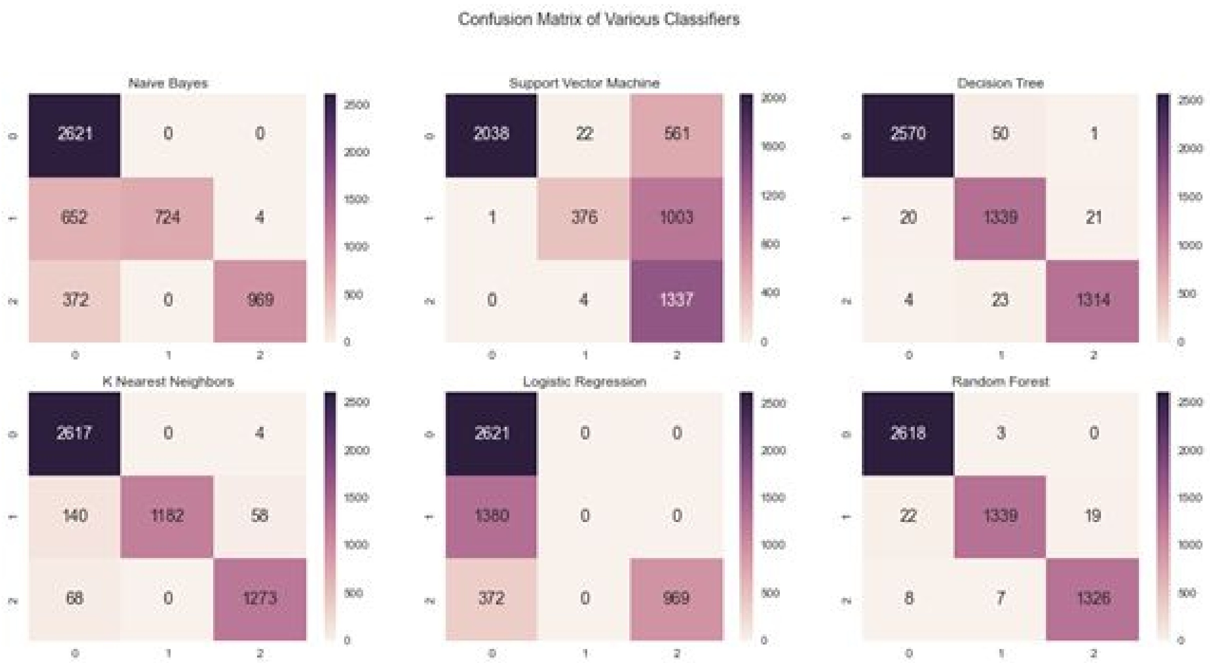


Figure 3-15: Confusion matrices of different classifiers.

Class\ Metrics	Precision	Recall	F1-score	Class\ Metrics	Precision	Recall	F1-score
MMC	0.99	1.00	0.99	MMC	0.99	0.97	0.98
MCC	0.99	0.97	0.98	MCC	0.94	0.97	0.96
MBB	0.99	0.99	0.99	MBB	0.98	0.99	0.99
Avg/total	0.99	0.99	0.99	Avg/total	0.98	0.98	0.98

Figure 3-16: Evaluation metrics for selected classification mechanisms.

3.6 Conclusion

5G is the next frontier of innovation for entire mobile industry. Consequently the three major objectives for 5G, are support of massive capacity and massive connectivity; support for an increasingly diverse set of services, applications and users; and in addition flexible and efficient use of all available non-contiguous spectrum for widely different network deployment scenarios. Framed in this context, this chapter elaborated on the status & challenges in hardware/ software development and in 5G wireless communications by focusing on physical layer, MAC and RRM. Also the benefits of machine learning in 5G network management were discussed. By taking into account the diversity of infrastructure, radio resources and services that will be available in 5G, an adaptive network solution framework will become a necessity. Breakthrough developments in several RAN technologies will be required for realizing novel, 5G solutions. Such technologies include among others multiple access and advanced waveform technologies combined with coding and modulation algorithms, massive access protocols, massive MIMO and virtualized and cloud-based radio access infrastructure.

3.7 References

[1] J. Elliott, S. Sharma, “Dynamic End-to-end network slicing unlocks 5G possibilities”, available online: https://insight.nokia.com/dynamic-end-end-network-slicing-unlocks-5g-possibilities, accessed Nov. 2016

[2] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G Wireless HetNet: Potential Benefits and Challenges," in IEEE Vehicular Technology Magazine, vol. 11, no. 1, pp. 64-75, March 2016

[3] http://networks.cttc.es/mobile-networks/software-tools/lena/

[4] http://www.openairinterface.org/

[5] https://sourceforge.net/projects/openlte/

[6] https://github.com/srsLTE/srsLTE

[7] http://www.etsi.org/deliver/etsi_ts/136200_136299/136212/10.01.00_60/ts_136212v100100p.pdf

[8] A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II Kalyanmoy Deb, Associate Member, IEEE, Amrit Pratap, Sameer Agarwal, and T. Meyarivan, IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 6, NO. 2, APRIL 2002

[9] S. Saur, A. Weber, and G. Schreiber, “Radio access protocols and preamble design for machine-type communications in 5G,” in Signals, Systems and Computers, 2015 49th Asilomar Conference, Nov 2015, pp. 3–7.

[10] 3GPP TR 36.814, “Further advancements for E-UTRA physical layer aspects”, March 2010.

[11] N. Namdev, S. Agrawal, and S. Silkari. "Recent Advancement in Machine Learning Based Internet Traffic Classification." Procedia Computer Science 60 (2015): 784-791.

[12] S. Valenti et al. “Reviewing traffic classification”, Data traffic monitoring and analysis, pp.123-147, Springer, 2013

[13] T. T.T Nguyen and G. Armitage. "A survey of techniques for internet traffic classification using machine learning." IEEE Communications Surveys & Tutorials 10.4 (2008): 56-76.

[14] J.S. Aafa et al. “A Survey on Network Traffic Classification Techniques”, IJERT, vol. 3, no. 3, Mar. 2014

[15] A. Moore and D. Zuev, “Internet traffic classification using Bayesian analysis techniques,” in ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS) 2005, Banff, Alberta, Canada, June 2005.

[16] M. Crotti, M. Dusi, F. Gringoli, and L. Salgarelli, “Traffic classification through simple statistical fingerprinting,” SIGCOMM Comput. Commun. Rev., vol. 37, no. 1, pp. 5–16, 2007

[17] X. Peng, L. Qiong, and L. Sen. "Internet traffic classification using support vector machine [j]." Journal of computer research and development 3 (2009): 010.

[18] T. Auld, A. W. Moore, and S. F. Gull, “Bayesian neural networks for Internet traffic classification,” IEEE Trans. Neural Networks, no. 1, pp. 223–239, January 2007.

[19] N. Williams, S. Zander, and G. Armitage. "A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification." ACM SIGCOMM Computer Communication Review 36.5 (2006): 5-16.

[20] IEEE 802.16p Machine to Machine (M2M) Evaluation Methodology Document (EMD)

[21] P. Ameigeiras et al. "Analysis and modelling of YouTube traffic." Transactions on Emerging Telecommunications Technologies 23.4 (2012): 360-377

[22] Biau, G., Devroye, L., & Lugosi, G. (2008). Consistency of random forests and other averaging classifiers. Journal of Machine Learning Research, 9(Sep), 2015-203

4 Management of shared resources in a multi-provider environment

In this chapter we present a hierarchical radio resource management scheme, which, by utilizing small and macro cells, allows cellular users to reuse and exploit potentially underutilized spectrum portions. In order to overcome the issue of limited resources that can be allocated to users, one can exploit the Spectrum Access System (SAS), a technology that enables better allocation of the available spectrum to users in below 6 GHz bands. Providers can utilize different spectrum with different license schemes with specific requirements. However, exploiting SAS also introduces certain challenges. A mechanism able to dynamically allocate network resources to different user categories in order to meet the satisfaction of the users while guaranteeing to meet the requirements of the category they belong to is presented. The proposed algorithm operates in the 3.5 GHz band taking into account the SAS constraints and requirements. Finally, through simulations, the effectiveness of this approach is shown.

4.1 Introduction

After the market launch of LTE-A in 2013, the cellular ecosystem started working on defining the next generation of cellular networks, the so called fifth generation (5G). The main focus of the 5G definition phase was on creating a common vision among the different stakeholders, on defining a set of use cases, scenarios and services expected to be launched around 2020. 3GPP started the specification work on what is known as 3GPP 5G phase 1, a first set of features of the forthcoming 5G system, targeting completion around mid of 2018. The newly identified services do stress the importance of reducing communication latency, increasing reliability, throughput, energy efficiency and other Key Performance Indicators (KPI). As a matter of fact, current 4G system techniques for managing radio resources are able to maximize the Quality of Service (QoS) of the served users but are not able to perform the resource allocation in user categorization or slicing environments [4]. This limitation is due to the fact that resource allocation in 4G systems is performed by associating a priority to the service requested by the user. For instance, different users may belong to different categories with different priorities into the same or different categories, and thus such users should be managed by considering the priority they have plus the priority of the category they belong. As a result, the current 4G approach is suboptimal and cannot fulfil the new demands of the forthcoming new services. In fact, the 5G system is expected to be able to serve a diverse range of services with different design requirements. Key enablers such as ultra-dense small cells with multiple antenna nodes are believed to be core system elements in meeting the chosen challenging requirements. The utilization of small cells will provide a more scattered connectivity, thus avoiding to route the entire data through a large, in terms of surface covered, centralized base station. As a matter of fact, 5G will only be possible by making available more spectrum, by enhancing the spectrum management adding intelligence and achieving a higher and better utilization factor, and finally by introducing real-time spectrum data management.

This subchapter provides an overview of the Spectrum Access System (SAS) Figure 4-1 technology in the 3.5 GHz band that will enable a more dynamic, secure and efficient allocation, management and sharing of spectrum resources. SAS will enhance the performance of wireless broadband networks, optimizing capacity through higher spectrum utilization, ultimately delivering a high quality of experience to the end users. As Federal Communications Commission (FCC) already provided the SAS system requirements [16], work has started in the wireless community towards turning the vision into a functioning reality. In this work, it is proposed an algorithm for satisfying the required quality of certain user categories while at the same time making it possible to increase the overall performance of the system, through better allocation of resources to other user categories.

SAS model (three layers).

- Incumbent Access
  - **Licensed**
- Priority Access Licenses (PAL)
  - **Lightly-Licensed**
- General Authorized Access (GAA)
  - **Unlicensed**

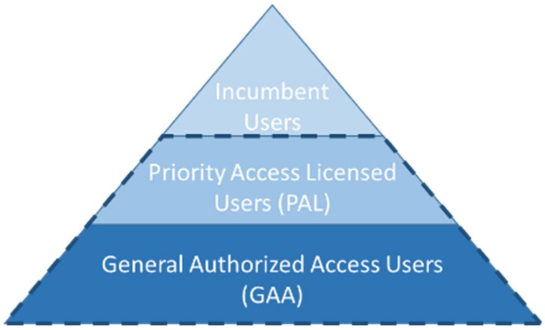


Figure 4-1: SAS three Tier Model.

This work is also proceeds to the development of a hierarchical (that is, blending distributed and centralized) management of ultra-dense multi-RAT and multiband networks. Centralized management is used as a baseline, as shown in Figure 4-2(a) which can be expanded with distributed management by moving management decisions related to RAT/spectrum/channel selection closer to the node level as depicted in Figure *4-2*(b). Specifically, the proposed algorithm will initially run in a distributed manner in order to limit the excessive signaling of centralized solutions in dense environments. However, in the case where the distributed approach does not provide satisfactory solutions, then a centralized approach could be used.

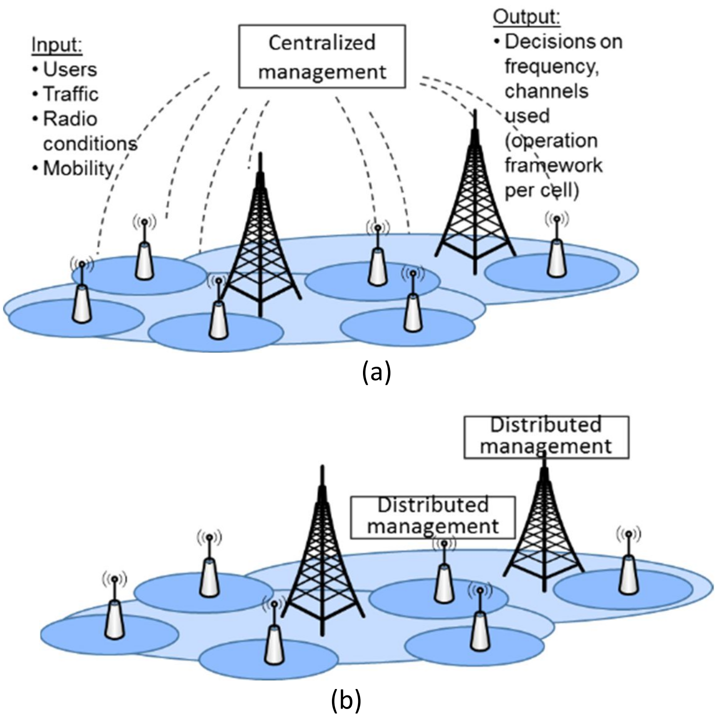


Figure 4-2 :Distributed and centralized radio resource management.

The rest of the subchapters are structured as follows. A survey of related works in the dynamic resource allocation area and the status of regulations about the utilization of the 3.5 GHz band are given. An introduction of the channel selection mechanism and our proposed solution is presented. Evaluation of potential solutions and some preliminary system-level simulation results are discussed. Finally, a conclusion and an analysis of what machine learning techniques can bring to this solution.

4.2 Related Work

There is extensive literature around dynamic resource allocation. For instance, authors in [5] develop a super-radio formation algorithm to identify valid super-radios, i.e., a set of radios that can coexist on the same channel(s) via WiFi-like carrier-sensing mechanisms. In this sense, they elaborate on the coexistence-aware dynamic channel allocation for 3.5 GHz shared spectrum systems. Also, interference between cells, interference management (inter-cell interference coordination) and power control approaches are of high importance, especially in ultra-dense heterogeneous environments that change dynamically and with high frequency; these aspects are extensively investigated in [6][7][8]. Authors in [9] claim that cognitive wireless networks enable dynamic access to underutilized spectrum in the licensed band. This type of network provides higher values for data rates and capacity, as users tend to use data links instead of making voice calls, using enhanced radio access and decision making techniques. In addition, spectrum efficiency is determined by a variety of parameters such as successful assessment of channel availability, selection of suitable and short transmission links, and traffic distribution between site access points. Authors in [13] present the implementation of a SAS capable of dynamic frequency assignment and interference management which is critical for the success of the proposed framework. They also describe the efforts toward a spectrum sharing system by summarizing different interest groups' standpoint on the FCC proposed framework as well as an exemplary SAS architecture that accommodates the tiered access to shared spectrum and suitable approaches to achieve important SAS capabilities. Finally various applications of this specific framework in the areas of smart networks and armed forces are presented in [14] and [15], respectively.

Also the arena of collaborative research funded projects has been focusing on using more effectively the available spectrum. For instance authors in [17], starting with an analysis of the regulator regimes and the available standards coming from the ETSI bodies, propose extensions to the logical blocks of the ETSI-defined architecture for Licensed Shared Access (LSA) and introduce and validate some interesting new allocation algorithms. Even if our work focuses on bands below 6 GHz, it is worth mentioning that important efforts have already started on improving the overall system dynamicity by also using the new bands to be made available for access in the mmWave domain up to 100 GHz. For instance, authors in [18] stress in different key use cases and scenarios the benefits that can derive in a smarter and more dynamic usage of several bands, specifically analyzing potential impacts on relevant standards bodies as well as regulatory aspects.

4.3 System Aspects

Centralized management of the cellular system can be expanded with distributed management by moving management decisions related to RAT/spectrum/channel selection closer to the node level as depicted in Figure 4-3 In this figure our work is focused on the block titled “RAT / Spectrum/ Channel Selection”.

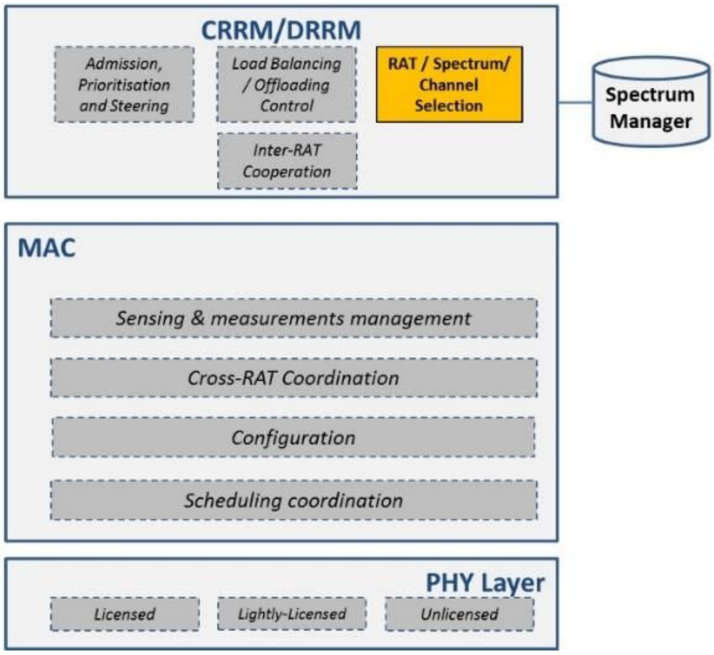


Figure 4-3: RAT/spectrum/channel selection component.

Specifically, the proposed algorithm initially runs in a distributed manner in order to limit the excessive signaling of centralized solutions in dense environments. However, in the case where the distributed approach does not provide satisfactory solutions (i.e., not reaching the desired KPI), a centralized approach can also be used. Based on the increased demand for sub-6 GHz spectrum, stemming from the new planned services in future wireless networks that will cause exponential growth in wireless data traffic, changes must be done in order to obtain QoS and capacity expansion. Therefore, standards and regulatory bodies are increasingly pursuing policy innovations based on the paradigm of shared spectrum, which allows spectrum bands that are underutilized by primary owners to be exploited opportunistically by secondary devices. Specifically, our algorithmic solutions focuses on the SAS in the 3.5 GHz band, which is a technology consisting of the following three-tier model:

- Incumbent Access.
- Priority Access Licenses (PAL).
- General Authorized Access (GAA).

Incumbent users are able to utilize a channel over the user that is at a lower level (PAL or GAA). PAL and GAA users are controlled by the SAS system and thus must register and check all of their operations in order to provide a secure and interference-free environment to higher-tier users. The basic rules defined by the FCC [10], which are required for the SAS model that must be introduced for the usage in the 3.5 GHz band, are:

Incumbent Access:

- These users will be protected from harmful interference from PAL and GAA users.
- Authorized federal and grandfathered Fixed Satellite Service (FSS) users currently operating in the 3.5 GHz band.
- Other incumbents including for example Department of Defense (DoD) radar.

Priority-Access Licenses (PAL):

- PAL will use competitive bidding within the 3550-3650 MHz portion of the band, as illustrated in Figure 4-4
- Each PAL will use a 10 MHz channel in a single census tract for three-years (a specific license is to be acquired).
- Up to seven total PALs may be assigned in any given census tract with up to four PALs going to any single applicant.
- Applicants may acquire up to two-consecutive PAL terms in any given license area during the first auction.

General Authorized Access (GAA):

- GAA users are permitted to use any portion of the spectrum.
- 3550-3700 MHz band is not assigned to a higher tier user.
- May also operate opportunistically on unused Priority Access channels 3650-3700 MHz, as illustrated in Fig. 2 (Light-gray spectrum).
- GAA users are permitted to use 80 MHz of all the available and not assigned to any higher tier in the 3.5 GHz range.
- GAA users do not have to obtain an individual spectrum license.

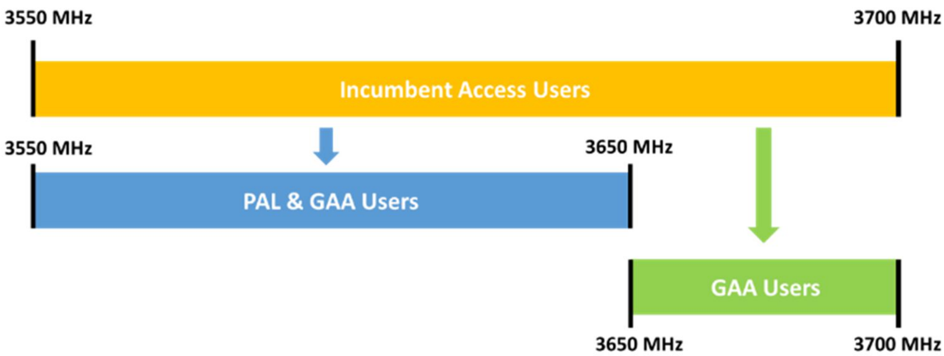


Figure 4-4 : 3.5GHz spectrum allocation for each tier of the SAS model.

The regulations and specifications listed above for the SAS model on the 3.5 GHz are quite complicated but are a necessity in order for the model to work properly and therefore are implemented in our 5G simulator. This is essential for the evaluation of the proposed solution as it allows that the obtained results are as close to the real-life implementation as possible. Finally, it has to be mentioned that the SAS system does not work as a real time scheduler, it can only instruct the cease of transmissions of lower tiers to the area, if a higher tier is present and transmitting.

4.4 Proposed Algorithm

Figure 4-5 presents a chart of the messaging sequence of the proposed algorithm. Initially, the algorithm first runs in a distributed way, in every cell inside the Distributed Radio Resource Management (dRRM) functional block, which is located at every Base Station (BS). From there, information about several parameters, like the availability of RAT/spectrum/channels cell of the users, utilized license schemes and neighboring cells frequency assignments, can be gathered. After the successful selection of a particular channel, all the appropriate information and specifications are requested and received to and from the Medium Access Control (MAC) layer. At the MAC layer, scheduling and inter-RAT coordination mechanisms are enabled, run if necessary and send information to the dRRM. Then, the physical layer is reconfigured to the new selected RAT, band, and channel. Finally, the centralized Radio Resource Management (cRRM) is invoked whenever the dRRM algorithm does not achieve the desired KPIs, or in general the thresholds required by the system for optimal results. Hence, the dRRM sends the available information from all cells to the cRRM where a centralized channel selection runs, in order to produce the best possible solution. At that time, the MAC is called to perform the scheduling and inter-RAT coordination, sending new configuration instructions to all dRRM.

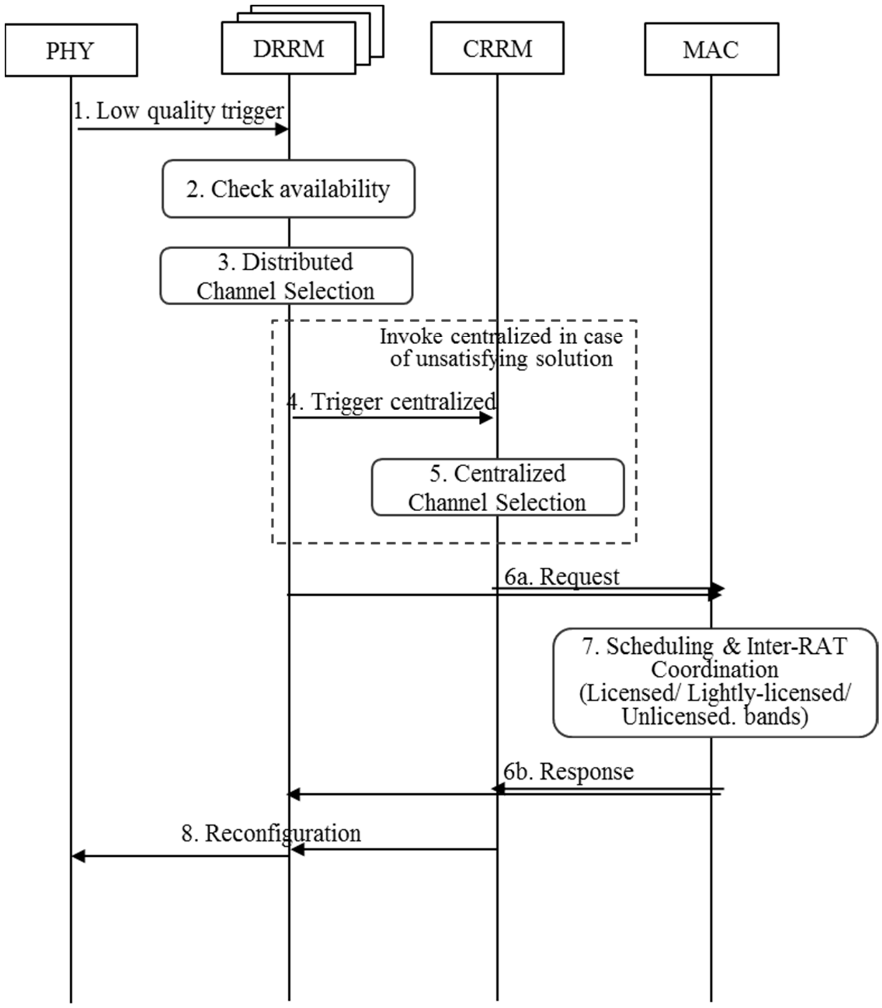


Figure 4-5 : Message sequence chart for RAT/spectrum/channel selection.

Each BS is able to acquire information about the spectrum bands, channels and UEs and can choose which band can be selected at any time either by utilizing a licensed or unlicensed channels, or by selecting the 3.5 GHz band Figure 4-6. Then, the SAS mechanism is enabled in order to provide information about the availability of channels and to select the proper one. In addition, after utilizing a 3.5 GHz channel, the algorithm keeps checking for any information about the channel that is given to a specific tier. In particular, the PAL and GAA users may receive instructions to change channel (or even band if no channels are available) whenever a higher tier user needs to use the specific channel. Most importantly the SAS is not a real time scheduler and this is why the proposed algorithm is needed, which provides the appropriate efficiency and effectiveness of faster and better selection of the channels without any consequences on the proper functioning of the system.

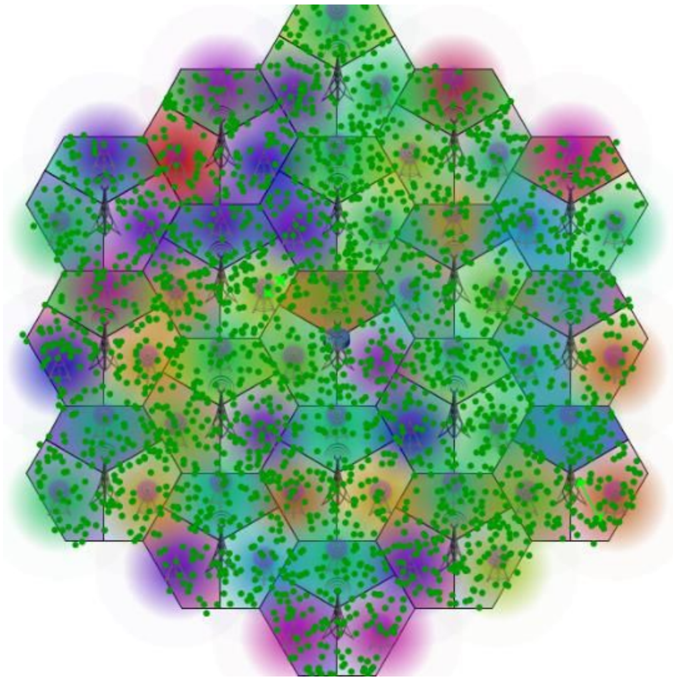


Figure 4-6: Map of enviroment.

As a next step we have designed an extension of the proposed solution in order to find appropriate resources and allocate resources among the different user categories. Specifically, through our tests we have seen that incumbent users experience higher throughputs and lower latencies than the required ones. This may lead to worse results for the other categories (PAL and GAA). Thus, we consider a more dynamic approach by taking into account the minimum required throughput of each user category and particularly of the incumbent users that need to achieve a minimum throughput. In our solution we consider that no service degradation should occur for the incumbent, priority users. To achieve this, the algorithm uses as input the required throughput of the incumbent users and then proceeds to the optimum allocation of Resource Blocks (RB), in order to achieve the desirable/ required throughput. When assigning the physical resources, we consider the traffic requirements of the specific user. We acquire the traffic requirements (i.e. throughput that the user can achieve) by determining the transmission path loss and shadowing effects in order to assign the optimum resources. Having considered these aspects, we implement a resource allocation algorithm where network resources are allocated in order to maximize the overall quality of all the user categories by taking into account their requirements (e.g., required throughput), as well as the priority of each user. In general, the proposed algorithm offers a functionality that helps into respecting certain QoS requirements of the incumbent, priority users and at the same time achieving better overall resource allocation in the system. Figure 4-7 provides a flowchart of the proposed algorithm.

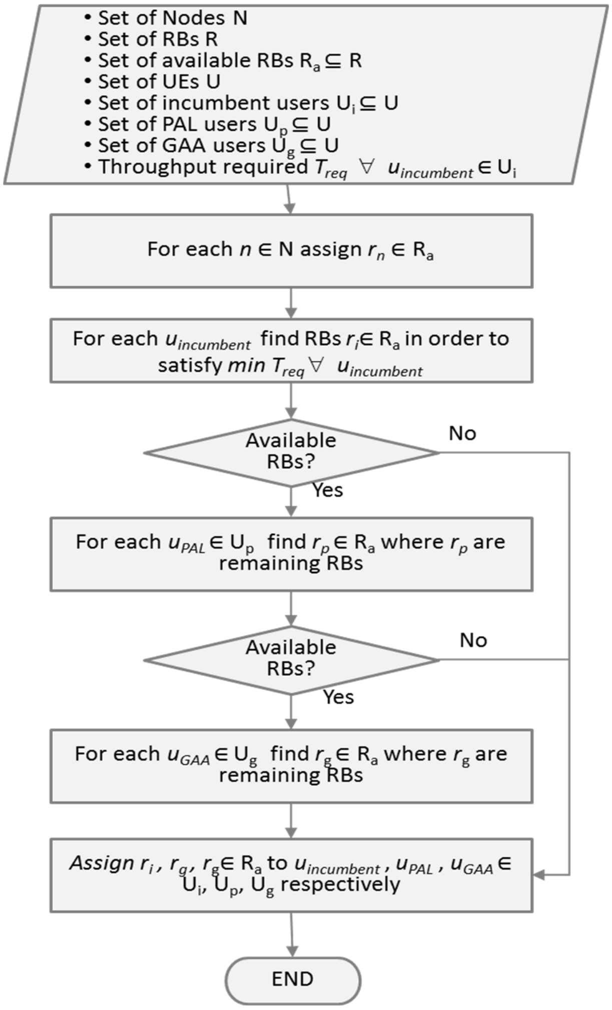


Figure 4-7 : Algorithm overview.

4.5 Evaluation

For the evaluation of the proposed algorithm, system-level simulations have been conducted. The implementation of our solution was performed using a proprietary 5G, system-level simulation tool which is fully developed in Java with various capabilities and has been calibrated according to the 3GPP TR 36.814 [11]. The simulator takes into account various parameters such as traffic level, available infrastructure elements, available channels, and then runs the most fitting test cases. As a result, the Cumulative Distribution Function (CDF) of coupling loss and downlink SINR has been checked in order to calibrate the tool with leading vendors and operators, for example Nokia, Huawei, Orange etc. The configuration is fully customizable so as to include various types of cells (i.e., macro and small cells). The system-level simulator takes into account a list of particularly interesting aspects, like:

- *Environment and configuration:* Environment concerns aspects related to traffic (e.g. proper modeling of 5G use cases like eMBB, mMTC etc.), anticipated load, mobility and radio conditions (e.g. propagation models). This is triggered by the fact that different traffic characteristics may apply depending on the specific considered use case.

- *System:* System aspects include considerations relevant to network deployment (e.g. small cells and macro cells). Spectrum aspects are considered for utilization of bands below 6 GHz and to be expanded in mmWave as well. Abstraction of the physical/MAC layer and RRM algorithms are also considered.
- *Analytics:* The simulation results are evaluated against KPI targets (e.g. in terms of throughput and success ratio of messages). The results are analyzed and visualized in order to provide a useful insight of the performance.

In order to evaluate the proposed algorithm in a heterogeneous system, our simulator can be utilized with all the above aspects implemented. Different scenarios and test cases are introduced to the simulation environment and will be further described and analyzed in detail. Additionally, we experiment with various traffic mix situations, specifically the percentage of the different licensed users in order to obtain an even broader knowledge of the algorithm capabilities, and overall performance for the specific network environment that is introduced to the simulator.

Initially, our algorithm is deployed to the system-level simulator that utilizes a number of macro and small cells at 2 GHz and 3.5 GHz bands, respectively, with a number of different channels. Specifically, the parameters imported to the simulator are: 19 macro BS, each one with three cells and 9 small cells, giving us a number of 171 small cells and a total of 228 macro and small cells throughout the network. In addition, we have utilized 15 channels at 10 MHz bandwidth for every cell and 10 of them can be assigned to the PAL users which is the 2/3 of the total channels. The Incumbent users are able to use any of the 15 channels of the system and of course are able to transmit to all of them at the same time, if needed. The GAA users are able to use all of the available spectrum but are permitted into using only 80 MHz of the 150 MHz at each time. Thus, only 8 channels can be assigned to the GAA users of the total of 15 that are available, but they can choose which of them will be utilized dynamically. In order to better assess our algorithm, a number of test cases and scenarios has been created. Table 4-1 and Table 4-2 present the parameters of the system-level simulation. It should be mentioned that the test cases in Table 4-2 are distinguished by the number of sessions simulated per day per user. This is reflected to the overall traffic of the network (since more sessions per day mean increased network traffic) and this leads to interesting findings with respect to the algorithm performance as more load is introduced in the system.

Table 4-1: Simulation parameters.

Parameter	Value
Macro BSs (with 3 cell each)	19
Inter-site distance	500 m
Small BSs	171
Total RAT Devices	228
Total Number of UEs	1500
File size (MByte)	15.0
Traffic model	FTP download
Total Number of channels	15 (3550-3700 MHz)

PAL Number of channels	10 (3550-3650 MHz)
GAA Number of channels	8 (3550-3700 MHz)
Traffic Mix	Incumbent:33% PAL:33% GAA:33%
Number of resource blocks:	50 PRBs per channel
Transmission Time Interval (TTI) length :	1ms
Simulation time	30s

Table 4-2: Test cases.

Test case	Sessions/day/user
1	23040
2	27360
3	31680
4	36000
5	40320
6	44640
7	48960

To evaluate the impact of our proposed algorithm we assume that incumbent users have a recommended bandwidth for file transfer with the use of FTP protocol of 20-25 Mbit. By setting a minimum quality level for incumbent users, the system proceeds to a better resource allocation of spectrum to other user categories (PAL/GAA). Moreover, even though the same resources have been utilized, meaning that no new channels or spectrum resources have been used, the proposed algorithm has achieved a better overall result for the whole system.

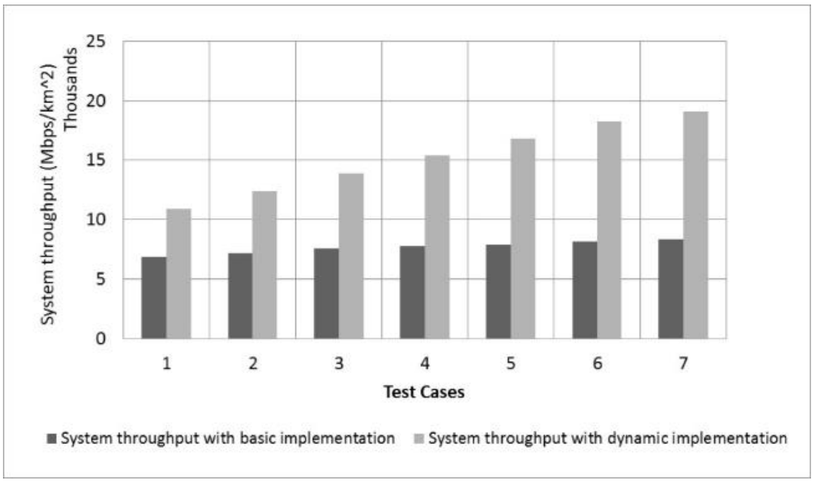


Figure 4-8 : Evaluation of different test cases (System throughput).

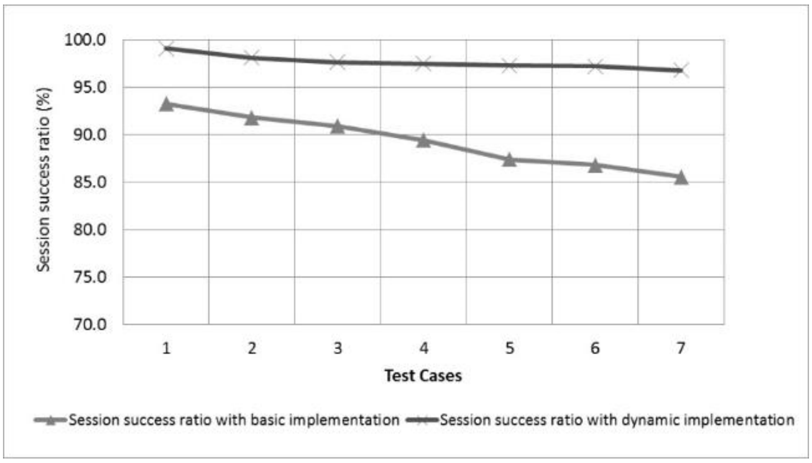


Figure 4-9 : Evaluation of different test cases (Session success ratio).

Figure 4-8 and Figure 4-9 show the impact of dynamic resource allocation between incumbent users, PAL and GAA (referred to as 'dynamic implementation') compared to basic implementation (which means that the system has fixed provisioning of resources for incumbent, PAL and GAA regardless of demand and without taking into account required QoS by users). Based on the load of the system for the cases 1 to 7 as described in Table 4-2, it can be observed that the dynamic algorithm managed to perform better than the basic one, at the overall throughput of the system, by making a better allocation of the same amount of resources (spectrum). Specifically, as mentioned before, the system has a pool of 15 channels for both basic and dynamic implementations and the only aspect that changes is the QoS provisioning for the incumbent users based on their needs. For example, when an incumbent user needs to stream high definition video, the recommended bandwidth is about 6 Mbit per second [12] in order to have a flawless streaming without any buffering or pauses of the video. If a user receives a much higher value than the recommended bandwidth for the consumed content, in our dynamic approach the extra speed (i.e. utilization of resources, RBs) can be given to other user categories.

4.6 Introduction of Machine Learning

In order to enhance the performance of the system, machine learning principles can be utilized. In particular, the PAL and GAA users may receive instructions to change channel (or even band if no channels are available) whenever a higher tier user wish to use the specific channel. By leveraging on the basic SAS principles the algorithm will be able to acquire knowledge about the channels utilization and specifically, if the band is occupied by Incumbent, PAL or GAA. In addition, band allocation, duration of usage, recurrence of usage and even location information can be collected in order to be able to select and change to a specific cell/channel faster and more reliably by predicting the utilization of the 3.5 GHz band of the Incumbent and Priority Access Licenses users.

With that in mind would be possible to locate the best channel that should be utilized each time. The information collected will be for each specific location for the algorithm to provide the CRRM and/or DRRM with the appropriate knowledge of the system when is required. Distributed interference management could even work as good as a centralized when the intelligent nodes are able to understand their environment and collect information about the neighbouring cells, only by analysing the information of their own. For example by evaluating the packet losses, modulation, throughput, SINR, etc, could be able to determine what channels the cells in the vicinity are transmitting. Also the system would be able to acquire information about the conditions (e.g., throughput, SINR, etc.) of the channels of current time and most importantly at a future time. The algorithm, could determine whether a channel is good or bad, i.e. to be selected or not by a user, or even schedule the channels allocations for future usage. By applying statistical learning techniques will be easier to automatically identify patterns in data that can be used to make more accurate predictions. Machine learning can offer faster reaction times to the system in cases of specific commands from the SAS, have to be implemented. Commands such as power control or switch of frequency range when a higher tier appears to the network in 60 seconds will cause significant problems. Many operators, wireless forums and companies that are pushing FCC on changing this regulation time to 600 seconds such as Motorola Solutions and Nokia Solutions; also WinnForum [3], Google and WISPA approve this approach. This is an important aspect of the whole system that the proposed algorithm with the introduction of machine learning techniques could possibly solve by predicting the appearance of a higher tier user.

4.7 Enabling Dynamic Spectrum Access

5G wireless networks are being designed to comprise and utilize Multi-Radio Access Technologies (multi-RATs) in a highly heterogeneous environment. Making use of the existing infrastructures and integrating the frequency bands of different wireless networks, is promising to improve the system data rates, achieving high capacity, Quality of Service (QoS) and help meet low latency requirements. Dynamic channel selection and band identification are challenging problems due to coexistence (of licensed, lightly-licensed and unlicensed users) requirements. In that manner, the available nodes may have a variety of RATs and spectrum that will be able to utilize and should be in position to do it in a reliable and optimum way, which is referred as to extended Dynamic Spectrum Access (eDSA). The eDSA idea consists of new interfaces and two main parts, which are the Radio Resource Manager (RRM) and the Medium Access Control (MAC) layer offering services and mechanisms that can facilitate efficient utilization of the available radio resources across different spectral bands using various RATs in real time.

4.7.1 Functional Framework Overview

The goal of this section is to give an overview of the system by discussing the considered protocol stack, which is mapped onto a virtualized architecture. In order to support eDSA, the protocol stack has to be capable of managing multiple RATs in an optimal way. represents the protocol stack developed in the SPEED-5G project [19]. The whole system is configured and managed by an RRM entity through the coordination interface. The RRM objective is the maximization of the system spectral efficiency, summed across the different available frequency bands, relying on cross RAT scheduling to efficiently aggregate carriers for capacity boosting or traffic offloading.

In order to natively support resource aggregation on different RATs, the user plane includes a Multi-Path-TCP layer, for scheduling TCP services over different radio bearers in parallel. These TCP paths are embedded in different GTP tunnels in order to redirect services to one or more RATs of a cell or of multiple cells, in order to support Coordinated Multipoint (CoMP) and virtual multi-cell Multi Input Multi Output (MIMO) schemes. The control plane is handled by the 5G Radio Resource Control (5G-RRC) function which manages the control procedures of the supported RATs, applying the eDSA procedures coordinated at the cRRM level. Figure 4-10 highlights how SPEED-5G protocol stack can handle multiple RATs at the MAC and PHY layers configured by the RRC layer, based on cRRM decisions.

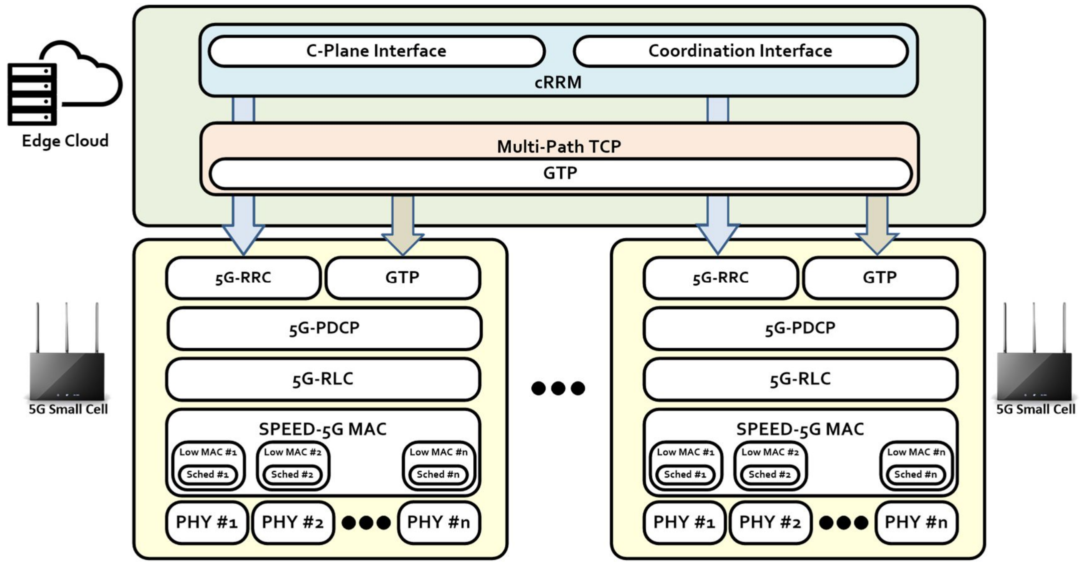


Figure 4-10: Multi-RAT protocol stack for eDSA support.

4.7.2 MAC aspects

As depicted in Figure 4-10, the MAC layer is foreseen to provide support for simultaneous operation of multiple RATs. In order to properly handle issues related with multi-RAT operation a decomposition of the MAC layer into two sublayers which are referred to as higher-MAC and lower-MAC is proposed. The higher-MAC is composed of a set of functions which do not require real-time execution and are not specific to any particular RAT. Their main aim is to coordinate and intelligently manage underlying lower-MAC entities, applying the long-term cRRM decisions at the MAC level. When decentralized RRM is considered higher-MAC and dRRM are collocated. As stated before, lower-MAC includes different RAT-specific functions (for example, KPI management, channel (de)multiplexing, and (de)framing) and functions which need to be executed in real-time, such as scheduling. In this way, the higher-MAC can be seen as a point-of-convergence for the protocol stack dealing with the control path, decoupled from the user plane and managing the sets of possible bearers. The introduction of higher-MAC intended to open the way for native support of multi-connectivity, integration of innovative multiple access techniques and use of different forms of cross-RAT spectrum and carrier-aggregation techniques (i.e. heterogeneous resource aggregation).

The following describes a list of MAC features which are intended to allow efficient operation of single-RAT and multi-RAT devices across different technology-specific and technology-neutral bands (which include licensed, lightly-licensed, and unlicensed bands):

- Common and dedicated management of control-related traffic is required for proper operation of cellular radio systems. In order to take advantage of multiple RATs which may operate over different frequency bands with different licensing regimes and potentially offer different QoS, the MAC facilitates coordination of control traffic transmission across multiple interfaces available for transmission. Additionally, as resources allocated to control channels may experience different levels of interference, the proposed MAC framework allows dynamic adaptation of control channel configuration.
- Devices which support multiple RATs often have access to resources not available to single RAT devices. In order to fully exploit the availability of these resources, the MAC framework manages distribution of load/steering of traffic across different RATs. The traffic can be

distributed based on various pieces of information such as channel load, traffic QoS requirements, or retransmission status. Such a feature also facilitates a cross-RAT spectrum/carrier aggregation as it enables user data to be simultaneously transmitted over multiple air interfaces (in both DL and UL); this feature essentially provides similar capability as in LTE WiFi link Aggregation (LWA) and enhance-LWA (eLWA) [20]. In addition, the proposed MAC allows dynamic adaptation of different system parameters. This is necessary as different frame formats or channel configurations may need to be used for optimal performance, considering the type of traffic and channel conditions.

- Support for coexistence and ensuring fair access to a shared medium is one of the main design principles of the MAC protocol framework. This is achieved either by implicit coordination or explicit coordination between nodes. The implicit coordination is usually based on different forms of sensing, duty cycle adaptation, or frame structure adaptation; whilst the explicit coordination requires exchange of messages between nodes for the reservation and identification of transmission opportunities. In order to enable efficient operation in scenarios where multi-RAT capable devices will operate using shared resources, MAC is able to support different coexistence mechanisms to coordinate coexistence functions of different RATs. For instance, the proposed MAC design is capable of tuning sensing parameters, adapting duty cycle, or changing MAC frame formats. In case of a multi-RAT device, the proposed MAC also supports the use of coexistence mechanisms supported in different RATs to enhance operation of another RAT.
- Different physical layers which operate using different waveforms (for example, Orthogonal Frequency Division Multiplexing (OFDM), Post-LTE filtered multicarrier modulations as well as Non-Orthogonal Multiple Access (NOMA) techniques) can all be supported by the proposed framework. In order to fully exploit this diversity, the SPEED-5G MAC is capable of tuning different parameters of underlying physical layers.
- Obtaining measurements is necessary for optimal allocation of resources in a wireless network. Inclusion of a dedicated monitoring interface between cRRM, higher-MAC and Physical Layer is proposed in order to i) enable efficient and dynamic coordination of RATs, and ii) provide sensing and KPI/QoS measurements to cRRM, so as to feed the coarse-grained centralized resource management, with information needed to coordinate small cell operations, over longer time-scales.

4.7.3 Multi-Rat MAC and RRM for Dynamic Spectrum Access

Figure 4-11 depicts the functional architecture of MAC layer in more detail, highlighting the higher-MAC and lower-MAC split. The former responsible for cell configuration, RAT coordination and coexistence in shared spectrum and the latter dealing with RAT-specific bearer management, data transmission and reception are also presented. In addition to the control plane and data plane, this block diagram interestingly includes a new kind of interface which carries all the sensing and KPI reporting data, used for both RAT coordination and eDSA enablement at cRRM, called the monitoring plane. An important aspect of this MAC design is that control and data planes are de-coupled, having the higher-MAC only dealing with control plane with no influence on the data plane (and allowing for independent scaling of control resources), which is handled at the lower-MAC level. Note that the higher layers include the 5G-RLC which handles the active logical channels carried over all the RATs supported by the small cell. Whereas in existing systems, the traffic queues are treated as separated entities corresponding to specific RATs (LTE, WCDMA, WiFi, etc.), this feature allows this block to expose to the cRRM (which is connected to a 5G-RLC entity), the list of all active logical channels with the associated QoS requirements. It also provides the buffer statuses, needed by resource management algorithms to take decisions on how to make an optimal use of spectrum and RAT resources for the group of small cells it has to coordinate.

The proposed design can not only support current RAN functional split options depending on the backhaul characteristics but also a new split option at the MAC level. Additionally, the proposed design provides sufficient flexibility to enable integration of any new RAT, any spectrum resource or new coordination algorithms hosted at higher-MAC level without impacting the lower-MAC layer design and entities.

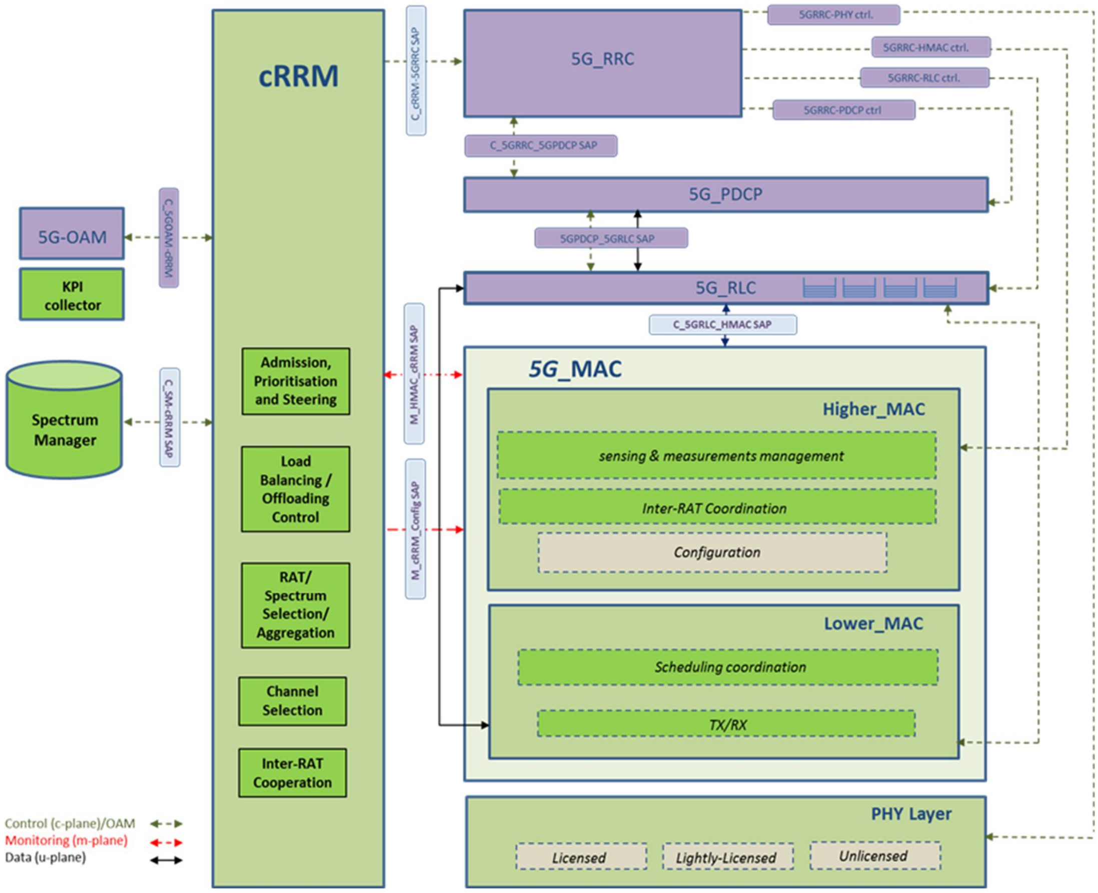


Figure 4-11: cRRM, Higher-MAC and Lower-MAC split.

The higher-MAC hosts RAT-independent functions designed to coordinate the cross RAT scheduling of the active logical channels, as well as to facilitate the operation in shared spectrum. In the case of a virtualized cRRM entity, the higher-MAC receives the eDSA decisions taken at cRRM level and applies them by configuring the lower-MAC instances. In the case of non-virtualized RRM, higher-MAC and dRRM functions can be co-located. This section deals with the description of the higher-MAC functions, depicted in Figure 4-12. Besides the configuration entity containing the common configuration parameters related to the cell, the UE provided by the cRRM or the OAM, higher-MAC is composed of the Sensing and measurement management, the Inter RAT coexistence coordination and management and the Inter RAT scheduling entities. The Inter RAT coexistence coordination and management entity is responsible for ensuring proper coordination of transmissions on different RATs when the medium has to be shared. This covers the following set of functions: Duty-cycle adaptation function is responsible for managing coexistence adopting a time-domain utilization pattern approach like the CSAT operation of LTE-U [21].

Listen-before-talk (LBT) function refers to coexistence methods based on the detection of on-going traffic on a shared frequency resource prior to the data transmission, mandated for LTE-LAA operation [26]. It manages the detection method (energy detection, preamble detection) and duration, and processes the sensing results comparing them with tunable detection thresholds. Frame format function is responsible for adapting the MAC frame structure to fit with the need of the different active bearers (e.g. uplink vs. downlink resource ratio, TTI or slot duration modification) or to react to the changes in the channel quality, enabling better coexistence The Contention coordination function has the role of managing the contention processes applied on the different RATs and frequency bands; it can control and adapt the random access schemes and resources on licensed bands or tune the contention access algorithms on shared spectrum. Channel configuration function copes with both identifying a suitable channel in non-licensed spectrum and forwarding the related parameters to lower-MAC (bandwidth, guard band, transmit power control, etc.). Multiple Access (MA) function covers the configuration and adaptation of lower-MAC components, taking advantage of orthogonal and non-orthogonal multiple access schemes an. In the case of NOMA, this function provides default parameters to manage the different actives UEs, given the UE localization, the QoS requirements of the active logical channels. Finally, the Sensing function enables coordination and effective use of different sensing mechanisms as well as sequential sensing of different bands. The parameters which can be configured by this block include sensing duration, minimum signal detection level, and sampling rate.

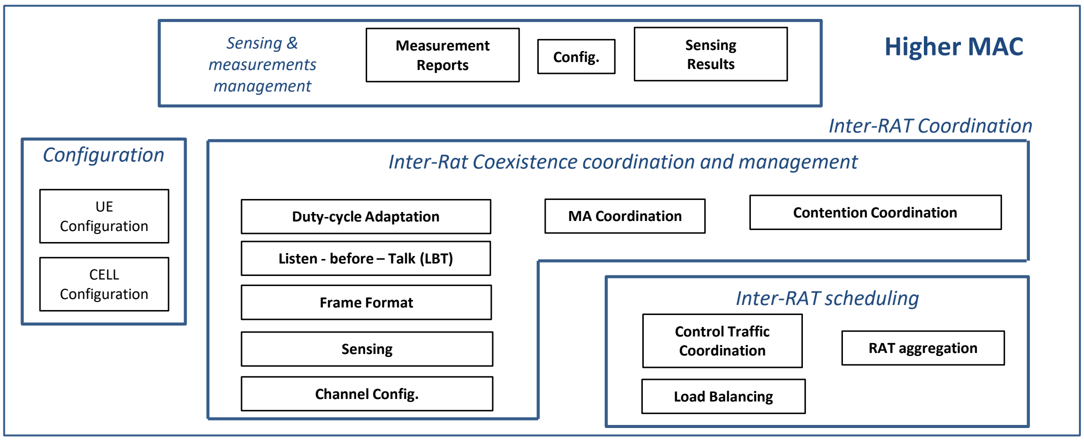


Figure 4-12: Higher-MAC functional architecture.

The Inter-RAT scheduling entity is capable of making short term decisions about the mapping of logical channels on the different RATs, depending on the load the small cell has to cope with and given the experienced QoS on the considered RAT. It covers the RAT Aggregation function that can take the decision of aggregating a set of lower-MAC different RATs according to the instantaneous cell loading and interference levels. This function works in cooperation with the Load Balancing function which makes a decision on which traffic should be steered and on which kind of spectrum band if RAT aggregation is required. It maintains a table on the active logical channels and their associated QoS requirements. Also in the scope of this higher-MAC entity, the Control Traffic function is responsible for steering and coordinating transmission of control traffic. Its main aim is to facilitate transmission of control traffic of one RAT over control channels of another RAT (this includes broadcast traffic and dedicated control traffic).

The Sensing and Measurement management entity is responsible for feeding the monitoring plane by collecting sensing results and measurement reports (link control KPIs) and forwarding them to the cRRM. It is composed of the Configuration function which sets the parameters (e.g. reported items, format and rate of reporting) for KPI collection and sensing reported from MAC to cRRM. The Measurement reports and Sensing results blocks are in charge of handling the actual data pertaining respectively to KPIs from MAC (BLER, latency, jitter of the active logical channels and CQI reports) and to sensing in different spectrum bands in which a small cell can operate. This data can be consumed locally within the higher-MAC as inputs to algorithms for real time adaptation of traffic offloading decisions, or can be forwarded to cRRM, feeding centralized spectrum monitoring and resource allocation modules.

The lower-MAC comprises of the different time constrained functions tied to the supported Air interfaces, like schedulers and Transmit/Receive (TX/RX) blocks. It embeds as many instantiations of lower-MAC entities shown in Figure 4, as the number of supported RATs. For instance, an LTE-A and a WiFi lower-MAC instances together with a post LTE-A lower-MAC instance, which is able to operate in licensed, lightly licensed and unlicensed spectrum can be hosted. The remainder of this subsection will detail only the latter instance, since LTE and WiFi legacy lower-MAC both can be effectively integrated into this MAC framework with little modifications. The lower-MAC is mainly composed of the Scheduling coordination and TX/RX entities.

The following functions are integrated into the Scheduling Coordination block: KPI Collection function collects and delivers different measurements and KPIs provided by the scheduler to higher-MAC through the monitoring plane, according to the configuration set by cRRM and higher-MAC. Carrier Configuration function applies the channel configuration according to the guidelines provided by the higher-MAC. Discontinuous Receive (CP-DRX) function provides the dynamic allocation of resources for common control traffic and the UE-specific control traffic. The behavior of CP-DRX can be dynamically adapted by the Control Traffic Coordination function located in the higher-MAC. The Coexistence Scheduler function applies the coexistence decisions taken by the higher-MAC by adapting different coexistence mechanisms supported by the lower-MAC, such as Listen-Before-Talk (LBT). The Ctrl function is responsible for applying changes in the scheduler configurations and policies provided by RRM while the Logical channel manager function maintains the scheduler configurations for the active logical channels, considering their required QoS. The Time and frequency domain scheduler function implements algorithms and mechanisms to determine which users are served in a given transmission time interval on set of spectral resources to these users (taking into consideration re-transmissions and QoS related information). The HARQ (Hybrid ARQ) function provides buffers management, and management of the layer-2 re-transmissions in both uplink and downlink. It reserves resources for re-transmissions, and indicates whether to use the same or different resources. The RACH function is responsible for reservation of resources for Random Access, handling contention resolution and prioritizing access requests. The function receives and applies decisions from the RACH Coordination function of the higher-MAC which allows for dynamic adaptation of random access strategies. The QoS function provides the QoS-aware operation for the scheduler prioritizing users in the time (and possible frequency) domain according to their QoS requirements in terms of latency and bandwidth. The Multiple Access (MA) function applies the guidelines provided by higher-MAC and provides the control over the multiple access strategies of all users in the cell during their session times, implementing a non-orthogonal multiple access (NOMA) strategy when it is required. Finally the Frame function receives frame configuration decisions taken in higher-MAC and provides the frame configuration to the PHY and also to the time-frequency domain scheduler. It defines the time transmission interval, the system slots and timings, the framing aggregation in terms of sub-frames (or slots) and defines the basic time-frequency grid for being used in the scheduler.

The high-level functions of the RRM and the interfaces to other layers and some other system components, are shown in Figure 4-11.

The “Admission / prioritization / steering“ block makes decisions about whether to admit a new traffic flow, what priority level it should have, and which base-station or base-stations to steer it to. The outputs will be to label the different types of traffic so that the MAC can steer to the appropriate RAT, and to keep a mapping between types of traffic and the available bands. Some pre-determined association rules could be established, depending on the requirements of each specific traffic type and depending on what spectrum bands or RATs are available. Additional criteria might be included in the traffic steering besides QoS and requirements, such as the expected data rate or the expected coverage provided by a specific RAT or band.

Load balancing (LB) aims at making efficient use of the limited spectrum to deal with unequal loads in order to improve network reliability by reducing the congestion probability in hot spot areas of cellular networks. LB algorithms may therefore trigger handover or cell reselection procedures.

The objective of the “RAT/spectrum selection and aggregation” functional block is to select a suitable band and RAT to be used by each type of traffic. It also selects the number of channels to be used within a band, if needed. Additionally, it may suggest, for each band, the operational configuration per each band. If needed, it may also suggest a MAC Frame configuration. This functional block takes its decisions based also on specific regulations inherent to each band.

The “Inter-RAT cooperation“ block is aimed at improving the coexistence with other RATs in the same band. As an example, in the 5GHz unlicensed band where the data transmission must coexist with WiFi, this block may manage the transmission and reception of the Request-To-Send and Clear-To-Send (RTS/CTS) control messages as well as the Network Allocation Vector (NAV), which is used for virtual sensing.

The process of selecting the best available channel for transmission can be performed either in RRM or MAC layer, depending on the time scale and on the available information. It can work at a finer level of granularity than the spectrum selection. The dual blocks may interact with each other. For instance, in the case of autonomous small cell operation, channel selection may be performed autonomously in the MAC layer, and the results may be optionally verified by the RRM.

The Spectrum Manager is essentially a database that keeps track of the spectrum that is available to the network operator, and also the policies and restrictions on the use of the spectrum. The KPI block keeps a record of the target KPIs and also keeps track of current performance in terms of the KPIs.

4.7.4 Hierarchical management of multi-RAT networks

A hierarchical management which blends distributed and centralized solutions for ultra-dense multi-RAT and multiband networks could be beneficial in order to overcome the challenges generated in these environments. The centralized management used as a baseline, can be expanded with distributed management by moving management decisions related to RAT/spectrum/channel selection closer to the node level as depicted in Figure 4-13. Thus, a centralized approach can trigger the distributed one and vice versa as already discussed. The proposed algorithm is based on a learning technique for RAT/spectrum/channel selection in the 3.5 GHz band for achieving better performance especially, in dense and congested 5G environments. According to various studies [22][23][24][25] this band is deemed suitable to boost capacity and for use by small cells with relatively limited transmission power and range. Also, similar studies in [27][28][29][30][31][32] have elaborated on capacity expansion aspects and 5G including usage of MIMO and mmWave. The proposed algorithm assumes availability of a pool of bands within various licensing schemes (licensed/ unlicensed/ lightly-licensed) to fulfill certain traffic requirements. The proposed algorithm will initially run in a distributed manner in order to avoid the excessive signalling overheads of centralized solutions in dense environments.

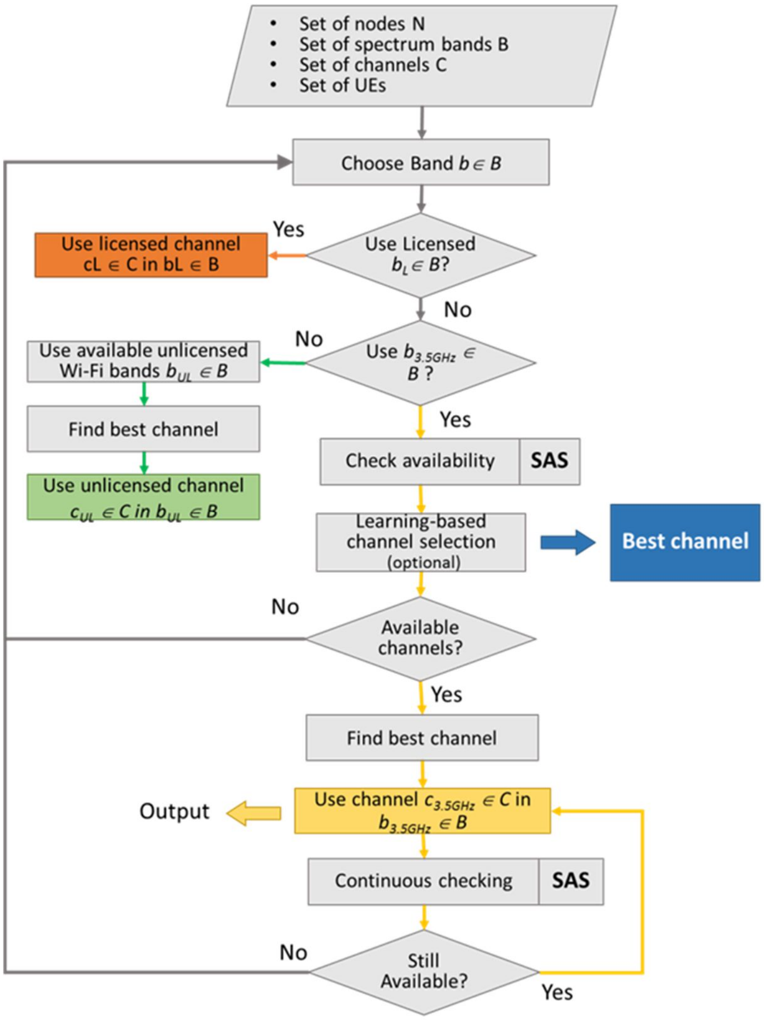


Figure 4-13 : Flowchart with learning capabilities.

4.8 Evaluation of learning algorithm

For the evaluation of such concepts, system level simulations are conducted. The implementation of our suggested solution was performed under a proprietary system-level simulation tool as discussed. In general, test cases with variable traffic loads based on the frequency of the arrival requests to the system. Additionally we will experiment with various traffic mix situations, specifically the percentage of the different licensed users in order to obtain an even broader knowledge of the algorithm capabilities, and overall performance for the specific network environment that will be introduced to the simulator.

In order to better assess our algorithm, a number of test cases and scenarios has been created. Table 4-3 presents the simulation parameters of the system level simulation with the different loads of the traffic expressed with a variety of request inter-arrival number per user per minute from users ranging from 6 up to 19. Also Table 4-4 introduces two different sets for the appearance of the three tier users to the system. This variable was introduced to our tests in order to simulate better the presence of this category which is responsible of creating the most challenges to the other two users.

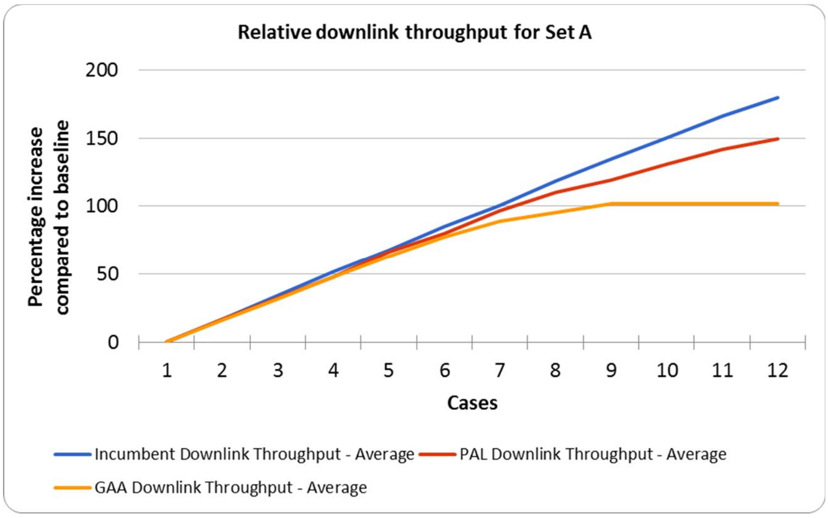
Table 4-3: Summary of simulation parameters.

Parameter	**Value**
Macro BSs (with 3 cell each)	19
Inter-site distance	500 m
Small BSs	285
Total RAT Devices	345
Total Number of UEs	58000
Request inter-arrival number per user per minute which represent cases 1-12	Exponential (6 - 19)
File size (MByte)	2.0
Traffic model	FTP download
Total Number of channels	15 (3550-3700 MHz)
PAL Number of channels	10 (3550-3650 MHz)
GAA Number of channels	8 (3550-3700 MHz)
Number of resource blocks:	50 PRBs per channel
Transmission Time Interval (TTI) length :	1ms
Simulation time	60s

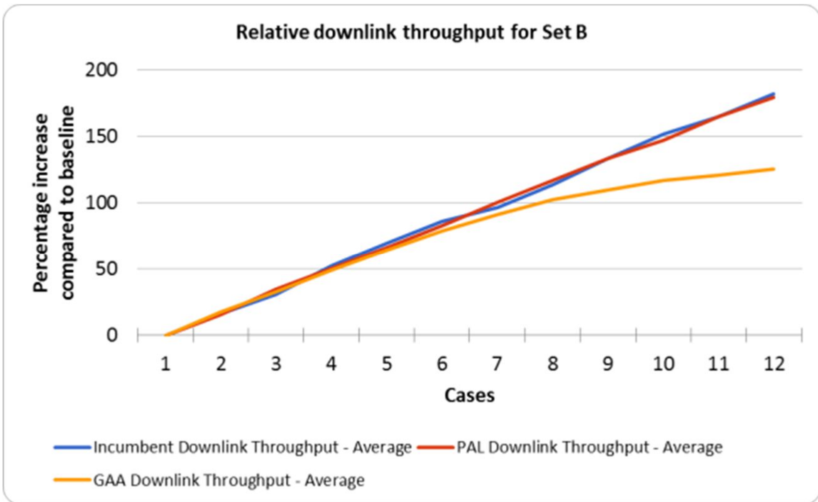
Table 4-4: Considered traffic mixes for sets.

Set	**Incumbent**	**PAL**	**GAA**
A	50%	17%	33%
B	10%	30%	60%

Figure 4-14 illustrates the relative downlink throughput for (a) set A, (b) set B. Cases 1-12 are also explained in Table *4-4*. Specifically it is shown that UEs of set B have higher throughput as request inter-arrival number per user per minute increases due to the fact that more channels are available due to less incumbents (according to the traffic mix) compared to set A. Moreover, relative latency for (a) set A, (b) set B which is shown in Figure 4-15 is better for set B compared to set A as request inter-arrival number per user per minute increases.



(a)



(b)

Figure 4-14 : Relative downlink throughput for (a) set A, (b) set B.

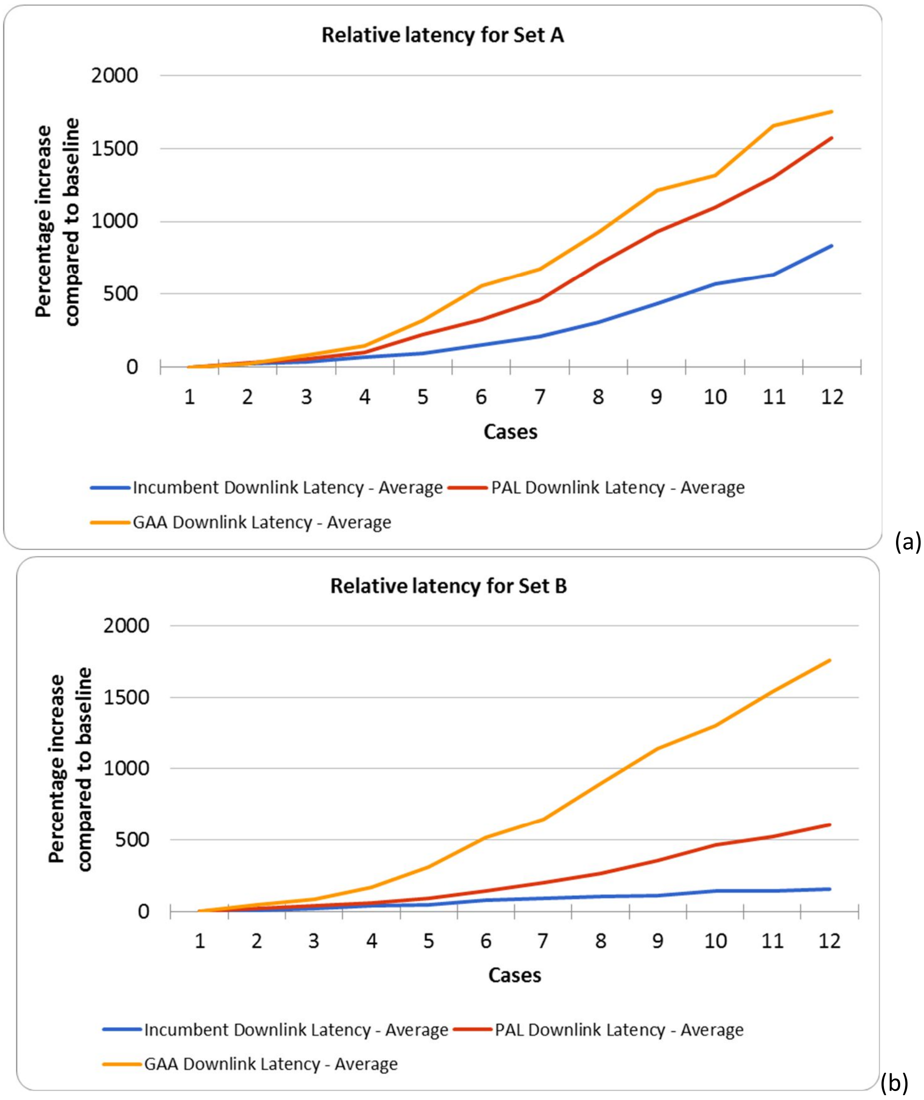


Figure 4-15: Relative latency for (a) set A, (b) set B.

4.9 Conclusions

Through this work and the evaluation, useful insights on the operation of SAS in 3.5 GHz systems have been provided. By satisfying the required quality of certain user categories it is possible to increase the overall performance of the system by better allocation of resources to other user categories as well. To enhance the performance of the system, machine learning principles have be utilized as discussed. By leveraging on the basic SAS principles, our enhanced algorithm will be able to acquire knowledge about the channels utilization and, specifically, if the band is occupied by incumbent, PAL or GAA users. In addition, by collecting the band allocation, duration of usage, recurrence of usage and even location information the selection and change to a specific cell/channel is faster and more reliably by predicting the utilization of the 3.5 GHz band of the incumbent and PAL users. With that information, it is possible to locate the best channel that should be utilized each time. By applying statistical learning techniques it will be easier to automatically identify patterns in data that can be used to make more accurate predictions.

4.10 References

[1] Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos, Valerio Frascolla, Panagiotis Demestichas, “Management of 3.5-GHz Spectrum in 5G Dense Networks: A Hierarchical Radio Resource Management Scheme”, IEEE Vehicular Technology Magazine

[2] Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Panagiotis Demestichas, Benoit Miscopein, Marcin Filo, Seiamak Vahid, Bismark Okyere, Michael Fitch, “Multi-RAT Dynamic Spectrum Access for 5G Heterogeneous Networks: The SPEED-5G Approach”, IEEE Wireless Communications, vol. 24, no. 5, pp. 14-22, October 2017

[3] Document WINNF-15-R-0045, Version 1.0.0, 22 July 2015

[4] Y. L. Lee, T. C. Chuah, J. Loo, A. Vinel, “Recent advances in radio resource management for heterogeneous lte/lte-a networks,” IEEE Communications Surveys Tutorials, vol. 16, pp. 2142–2180, Fourthquarter 2014.

[5] X. Ying, M. M. Buddhikot, S. Roy, "Coexistence-aware dynamic channel allocation for 3.5 GHz shared spectrum systems," 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Piscataway, NJ, 2017, pp. 1-2.

[6] Yida Xu et al., "Evaluation of Inter-Cell Interference Coordination with CAP model," 2015 11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE), Taipei, 2015, pp. 160-165.

[7] X. Lyu, H. Tian, W. Ni, R. P. Liu, P. Zhang, "Adaptive Centralized Clustering Framework for Software-Defined Ultra-Dense Wireless Networks," in IEEE Transactions on Vehicular Technology, vol. 66, no. 9, pp. 8553-8557, Sept. 2017.

[8] J. Ellenbeck, C. Hartmann, L. Berlemann, "Decentralized inter-cell interference coordination by autonomous spectral reuse decisions," 2008 14th European Wireless Conference, Prague, 2008, pp. 1-7.

[9] Al-Dulaimi, S. Al-Rubaye, J. Cosmas and A. Anpalagan, "Planning of Ultra-Dense Wireless Networks," in IEEE Network, vol. 31, no. 2, pp. 90-96, March/April 2017.

[10] Federal Communications Commission, “3.5 GHz Band / Citizens Broadband Radio Service”, Apr. 2015, available online at: https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-47A1.pdf.

[11] 3GPP TR 36.814 (V 9.2.0), “Further advancements for E-UTRA physical layer aspects”, Mar. 2017.

[12] K. Bilal, A. Erbad, "Impact of Multiple Video Representations in Live Streaming: A Cost, Bandwidth, and QoE Analysis," 2017 IEEE International Conference on Cloud Engineering (IC2E), Vancouver, BC, 2017, pp. 88-94.

[13] M. M. Sohul, M. Yao, T. Yang, J. H. Reed, "Spectrum access system for the citizen broadband radio service," in IEEE Communications Magazine, vol. 53, no. 7, pp. 18-25, July 2015.

[14] Parvez, T. Khan, A. I. Sarwat, Z. Parvez, "LAA-LTE and WiFi based smart grid metering infrastructure in 3.5 GHz band," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, 2017, pp. 151-155.

[15] T. Tuukkanen, S. Yrjölä, M. Matinmikko, P. Ahokangas, M. Mustonen, "Armed forces' views on Shared Spectrum Access," 2017 International Conference on Military Communications and Information Systems (ICMCIS), Oulu, 2017, pp. 1-8.

[16] Federal Communications Commission, “3.5 GHz SAS Conditional Approval Public Notice,”, Dec. 2016, available online at: https://www.fcc.gov/document/35-ghz-sas-conditional-approval-public-notice.

[17] V. Frascolla, M. Butt, N. Marchetti, A.J. Morgado, A. Gomes, et al., "Dynamic Licensed Shared Access - A new architecture and spectrum allocation techniques," VTC-Fall 2016, Montreal, Canada, 2016.

[18] M. Tercero, P. von Wrycza, A. Amah, J. Widmer, M. Fresia, *et al.*, “5G systems: The mmMAGIC project perspective on use cases and challenges between 6–100 GHz,” IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Doha, 2016.
[19] SPEED-5G Public Deliverable, “D3.2: SPEED-5G enhanced functional and system architecture, scenarios and performance evaluation metrics”, ICT-671705, H2020-ICT-2014-2, June 2016.
[20] LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (3GPP TS 36.300 version 13.3.0 Release 13)
[21]LTE-U Forum, LTE-U CSAT Procedure TS V1.0, http://www.lteuforum.org/uploads/3/5/6/8/3568127/lte-u_forum_lte-u_sdl_csat_procedure_ts_v1.0.pdf, October 2015
[22]R. Matsukawa, T. Obara, and F. Adachi., “A dynamic channel assignment scheme for distributed antenna networks”, Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th. IEEE, 2012.
[23]S.-J. Kim, I. Cho, Y.-K. Kim, and C.-H. Cho, “A two-stage dynamic channel assignment scheme with graph approach for dense femtocell networks,” IEICE Trans. Commun., vol.E97-B, no.10, pp.2222–2229, Oct. 2014.
[24]M. Ismail et al., “A Distributed Multi-Service Resource Allocation Algorithm in Heterogeneous Wireless Access Medium”, IEEE Journal on Selected Areas in Communications, vol.30, n.2, pp.425-432, February 2012.
[25]G. Yu, “Multi-Objective Energy-Efficient Resource Allocation for Multi-RAT Heterogeneous Networks”, IEEE Journal on Selected Areas in Communications, vol.33, n.10, pp.2118-2127, October 2015.
[26]3rd Generation Partnership Project. Technical Specification Group Radio Access Network; Study on Licensed-Assisted Access to Unlicensed Spectrum; (3GPP TR 36.889 version 13.0.0 Release 13).
[27]P. Aquilina, A. C. Cirik and T. Ratnarajah, "Weighted Sum Rate Maximization in Full-Duplex Multi-User Multi-Cell MIMO Networks," IEEE Transactions on Communications, vol. 65, no. 4, pp. 1590-1608, April 2017.
[28]D. Zhang; T. Muhammad; S. Mumtaz; J. Rodriguez; and Takuro Sato, “Integrating energy efficiency analysis of massive MIMO-based C-RAN,” EURASIP Journal on Wireless Communications and Networking, vol. 2016, no.1, pp. 277-285, Dec. 2016.
[29]D. Zhang; Z. Zhou; C. Xu; Y. Zhang; J. Rodriguez; T. Sato, "Capacity Analysis of Non-Orthogonal Multiple Access with mmWave Massive MIMO Systems," to appear in IEEE Journal on Selected Areas in Communications, doi: 10.1109/JSAC.2017.2699059
[30]Z. Ding, L. Dai and H. V. Poor, "MIMO-NOMA Design for Small Packet Transmission in the Internet of Things," in IEEE Access, vol. 4, no. , pp. 1393-1405, 2016.
[31]S. Mumtaz, A. Alsohaily, Z. Pang, A. Rayes, K. F. Tsang, J. Rodriguez, "Massive Internet of Things for Industrial Applications: Addressing Wireless IIoT Connectivity Challenges and Ecosystem Fragmentation," in IEEE Industrial Electronics Magazine, vol. 11, no. 1, pp. 28-33, March 2017
[32]Ali, M., Mumtaz, S., Qaisar, S. et al. "Smart heterogeneous networks: a 5G paradigm" Springer,Telecommun Syst (2017).

5 Network slicing

This chapter summarizes the main findings of the work in 5G RAN Slicing. Specifically, it covers the work related to 5G slicing management functionality and validation. Therefore it describes the implementation actions that have to been taken in order to support slicing functionality based on the requirements and modeling.

5.1 Introduction

The wireless world has exhibited a huge progress over the past three decades. Currently, tremendous resources are allocated for conceptualizing and realizing the 5th generation (5G) of wireless/mobile communications. This push towards 5G is motivated by a combination of business requirements and of technology trends that can efficiently boost the performance of various parts of the infrastructure. The business requirements are to accomplish:

- A higher valorization of the infrastructure, through the generation of new revenue streams, which are possible through the introduction and provision a much larger set of services compared to today; these services have highly demanding and heterogeneous QoS (Quality of Service) requirements;
- Drastic improvements in resource usage and therefore cost efficiency.

Services are associated with numerous verticals sectors, e.g., energy, health, media provision, water/environment management, etc. Due to the heterogeneous QoS requirements, services are classified with respect to aspects like: whether they involve extreme mobile broadband (eMBB), or whether they require ultra reliable and low latency communications (URLLC) or even massive amount of sensors and machines that need to communicate with each over (mMTC) Figure 5-1.

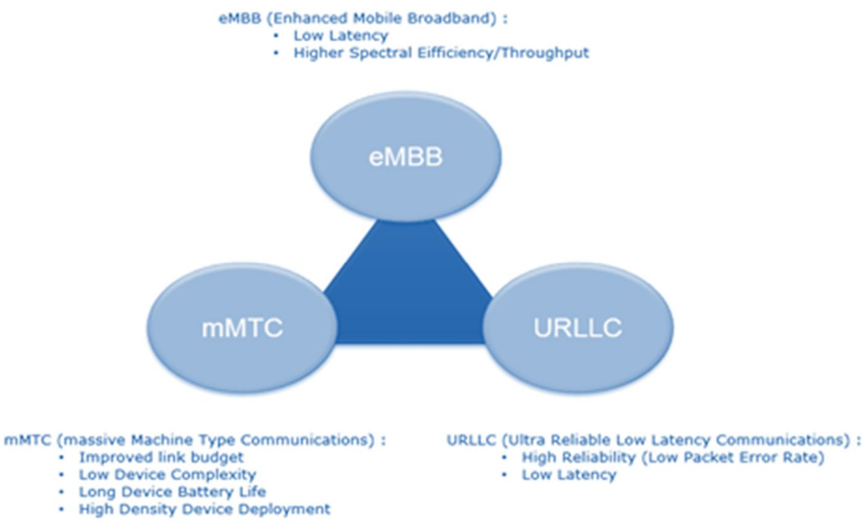


Figure 5-1: Overview of services and requirements (Source: 5G-PPP).

From the technology perspective, emerging trends are: the introduction of a new radio (NR) air interface, the use of more spectrum (below and above 6 GHz), the constant shrinking of cell sizes and the eventual elimination of cells (cell-less architecture), the “softwarization” of large parts of the infrastructure in conjunction with the intelligent and flexible splitting of functionality in hardware and software, the realization of wireless infrastructures through “cloud” technologies (in order to benefit from the merits of the cloud approach). In parallel, there are pushes to essentially advance the management intelligence, in order to achieve a constantly agile (reactive/proactive, automated/prescriptive, fast, reliable, trustworthy), and, therefore, more efficient system behaviour. At first, all these will lead to a most powerful, yet complex, 5G Radio Access Network (5G RAN). In order to efficiently provide the diverse services/QoS levels, through the complex and powerful network, the notion of slicing has been introduced. In general terms, a slice can be seen as a logical network that relies on a subset of the physical resources of a network. In this respect, there can be 5G RAN slicing, as well as the slicing of other segments, in order to support the end-to-end connectivity. In this respect, a network is partitioned into a set of slices. Each slice should be adequate for delivering a specific service. Beyond the general discussion above, the following issues need to be meticulously addressed:

- The detailed definition of “what is a slice”, taking into account the current work done in standards, and the delineation between RAN slices and the slicing of other segments;
- The integration of RAN slicing concepts in the context of a 5G architecture;
- The specification and development of the appropriate management functionality for 5G RAN slices;
- The realization of activities for validating (through simulation) the 5G-RAN slice-management functionality;
- The realization of activities for demonstration, dissemination, and, potentially, impacting standards.

A network slice is a virtual network that’s created on top of a physical network in such a way that it gives the illusion to the slice tenant of operating its own dedicated physical network. A network slice is a self-contained network with its own virtual resources, topology, traffic flow and provisioning rules. In Figure 5-2**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** certain layers are depicted. Specifically, the Service Instance Layer represents the services. Each service is represented by a Service Instance. A Network Slice Instance Layer provides the network characteristics which are required by a Service Instance. It can be a set of network functions. Finally, the resource layer is the layer of physical assets for computation, storage or transport including radio access. Some main characteristics of network slice instance are provided below:

- A network slice instance may be fully or partly, logically and/or physically, isolated from another network slice instance.
- The resources comprises of physical and logical resources.
- A Network Slice Instance may be composed of Sub-network Instances, which as a special case may be shared by multiple network slice instances. The Network Slice Instance is defined by a Network Slice Blueprint.
- Instance-specific policies and configurations are required when creating a Network Slice Instance.
- Network characteristics examples are ultra-low-latency, ultra-reliability etc.

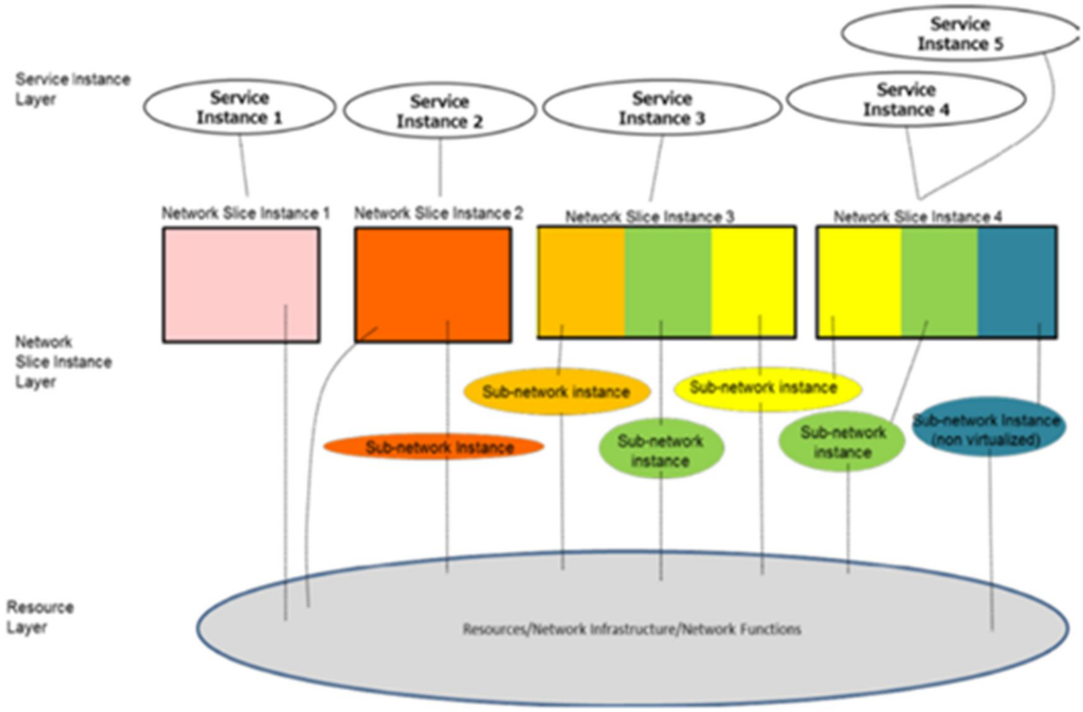


Figure 5-2: Network Slicing Concept [6].

5.1.1 Requirements from standards

The following requirements are related to slicing operations and have been documented by 3GPP in[12][13]. The system shall allow the operator to:

- Create, modify, and delete a slice
- Define and update the set of services supported in the network slice
- Assign a device to a network slice based on subscription, device type and services provided by the network
- Configure the information which associates a device or a service to a new slice
- Assign a device to a network slice, to move a device from one network to another and to remove a device from a network slice

Other requirements shall include also:

- Enable a device to simultaneously assigned to and access services from more than one network slice of one operator.
- Support the adaptation of capacity, i.e. elasticity of capacity of a network slice.
- Ensure that elasticity of capacity of one slice has no impact on services provided by other slices.
- Support means by which the operator can add and remove network functions to the network such that they can be in the network slice.
- Support a mechanism to assign a device to a network slice with the needed services and authorization or to a default network slice.

- Traffic and services in one network slice shall have no impact on traffic and services other network slices in the same network.
- Creation, modification and deletion of a network slice shall have no or minimal impact on traffic and services in same network.

Moreover, as defined by 3GPP in [14], a network slice always consists of a RAN part and a CN part. The support of network slicing relies on the principle that traffic for different slices is handled by different protocol data unit (PDU) sessions. Network can realise the different network slices by scheduling and also by providing different L1/L2 configurations. The UE provides assistance information for network slice selection in RRC message, if it has been provided by NAS. While the network can support large number of slices (hundreds), the UE need not support more than 8 slices simultaneously. Network slicing is a concept to allow differentiated treatment depending on each customer requirements. With slicing, it is possible for Mobile Network Operators (MNO) to consider customers as belonging to different tenant types with each having different service requirements that govern in terms of what slice types each tenant is eligible to use based on Service Level Agreement (SLA) and subscriptions. Furthermore the following aspects on slice availability and support of multiple slices are being defined by 3GPP in [14]:

- Some slices may be available only in part of the network. Awareness in the NG-RAN of the slices supported in the cells of its neighbours may be beneficial for inter-frequency mobility in connected mode. It is assumed that the slice availability does not change within the UE's registration area.
- The NG-RAN and the core are responsible to handle a service request for a slice that may or may not be available in a given area. Admission or rejection of access to a slice may depend by factors such as support for the slice, availability of resources, support of the requested service by NG-RAN.
- In case a UE is associated with *multiple network slices simultaneously* only one signalling connection is maintained and for intra-frequency cell reselection, the UE always tries to camp on the best cell. For inter-frequency cell reselection, dedicated priorities can be used to control the frequency on which the UE camps.
- Slice awareness in NG-RAN is introduced at PDU session level, by indicating the S-NSSAI corresponding to the PDU Session, in all signalling containing PDU session resource information.
- It is the responsibility of the 5GC to validate that the UE has the rights to access a network slice. Prior to receiving the Initial Context Setup Request message, the NG-RAN may be allowed to apply some provisional/local policies, based on awareness of which slice the UE is requesting access to. During the initial context setup, the NG-RAN is informed of the slice for which resources are being requested.
- In addition, charging issues are discussed in [18], where aspects of charging for UE served by one or more Network Slice instances are elaborated including offline and online charging.

5.1.2 Architectural aspects

Architectural aspects have been considered in [7] and [8] where entities such as slice tenant, slice provider, slice manager are provided. In this sense, a *network slicing provider* (NSP), typically a telecommunication service provider, is the owner or tenant of the network infrastructures from which network slices are created. Also, a *network slice endpoint* (NSE) is a network-slice-aware terminal, typically subscribed to the service which is hosted in a network slice instance. A network slice endpoint may be capable of subscribing to multiple services hosted independently in different network slice instance simultaneously. In addition, a *network slice tenant* (NST) is the user of specific network slice instances, in which specific services are hosted and can be provided to NSEs. Network slice tenants can make requests of the creation of new network slice instances. Certain level of management capability should be exposed to network slice tenant from network slice service provider by pre-allocated outsource management entities. Furthermore, a *network function* (NF) is a processing function in a network. It includes but is not limited to network nodes functionality, e.g. session management, mobility management, switching, routing functions, which has defined functional behaviour and interfaces. Network functions can be implemented as a network node on a dedicated hardware or as a virtualized software functions. Data, Control, Management, Orchestration planes functions are Network Functions. Also, a *virtual network function* (VNF) is defined as a network function whose functional software is decoupled from hardware. One or more virtual machines running different software and processes on top of industry-standard high-volume servers, switches and storage, or cloud computing infrastructure, and capable of implementing network functions traditionally implemented via custom hardware appliances and middle-boxes (e.g. router, NAT, firewall, etc.). *Figure 5-3* illustrates the aforementioned entities and functions.

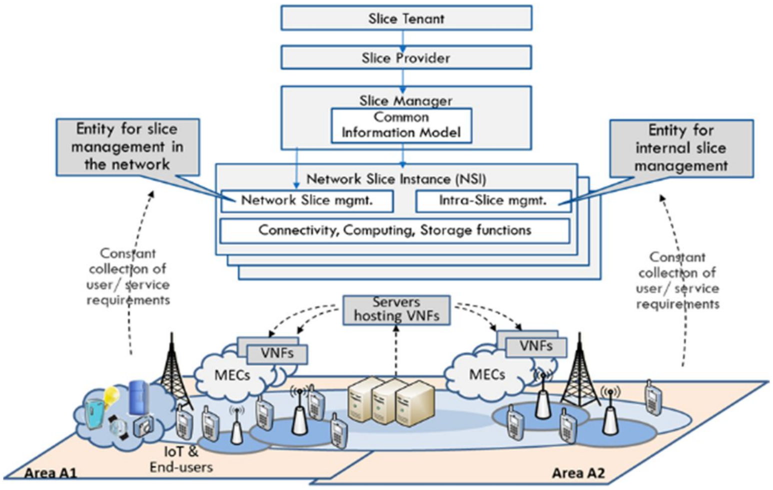


Figure 5-3: Architectural aspects.

5.1.3 Network functions

The section that follows provides an overview of 5G system architecture which consists of the following network functions (NF)[21]:

1. Authentication Server Function (AUSF)
    - Supporting authentication services.
2. Access and Mobility Management Function (AMF)
    - AMF supports functionality such as Registration management; Connection management; Reachability management; Mobility Management; Access Authentication and Authorization etc.
3. Unstructured Data Storage Function (UDSF)
    - The UDSF is an optional function that supports the functionality of storage and retrieval of information as unstructured data etc.
4. Network Exposure Function (NEF)
    - NEF supports the following independent functionality: Exposure of capabilities and events; Secure provision of information from external application to 3GPP network; Translation of internal-external information etc.

5. NF Repository Function (NRF)
    - NRF supports the functionality of service discovery; Maintains the NF profile of available NF instances and their supported services etc.
6. Network Slice Selection Function (NSSF)
    - It is used for Selecting the set of Network Slice instances serving the UE; Determining the Allowed NSSAI and, if needed, the mapping to the Subscribed S-NSSAIs; Determining the AMF Set to be used to serve the UE, or, based on configuration, a list of candidate AMF(s), possibly by querying the NRF etc.
7. Policy Control Function (PCF)
    - Supports unified policy framework to govern network behaviour; Provides policy rules to Control Plane function(s) to enforce them; Accesses subscription information relevant for policy decisions in a Unified Data Repository (UDR) etc.
8. Session Management Function (SMF)
    - Some or all of the SMF functionalities may be supported in a single instance of a SMF including Session Management e.g. Session establishment, modify and release, including tunnel maintain between UPF and AN node; UE IP address allocation & management; DHCPv4 (server and client) and DHCPv6 (server and client) functions; Charging data collection and support of charging interfaces; Downlink Data Notification etc.
9. Unified Data Management (UDM)
    - User Identification Handling (e.g. storage and management of SUPI for each subscriber in the 5G system); Access authorization based on subscription data (e.g. roaming restrictions); UE's Serving NF Registration Management etc.
10. Unified Data Repository (UDR)
    - Supporting functionality of storage and retrieval of subscription data by the UDM; Storage and retrieval of policy data by the PCF; Storage and retrieval of structured data for exposure, and application data (including Packet Flow Descriptions (PFDs) for application detection, application request information for multiple UEs), by the NEF etc.
11. User Plane Function (UPF)
    - Supporting functionality such as anchor point for Intra-/Inter-RAT mobility (when applicable); external PDU Session point of interconnect to Data Network; packet routing & forwarding (e.g. support of Uplink classifier to route traffic flows to an instance of a data network, support of Branching point to support multi-homed PDU session); packet inspection (e.g. Application detection based on service data flow template and the optional PFDs received from the SMF in addition) etc.
12. Application Function (AF)
    - The Application Function (AF) interacts with the 3GPP Core Network in order to provide services, for example to support application influence on traffic routing; accessing Network Exposure Function; interacting with the Policy framework for policy control; based on operator deployment, application Functions considered to be trusted by the operator can be allowed to interact directly with relevant Network Functions.
13. 5G-Equipment Identity Register (5G-EIR)
    - Supports the check of the status of PEI (e.g. to check that it has not been blacklisted)
14. Security Edge Protection Proxy (SEPP)
    - It is a non-transparent proxy and supports the functionality of message filtering and policing on inter-PLMN control plane interfaces and topology hiding.
15. Location Management Function (LMF)
    - The LMF functionality supports location determination for a UE; obtains downlink location measurements or a location estimate from the UE; obtains uplink location

measurements from the NG RAN; obtains non-UE associated assistance data from the NG RAN.

16. Non-3GPP InterWorking Function (N3IWF)
    - Support of IPsec tunnel establishment with the UE: The N3IWF terminates the IKEv2/IPsec protocols with the UE over NWu and relays over N2 the information needed to authenticate the UE and authorize its access to the 5G Core Network; termination of N2 and N3 interfaces to 5G Core Network for control - plane and user-plane respectively; relaying uplink and downlink control-plane NAS (N1) signalling between the UE and AMF; handling of N2 signalling from SMF (relayed by AMF) related to PDU Sessions and QoS; establishment of IPsec Security Association (IPsec SA) to support PDU Session traffic.
17. SMS Function (SMSF)
    - The SMSF supports the following functionality to support SMS over NAS; SMS subscription checking; SM-RP/SM-CP with the UE.
18. Network Data Analytics Function (NWDAF) [22]
    - NWDAF represents operator managed network analytics logical function. NWDAF provides slice specific network data analytics to the PCF and NSSF. NWDAF provides network data analytics (i.e., load level information) to PCF and NSSF on a network slice level. NSSF may use the load level information provided by NWDAF for slice selection.

5.1.4 Lifecycle of network slice instances

The provisioning of network slicing includes the four phases which are preparation, commissioning, operation and decommissioning [20]:

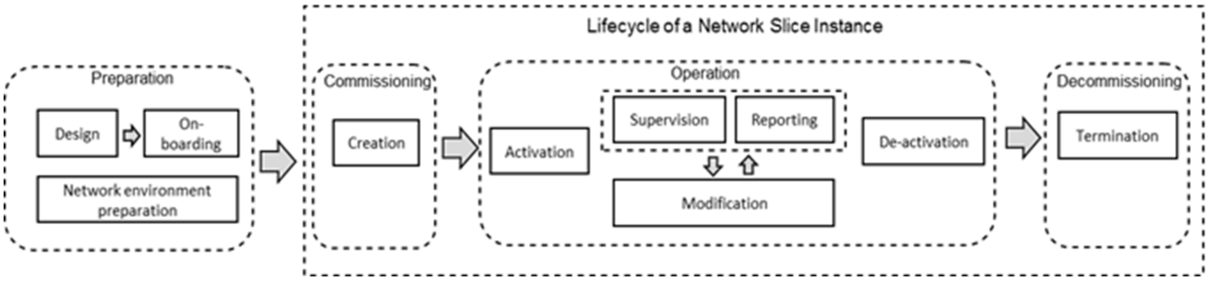


Figure 5-4: Lifecycle of Network Slice Instance [19].

- In the preparation phase the network slice instance (NSI) does not exist. The preparation phase includes network slice design, on-boarding, evaluation of the network slice requirements, preparing the network environment and other necessary preparations required to be done before the creation of an NSI.
- During the NSI lifecycle stage which include commissioning phase, operation phase and decommissioning phase, the NSI provisioning operations include:
    - Create an NSI;
    - Activate an NSI;
    - De-active an NSI;
    - Modify an NSI;
    - Terminate an NSI.

The operations of the provisioning of an NSI occur during different phases of a NSI:

- During the commissioning phase;
    - Create an NSI.
    - During NSI creation all resources to the NSI have been created and configured to satisfy the network slice requirements. NSI creation may trigger NSSI(s) creation or using existing NSSI(s) and setting up the corresponding associations.

- During the operation phase:
    - Activate an NSI;
    - Modify an NSI;
    - De-active an NSI.

NSI activation includes any actions that make the NSI active to provide communication services. NSI activation may trigger NSSI activation. NSI modification in operation phase could map to several workflows, e.g. changes of NSI capacity, changes of NSI topology, NSI reconfiguration. NSI modification can be triggered by receiving new network slice related requirements, new communication service requirements, or the result of NSI supervision automatically. NSI modification may trigger NSSI modification. The NSI deactivation operation may be needed before NSI modification operation and the NSI activation operation may be needed after the NSI modification operation. NSI deactivation includes any actions that make the NSI inactive and not providing any communication services. NSI deactivation trigger NSSI deactivation to deactivate constituent NSSI(s) which is not used by other NSI(s). Operator may decide to keep the NSI without termination after deactivation and reactivate it when receives new communication service request.

- During the decommissioning phase
    - Terminate an NSI.
    - NSI termination step includes any action that make the NSI doesn't exist anymore and release resources that are not used by other NSI(s). NSI termination may trigger NSSI termination to terminate constituent NSSI(s) which is not used by other NSI(s).

We should note that the network slicing requirements illustrate a need for centralized management and orchestration of various network slice instances. The existing Element Managers (EM), Network Manager (NM), and OSS/BSS do not have this capability. The figure that follows [17] presents a network slicing management and orchestration architecture, extending the 3GPP management reference framework with the required capabilities.

It is shown that for the realization of the missing functionalities while ensuring interworking with the legacy operations and management systems, entities in the big rectangle box are logically placed between the 3GPP management entities and the network infrastructure resource domains. It is proposed that entities such as network slice manager and other functions are added to an extended 3GPP management framework.

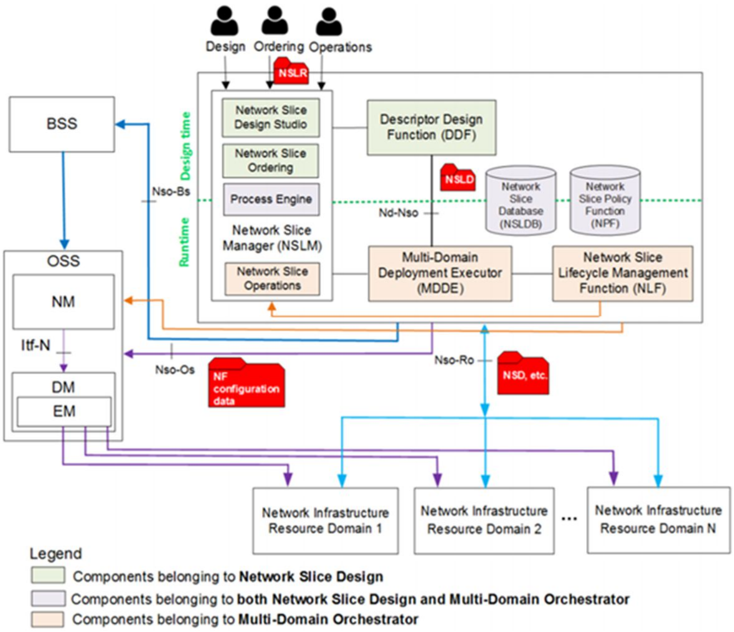


Figure 5-5: Proposed Extensions of 3GPP Management Framework [17].

5.1.5 Considered use cases

Overall, the 5G community is considering the implementation of slicing concept for facilitating various services such as mMTC, eMBB, V2X (e.g. V2V/V2I). In this context various types of devices may be accommodated by slices such as devices with very low throughput, long sleep cycles (and latency), non-critical (e.g. mMTC), smartphones for demanding traffic (e.g. HD/4K video etc. – eMBB), other connected devices with continuous control and communication (e.g. URLLC), closed-loop communication systems etc.

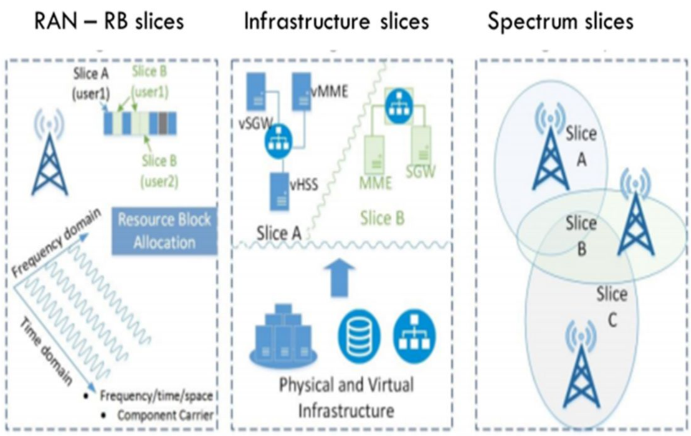


Figure 5-6: RB, infrastructure, spectrum slices as proposed in [9]

However, in this particular study we are focusing primarily on vehicle-to-infrastructure (V2I) related use cases. In V2I use cases, mobility of users is a key challenge which needs to be handled accordingly in order to ensure proper usage of slicing resources while users are moving from place A to place B. In addition, V2I use cases can involve both eMBB and URLLC services. eMBB services can involve on-board infotainment and static or walking users in the street sidewalks. URLLC services can involve aspects related to automated driving, driving safety etc. Also, 3GPP in [10] has defined some specific scenarios for slice (re-)selection such as:

- *Slice (re-)selection due to mobility:* This scenario can happen when the UE moves into different service area which has connections with different network slice set. This means that the serving access network (AN) is changed and the target AN doesn't support all deployed slices of the network.
- *Slice reselection due to the network maintenance issue:* A node consisting a network slice fails for some reason (e.g. node failure, network congestion) and the network needs to change the serving node for the UE. For example, if a serving entity has network congestion, the network may trigger the network slice reselection for guaranteeing certain level of QoS.
- *UE requested service change:* The UE may request to change service that the UE is receiving, which may lead to the network slice reselection. This scenario can happen when the UE supports multiple services and the services cannot be served without change of network slice.

As previously mentioned in this study we focus on V2I scenarios combining eMBB and URLLC (such as road safety services). Such a scenario shall take into account:

- eMBB traffic (in car);
- eMBB traffic (static users in the side street);
- V2I-related service for road safety (URLLC);
- eMBB service can be served by a dedicated slice (e.g. more strict requirements related to throughput);
- eMBB can be provided from in car or static users in the side street URLLC service can be served by another slice (e.g. more strict requirements related to latency).

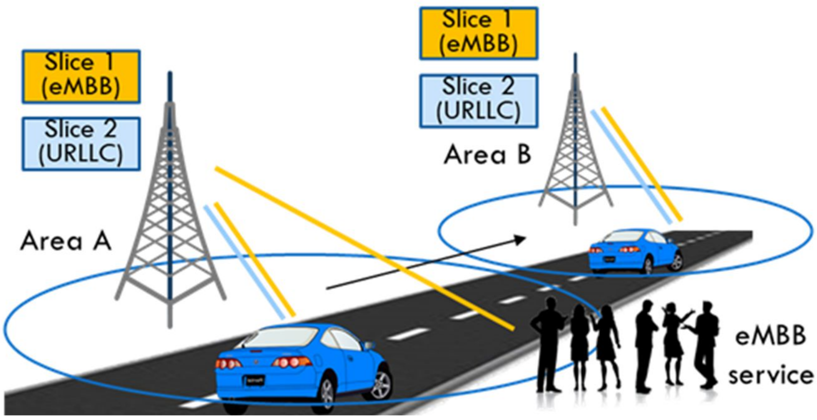


Figure 5-7: Considered V2I scenarios combining eMBB and URLLC slices

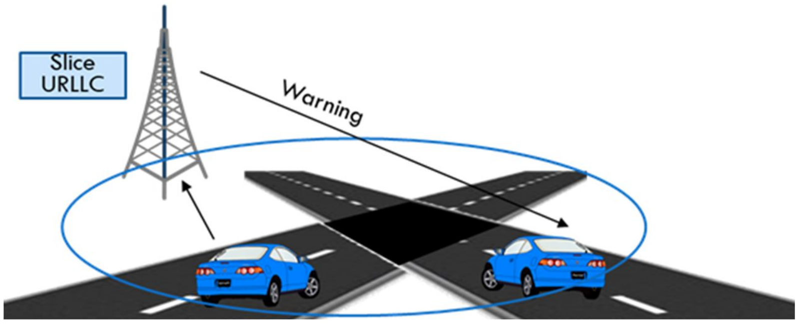


Figure 5-8: Road safety services as proposed by 3GPP TR 22.885 [11].

Also, road safety services have been proposed by 3GPP in [11] and consider the following indicative interaction between various entities:

- Vehicle A and Vehicle B's UEs determine a cell which is controlled by an BS.
- Vehicle A's UE triggers transmission of a V2I message periodically or based on a certain event that happens at the vehicle, e.g. collision risk warning.
- The BS receives one or more V2I messages from one or more vehicles including Vehicle A's UE.
- The BS may or may not filter out some V2I messages received from some UEs.
- The BS distributes a V2I message at the cell.
- Vehicle B's UE monitors transmission of the V2I messages and receives the V2I message at the cell.

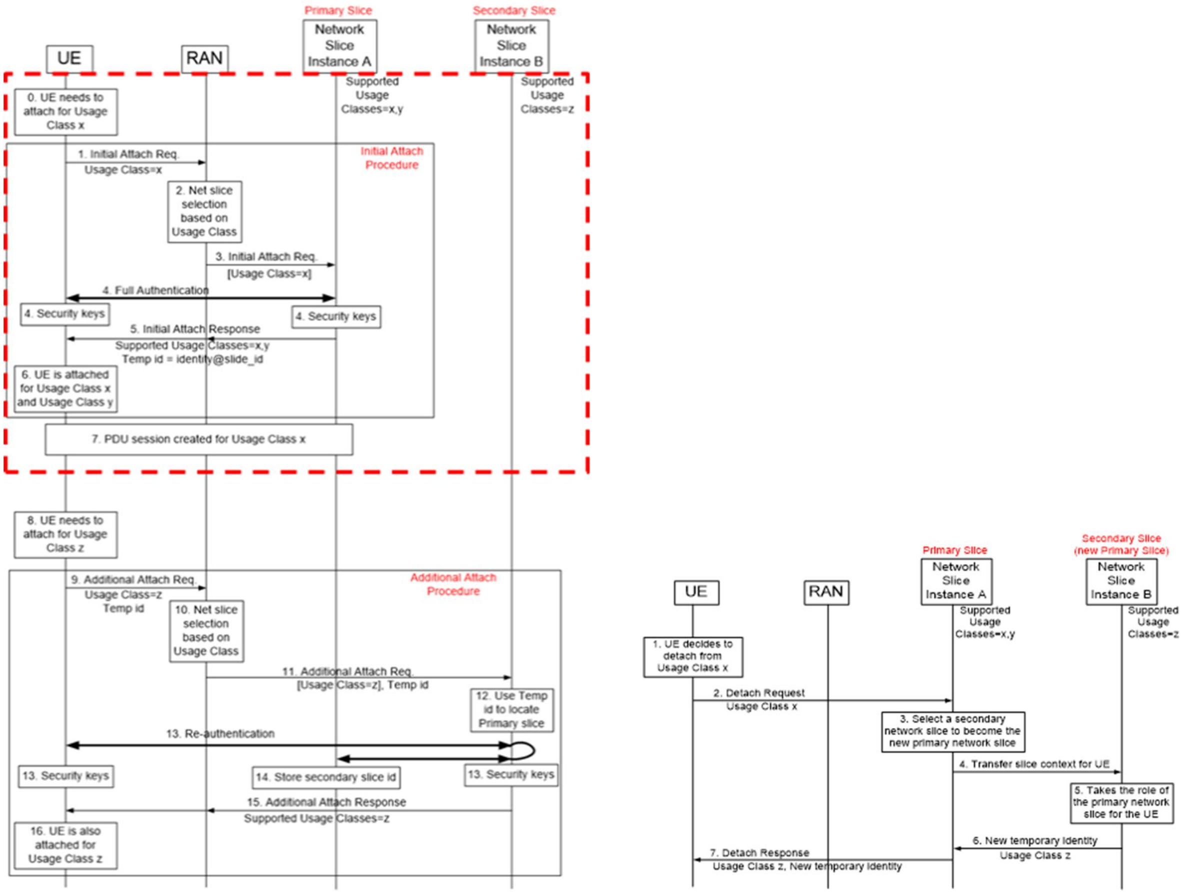
Specific test cases for simulation and demonstration, with particular focus on V2I and respective challenges, are also under consideration as outlined in the paragraphs that follow.

- Certain test cases could be considered in order to check potential *handover issues* in environments with mobility.
    - Assuming a case in which a BS that initially does not have any vehicles (or small number) served and at a later time there is a traffic jam in the area of the specific BS. Hence, increased network traffic is expected and appropriate handling of increased traffic from the system should be done.
    - A number of BS that do not have the required slices and need to create them which would mean re-arrangement of resources to existing slices, changes to resources, etc.
    - A number of BS that do not have the required slices but they are unable to create new slices (specific ones) and thus the traffic should be steered to another BS that has the available capacity and can support the required slices.
- Moreover, some challenges related to URLLC-related cases can be:
    - The so called problem of "Hidden node" where a node is not visible from another node, either due to an out of range or obstacle existence between the two nodes. An increase of transmit power of mobile nodes may not work if, for example, the reason of the hidden node is that there is a concrete or steel wall preventing communication with other nodes, such as buildings or even big trucks that exist on roads. There are also several software implementations of additional protocols that essentially implement a polling or token passing strategy. This may eliminate the hidden node problem, but at the cost of increased latency and less maximum throughput that creates problems to our URLLC use case where latency is of high criticality.
    - Roads without traffic lights or signs where the vehicles exchange information of their speed, direction of movement, next movement decisions etc. in order to coordinate at

junctions and create a seamless movement of traffic which could potentially be accident free.

- Also, challenges with respect to eMBB-related cases can be that a specific eMBB slice could provide to vehicles service with very high throughput in order to stream high quality services e.g. video for supporting cases when vehicles entering areas that have low or non-existence signal like tunnels, underground roads, between high-rise buildings etc.

The message sequence charts that follow (*Figure 5-9*), provide an indication of UE slice attach and detach.



(a) (b)

Figure 5-9: UE Slice Attach and Detach sequences [15].

Also, *Figure 5-10* illustrates the serving of eMBB (e.g. video) and URLLC (e.g. V2X/V2I communications) traffic services by different slices in the RAN and core and also as defined in 3GPP it introduces a “Network slice selection function”. A slice selection function is a key element in the future core network architecture, enabling the UE to be allocated to a proper slice. Furthermore, as it is shown in the figure, the eMBB service is served by a different slice compared to URLLC. In the problem statement and formulation that follows in this deliverable, we’ll show the allocation of available resources in network slices, in order to serve in the best possible manner the needed services. Due to the importance and criticality of URLLC (e.g. related to road safety, automated driving etc.), dedicated resources of macro cells will be allocated to such a slice, while the remaining resources will be assigned to other slices served by other macro and small cells. As a result, the formulation and problem solving will ensure that critical services are somehow prioritized (in terms of resource assignment) since service degradation is not easily tolerable for such services.

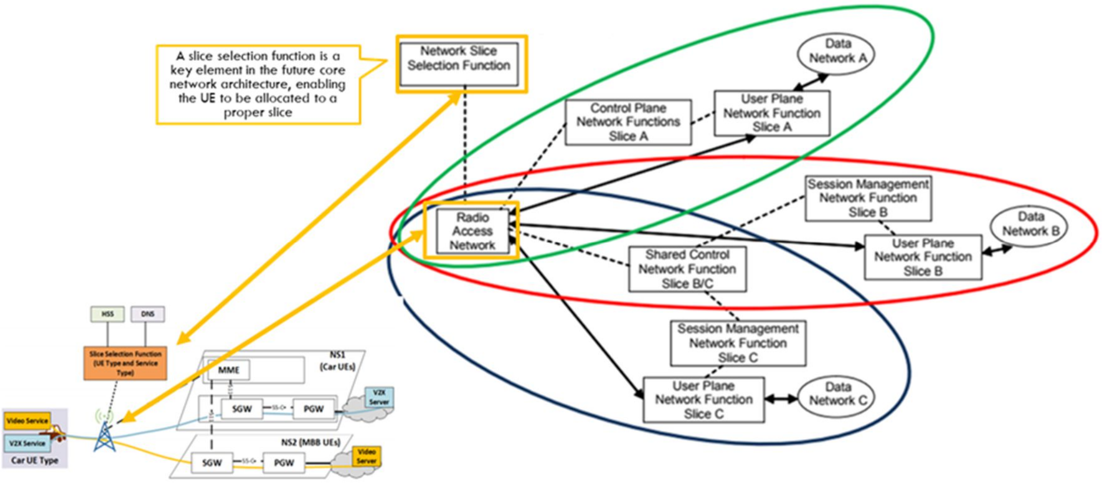


Figure 5-10: Slicing operations: selection based on usage class [20][21].

5.2 Problem statement and formulation

The problem statement and formulation which is presented in this subchapter, tries to solve the problem of the maximization of the bandwidth that is assigned – corresponding to quality levels and different weights for diversifying the importance of components.

The following inputs have been defined:

- Set of macro cells MC
- For each macro $k \in$ MC there is capacity M_k
- Set of small cells SC
- For each small cell $l \in$ SC there is capacity SC_l
- Set of cars/vehicles C
- For each car/vehicle $i \in$ C we have a requirement for eMBB traffic, $Cr_{i,eMBB}$, and for URLLC traffic, $Cr_{i,URLLC}$
- Set of pedestrians PED
- For each $j \in$ PED we have pedestrian requirement $pr_{j,eMBB}$

Rules - Policies

- Car with URLLC service assigned to macro (dedicated resources)
- Car with eMBB service assigned to macro or small (residual or shared resources)
- Pedestrian with eMBB service assigned to macro or small (residual or shared resources)

Decision variables, related to car and respective services

- X_{ik} : which macro serves the car/URLLC traffic; whether the URLLC traffic of car-i is served by macro-k
- Y_{ik} : which macro serves the car/eMBB traffic; whether the eMBB traffic of car-i is served by macro-k

- Z_{il} : which small serves the car/eMBB traffic; whether the eMBB traffic of car-i is served by small cell-l

Decision variables, related to pedestrians and respective services

- A_{jk} : which macro serves the pedestrian/eMBB traffic; whether the eMBB traffic of pedestrian-j is served by macro-k
- B_{jl} : which small serves the pedestrian/eMBB traffic; whether the eMBB traffic of pedestrian-j is served by small cell-l

Objective function

$$\max w_1 \cdot \sum_{i\in C}\sum_{k\in MC} X_{ik} \cdot Cr_{i,URLLC} + w_2 \cdot \sum_{i\in C}\sum_{k\in MC} Y_{ik} \cdot Cr_{i,eMBB} + w_3 \cdot \sum_{i\in C}\sum_{l\in SC} Z_{il} \cdot Cr_{i,eMBB} + w_4 \cdot \sum_{j\in PED}\sum_{k\in MC} A_{jk} \cdot pr_{j,eMBB} + w_5 \cdot \sum_{j\in PED}\sum_{l\in SC} B_{jl} \cdot pr_{j,eMBB}$$

where $w_1 > w_2 > w_3 > w_4 > w_5$

- Maximization of the bandwidth that is assigned – corresponding to quality levels
- Different weights for diversifying the importance of components

Explanation of objective function

$\sum_{i\in C}\sum_{k\in MC} X_{ik} \cdot Cr_{i,URLLC}$	Sum of each macro *k* ∈ MC serving each car *i* ∈ C which requests URLLC traffic (requirement $Cr_{i,URLLC}$)
$\sum_{i\in C}\sum_{k\in MC} Y_{ik} \cdot Cr_{i,eMBB}$	Sum of each macro *k* ∈ MC serving each car *i* ∈ C which requests eMBB traffic (requirement $Cr_{i,eMBB}$)
$\sum_{i\in C}\sum_{l\in SC} Z_{il} \cdot Cr_{i,eMBB}$	Sum of each small *l* ∈ SC serving each car *i* ∈ C which requests eMBB traffic (requirement $Cr_{i,eMBB}$)
$\sum_{j\in PED}\sum_{k\in MC} A_{jk} \cdot pr_{j,eMBB}$	Sum of each macro *k* ∈ MC serving each pedestrian *j* ∈ PED which requests eMBB traffic (requirement $pr_{j,eMBB}$)
$\sum_{j\in PED}\sum_{l\in SC} B_{jl} \cdot pr_{j,eMBB}$	Sum of each small *l* ∈ SC serving each pedestrian *j* ∈ PED which requests eMBB traffic (requirement $pr_{j,eMBB}$)

Constraints

$\sum_i X_{ik} \cdot Cr_{i,URLLC} \leq M_k$	Constraint which shows that capacity of macro cell k should be respected; total URLLC traffic assigned to macro *k* should be below the capacity M_k
$res_k = M_k - \sum_i X_{ik} \cdot Cr_{i,URLLC}$	Constraint which shows the residual resources of macro cell *k*
$\sum_k X_{ik} = 1$	Constraint which shows that car/URLLC traffic is definitely served; which URLLC traffic *i* is served by macro cell *k*
$\sum_k Y_{ik} + \sum_l Z_{il} \leq 1$	Constraint which shows that car/eMBB traffic is allocated either in macro, either small or nowhere;

if possible eMBB traffic of cars should be served

$$\sum_k A_{jk} + \sum_l B_{jl} \leq 1$$

Constraint which shows that pedestrian/eMBB traffic is allocated either in macro, either small or nowhere; if possible eMBB traffic of pedestrians should be served

$$\sum_i Z_{il} \cdot Cr_{i,eMBB} + \sum_j B_{jl} \cdot pr_{j,eMBB} \leq SC_l$$

Constraint which shows that capacity of small cell *l* should be respected

$$\sum_i Y_{ik} \cdot Cr_{i,eMBB} + \sum_j A_{jk} \cdot pr_{j,eMBB} \leq res_k$$

Constraint which shows that total eMBB traffic should not exceed the residual resources; total eMBB traffic assigned to macro *k* should be below the capacity res_k

5.3 Considered scenario and solution algorithm

For the purposes of the study a scenario comprising moving vehicles and users has been considered. Through such a scenario it is possible to show the impact of decisions to both URLLC and eMBB slices due to the fact that for vehicle commands, URLLC slices are of ultimate importance, while for passengers/pedestrians the usage of eMBB slices are utilized for transmission of content etc. Allocation of resources is initially prioritized for URLLC slice and the remaining resources are allocated to eMBB slices which may serve video traffic, browsing etc.

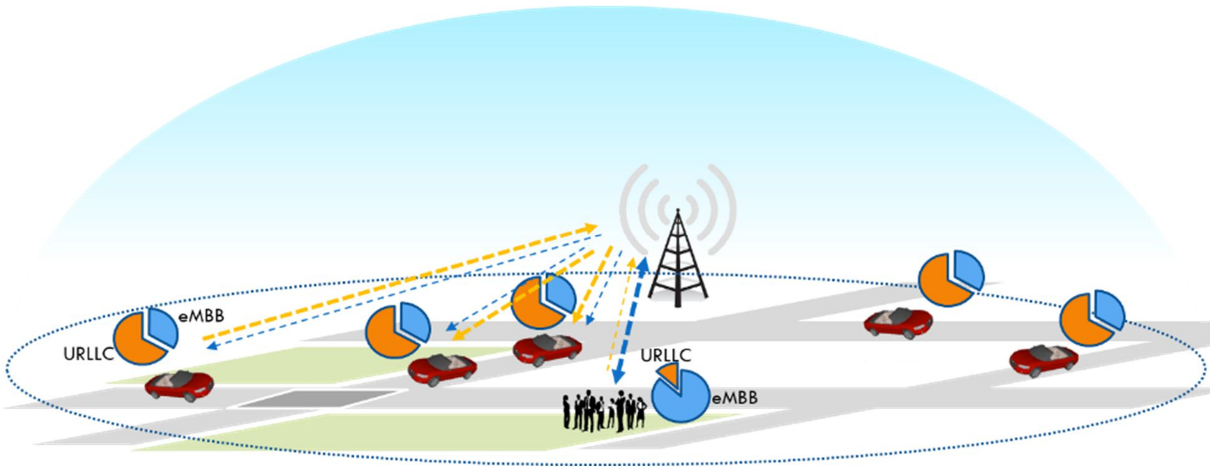


Figure 5-11: Considered scenario overview.

Moreover, the scenario, for proof-of-concept purposes and evaluation of resource allocation decisions made, takes into account the following phase of execution:

- *Initialization:* initial allocation of resources is made in URLLC and eMBB slices for accommodating traffic demand.
- *URLLC traffic increase:* An increase is triggered in order to assess that URLLC latency can respect the SLA constrains despite the fact that a significant amount of URLLC sessions are requested, due to preemption of resources from other non-critical slices.
- *eMBB traffic increase:* URLLC sessions are reduced to initial lower level, while eMBB traffic is increasing. In this case, either the eMBB is degraded or more resources will be made available to slices in order to avoid degradation of eMBB (or other slices).

5.3.1 Dynamic Resource Allocation

Based on the scenario that is already presented the dynamic resource allocation between the slices has been taken into account. For example at a football stadium Figure 5-12 where many people are gathered to see the match they intensively use eMBB application for uploading photos or watching videos from social media. At the other hand a limited use of URLLC is present at the same time, so the slices will be configure in a way to provide the quality that is needed in this situation.

Figure 5-12: eMBB Traffic.

In highway scenarios or in cases where traffic is high at a specific area Figure 5-13, multiple vehicles will be served by a cell utilizing mainly URLLC services. In cases of platooning and/or high traffic conditions critical messages about road, car and traffic status will be exchanged. Also, eMBB services (e.g. video watched by passengers) can be provided by respective slice.

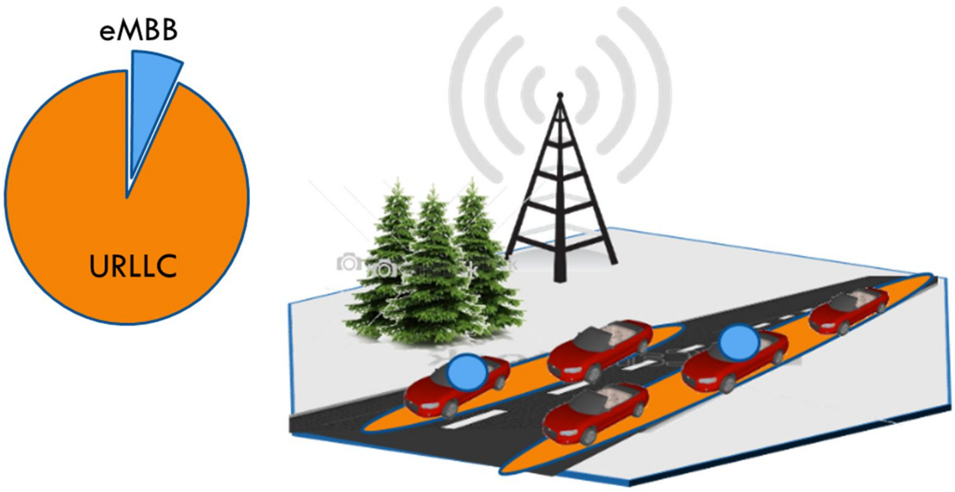


Figure 5-13: URLLC Traffic

The resources can change over time to handle different scenarios and situations whenever a specific traffic is in demand. The system will always have a minimum resource allocated to one of those two slices that we have in our scenario. As it can be seen at Figure 5-14 the resources can change from minimum to maximum in either way, but always the URLLC slice has the priority for acquiring the resources that it needs. Also, each network slice has no or minimal impact on traffic and services to the other slice at the same network.

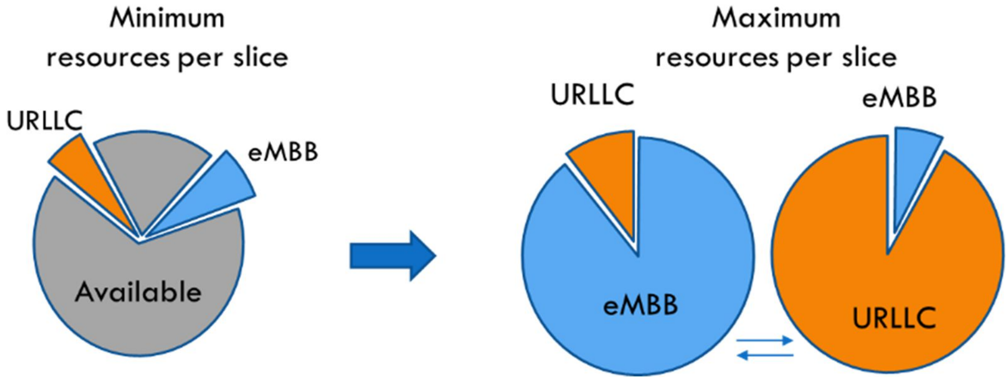


Figure 5-14: Resources per slice.

5.4 Solution algorithm

In addition, the following algorithm, which is represented in the flowchart of Figure 5-15 has been created in order to solve the problem of resource allocation to slices.

Input:

Set of macro cells MC where for each macro $k \in$ MC there is capacity M_k; set of small cells SC where for each small cell $l \in$ SC there is capacity SC_l; set of vehicles C where for each vehicle $i \in$ C we have a requirement for eMBB traffic, $Cr_{i,eMBB}$, and for URLLC traffic, $Cr_{i,URLLC}$; set of pedestrians PED where for each $j \in$ PED we have pedestrian requirement $pr_{j,eMBB}$; set of slices SL and a set of resources RB open to various optimization algorithms.

Process:

Each time a new event request is generated, the algorithm checks whether resources are available or not. If resources are available, then they are assigned to the request and the algorithm reiterates for the next event. If are not available, then the algorithm will select the slice from which preemption of resources can happen (selections should minimize the impact on SLAs, in conjunction with the achievement of fairness) and if this is still not possible, then an allocation re-computation will be made for redistributing “existing” resources to slices (e.g., through reinforcement learning) or obtain additional resources.

Output:

Each resource $rb \in$ RB is allocated to different slices $\{sl_x, sl_y, \ldots\} \in$ SL.

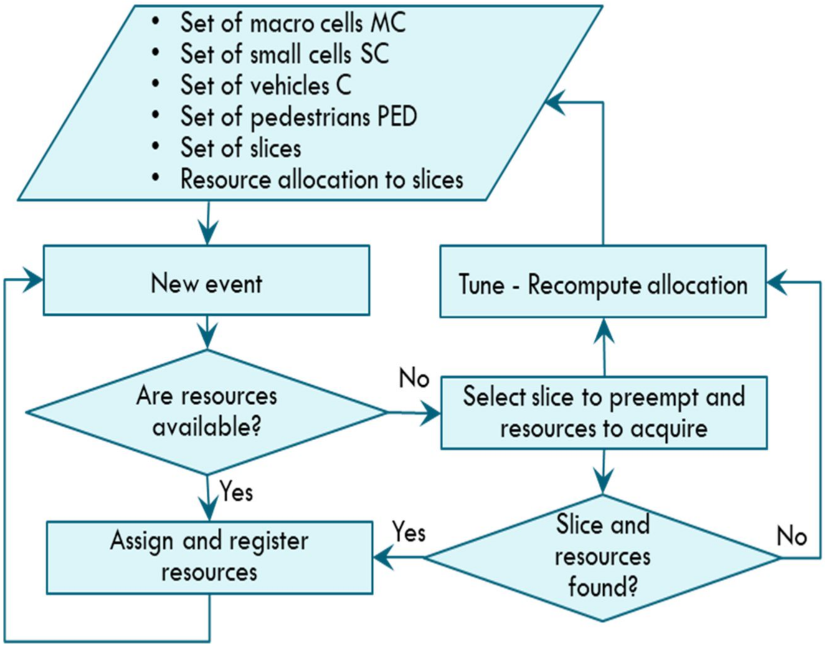


Figure 5-15: Algorithm flowchart.

5.5 Implementation aspects

5G is poised to support a set of ambitious use cases. For instance, use case families in NGMN include broadband access in dense areas and everywhere (eMBB), massive Internet of Things and machine-type communications (mMTC) as well as ultra-reliable communications (URLLC). For the needed representation/ modeling of such aspects, environment models shall take into account area aspects, traffic, mobility and propagation models depending on the considered area (e.g. urban etc.). 3GPP has defined various traffic models [3][4].

Traffic models:
In this study the focus in on URLLC and eMBB traffic. For URLLC the following assumptions are used [2]:

- Traffic model based on Beta distribution (α = 3, β= 4 within [0,10] seconds, measured from the time of event occurrence)
- Bursts of small packets (up to 200 bytes) are sent between traffic source and UE. The inter-arrival time of the packets may be not constant. The average bit rate in the long term could be small.
- Packet transmission periodicity: 140 km/h: 100ms; 70 km/h: 200ms

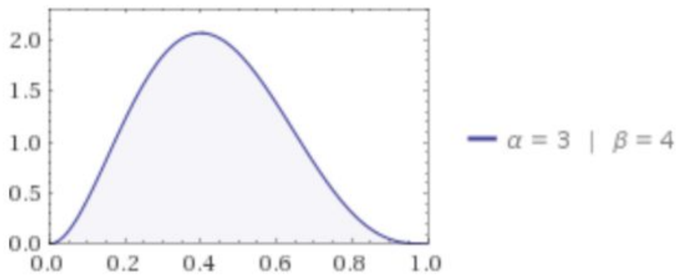


Figure 5-16: PDF for Beta distribution (α=3, β=4).

For eMBB (FTP model 1) the following assumptions are used:

- File size S: 2MB (16Mbit)
- User arrival rate λ=offered traffic/S
- Offered traffic = [2, 4, 6, 8, 10] Mbps
- Possible range of λ: [0.125, 0.25, 0.375, 0.5, 0.625] users/second
- Mean time: [8, 4, 2.6, 2, 1.6]seconds
- Drop file time: 32 seconds

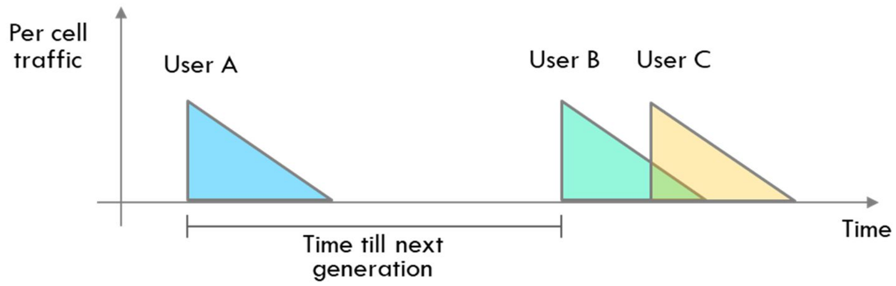


Figure 5-17: Packet generation in FTP model 1.

For eMBB (FTP model 2) the following assumptions are used:

- File size S: 0.5MB (4Mbit)
- Exponential Distribution for file arrival rate λ=0.2 (files/sec)
- Mean time: 5 seconds
- Number of Users, K: [2, 5, 8, 10, 14, ...] per cell
- Drop file time: 8 seconds

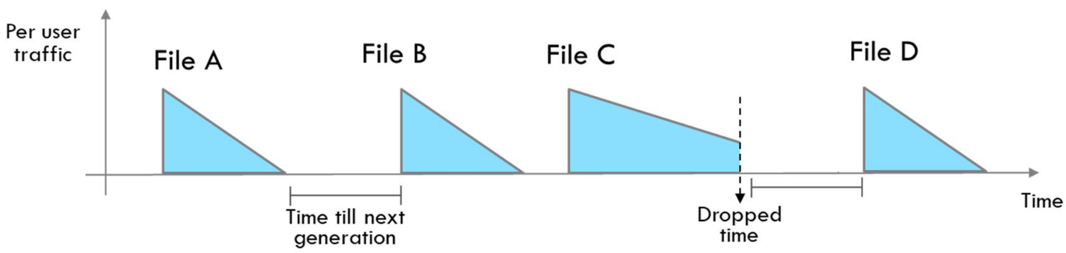


Figure 5-18: Packet generation in FTP model 2.

For eMBB (FTP model 3) the following assumptions are used:

- File size S: 0.5MB (4Mbit)
- Poisson Distribution for file arrival rate with λ =0.2 (files/sec)
- Mean time: 5 seconds
- Number of Users, K: [2, 5, 8, 10, 14, ...] per cell
- Drop file time: 8 seconds

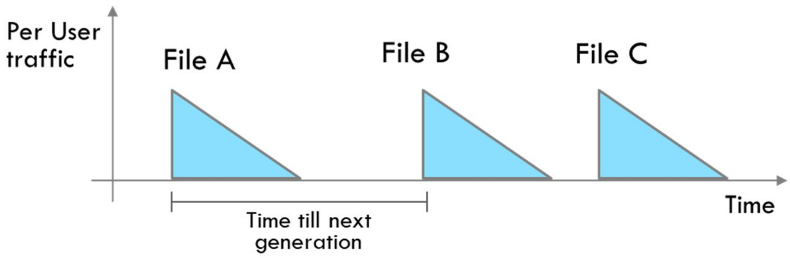


Figure 5-19: Packet generation in FTP model 3.

The table that follows (Table 5-1) summarizes the considered traffic models.

Table 5-1: Traffic models overview.

Service type	Traffic model
eMBB	FTP Model 1 (TR36.814): Poisson distributed with user's arrival rate λ
	FTP Model 2 (TR36.814): Packets arrive exponentially
	FTP Model 3 (TR36.889): Packets for the same UE arrive according to a Poisson process with arrival rate λ
URLLC	TR37.868: Small packets, bursty, Beta distribution (α=3, β=4)

The types of mobility models used:

- *Random Walk:* According to this model each UE changes its speed and direction at each time interval. The default value of time interval is 1s, while these values can be configured from the simulator's graphical interface. For every time interval, the direction is chosen from (0, 2π], while speed follows a uniform or Gaussian distribution from [0, v_{max}].
- *Linear Motion:* Each UE choose randomly a direction and moves along it, with a constant speed. UE continues to move in this direction even if it reaches the cell boundaries (wrap-around model).
- *Random Direction:* Each UE choose randomly a direction and moves along it, with a constant speed, until it reaches the boundary of the cell. Then UE chooses another direction to travel and moves with the same speed until it reaches the boundary of the cell again.

5.6 Evaluation

For performance evaluation of the defined scenarios, system-level simulations were conducted. The system-level simulator has been calibrated according to the 3GPP specifications. The simulator takes into account various parameters such as traffic level, available infrastructure elements, available channels and evaluates the various test cases. The calibration state of the simulator has been checked against the reference results of the 3GPP LTE calibration campaign (3GPP TR 36.814). As a result, the Cumulative Distribution Function (CDF) of coupling loss and downlink SINR have been checked in order to calibrate the tool with leading operators and vendors. The configuration is fully customizable so as to include various types of cells (i.e., macro and small cells). Specifically, it is possible to customize aspects related to the size of simulator playground; the area type (e.g., Dense Urban etc.); the number and position of macro base stations and their inter-site distances (ISDs); the number and position of small cells per macro base station; the number and position of end-user devices in the playground; the mobility of the end-user devices; the number of available channels and much more.

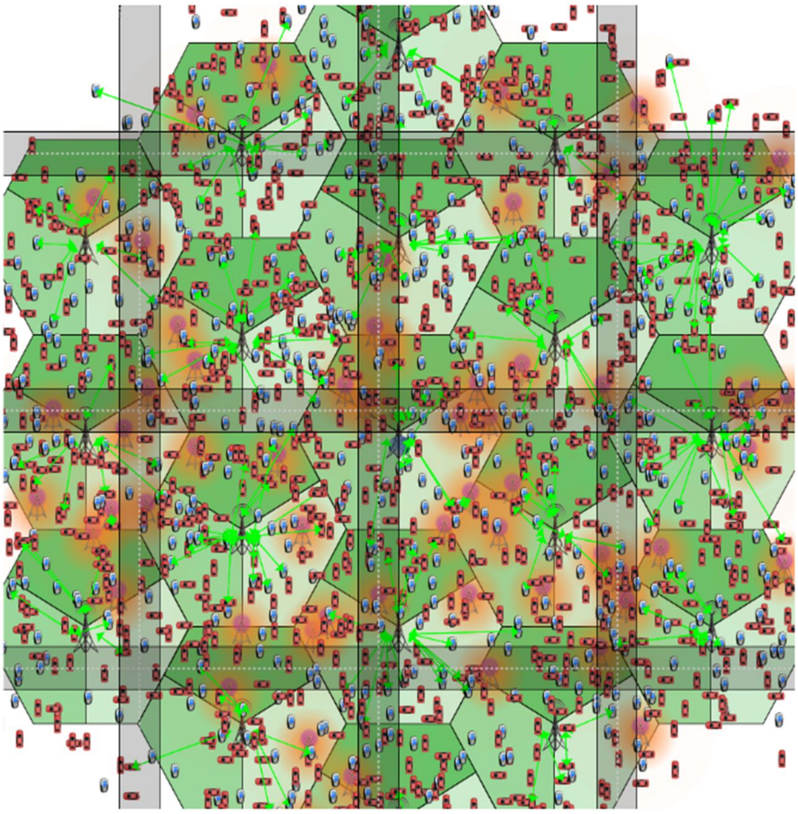


Figure 5-20: Considered topology.

Figure 5-20 demonstrates the topology of the simulation environment that has been developed for the evaluation of our proposed algorithm. The environment consists of macro and small cells shown as hexagonal tiles where vehicles and pedestrians are moving. Additionally, a grid of roads is also seen with 3 vertical and 3 horizontal roads that the vehicles inside these areas are moving in a straight line with freeway speeds. Summarized in Table 5-2 are the experimentation parameters used in our simulation for the performance evaluation of the proposed algorithm.

Table 5-2: Simulation parameters.

Parameter	Value
Macro BSs (with 3 cell each)	19
Inter-site distance	500 m
Small BSs	57
Total RAT Devices	114
Total Number of UEs	1500
Pedestrians & Passengers	500
Vehicles	1000
FTP file sizes	512-2048 Kbytes
URLLC file sizes	20 - 200bytes

Total Resources per cell	8 channels of 10MHz
Number of resource blocks	50 PRBs per channel
Simulation time	60s

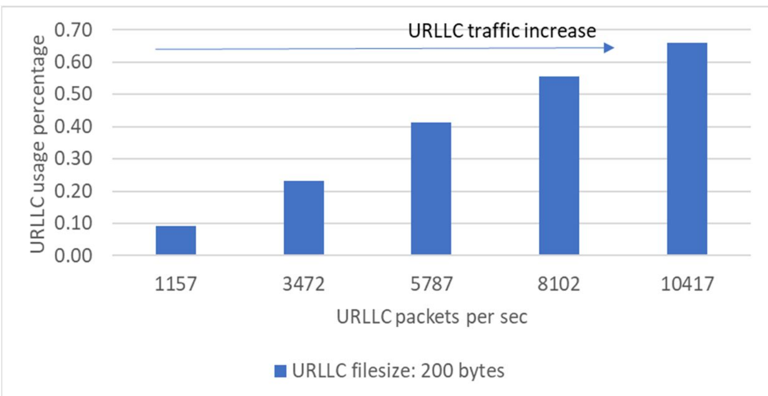
Table 5-3 summarizes the considered evaluation cases. There we can see that different traffic demand for URLLC and eMBB can result to different actions taken from the algorithm. Also the duration of traffic demand is also considered. Usually the algorithms only does some small fixes to the system resources, but when for example the traffic demand is high for both the URLLC and eMBB services then the system needs to go a step further and not only preempt resources or reassign them to different slices, but to acquire extra resources from licensed or even unlicensed frequencies that are available in the area and can be used by the users.

Table 5-3: Considered evaluation cases.

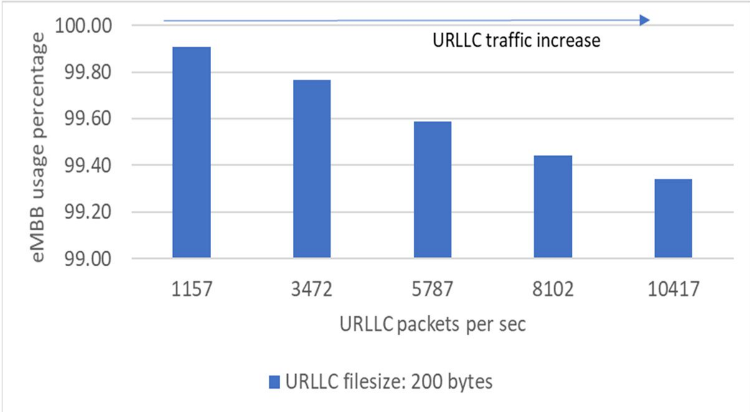
Cases	**URLLC**		**eMBB**	**Actions**
	Traffic Demand	**Duration of Demand**	**Traffic Demand**	
1	Low	Small	Low	None
2	Low	Small	High	Preempt eMBB
3	Low	Large	Low	None
4	Low	Large	High	None
5	High	Small	Low	Preempt URLLC
6	High	Large	Low	Reassign resources
7	High	Large	High	More resources
8	High	Small	High	Preempt eMBB

5.6.1 Results on slicing preemption

The subsection that follows shows the impact of URLLC traffic increase as well as eMBB file size increase. Specifically, Figure 5-21a and b, illustrate the impact of URLLC traffic increase (cases 5 and 6 in Table 5-3) in URLLC and eMBB usage percentage respectively. It is shown that as URLLC traffic increases, there is a small impact to eMBB usage percentage, which on the one hand decreases but on the other hand the decrease is not very large.



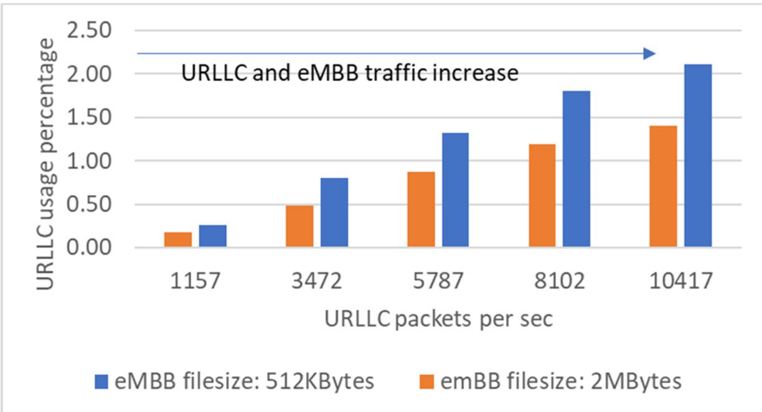
(a)



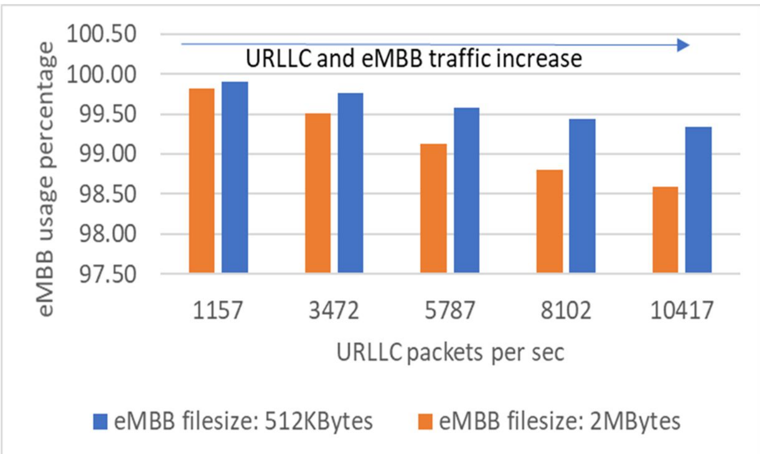
(b)

Figure 5-21: Impact of URLLC traffic increase (cases 5 and 6) in (a) URLLC usage percentage, in (b) eMBB usage percentage.

Similarly, Figure 5-22 a and b, illustrate the impact of URLLC and eMBB traffic increase (cases 7 and 8 in Table 5-3) in URLLC and eMBB usage percentage respectively. It is shown that as URLLC traffic increases, there is also a small impact to eMBB usage percentage. However, the impact is greater when eMBB traffic increases.



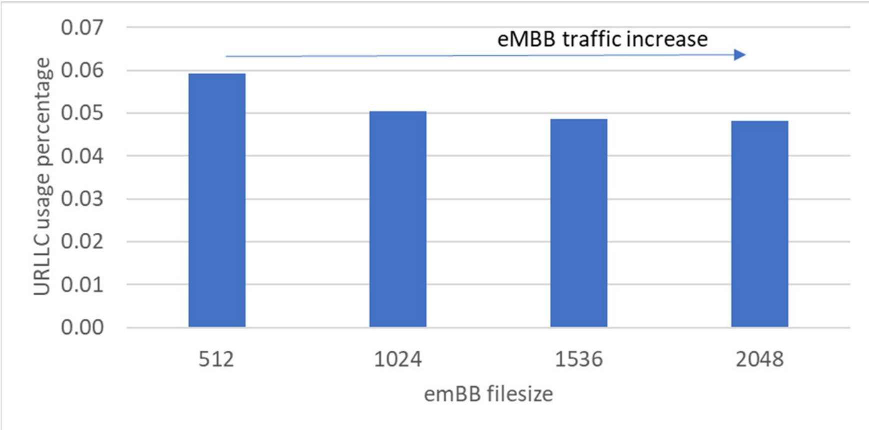
(a)



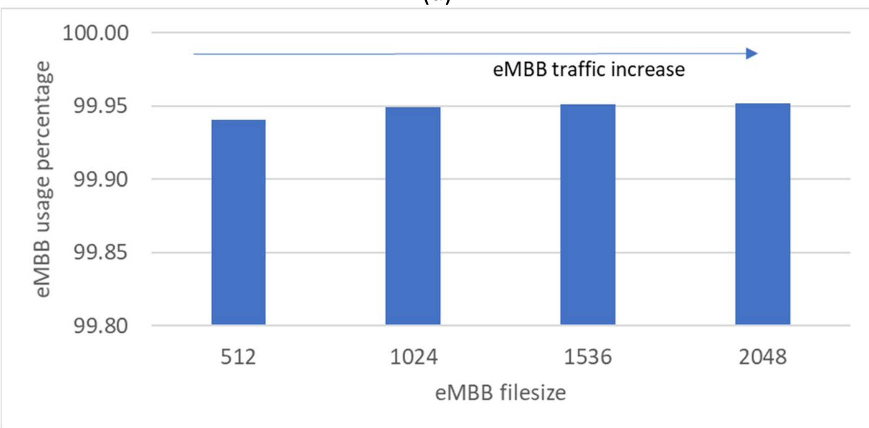
(b)

Figure 5-22: Impact of URLLC and eMBB traffic increase (cases 7 and 8) in (a) URLLC usage percentage, in (b) eMBB usage percentage.

Finally, Figure 5-23 a and b, illustrate the impact of eMBB traffic increase (cases 2 and 4 in Table 5-3) in URLLC and eMBB usage percentage respectively. It is shown that as eMBB traffic increases, there is a small impact to URLLC usage percentage.



(a)



(b)

Figure 5-23**:** Impact of eMBB traffic increase (cases 2 and 4) in (a) URLLC usage percentage, in (b) eMBB usage percentage.

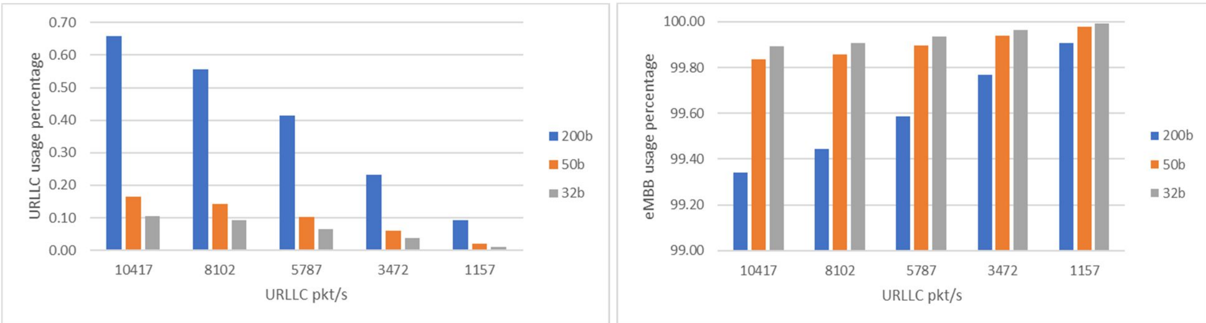
5.6.2 Results on mobility and packet size impact

The impact of our algorithm is also assesed in cases of URLLC traffic increase with different URLLC packet sizes such as 32, 50, 200 bytes and this time the users are moving inside our simulated playground in different speed levels ranging from 70 to 250km/h. By combining these test cases, we can evaluate the mobility aspects in combinations with various file sizes in order to provide a better view on the slices when are deloyed in a high and constantly changing environment. Either at highways or at urban streets congestions may appear due to mobility of the cars to a specific cell, thus resource outage and high traffic demand should be handled.

Table 5-4: Simulation parameters.

Parameter	Value
Macro base stations	19 3-sectorized (57 macro)
ISD	500 m
Small cells	57
Simulation time	60 s
eMBB packet size	2048 (KB)
eMBB users	500
URLLC packet size	32, 50, 200 (bytes)
URLLC users	1000
Vehicle speeds	70, 140, 200, 250 km/h

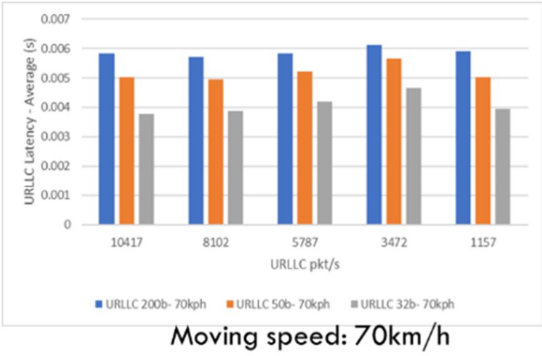
URLLC usage percentage is illustrated in Figure *5-24*a for increasing number of packets per second ranging from 1157 to 10417 and increasing size of files ranging from 32 to 200b. As anticipated the impact of URLLC usage percentage (usage of RBs) is higher when larger URLLC files are utilized. For eMBB, the usage percentage is shown in Figure *5-24*b for the same number of packets and file sizes as tested for the URLLC cases.

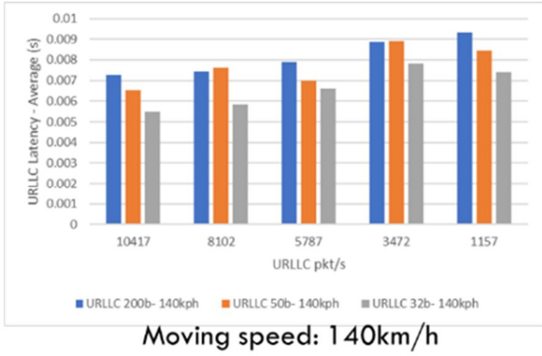


(a) (b)

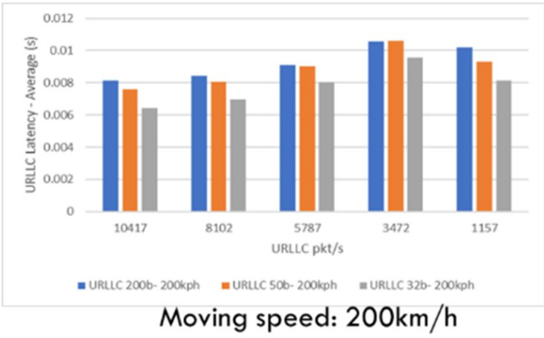
Figure 5-24: (a) URLLC usage percentage and (b) eMBB usage percentage for various packets per second and file sizes.

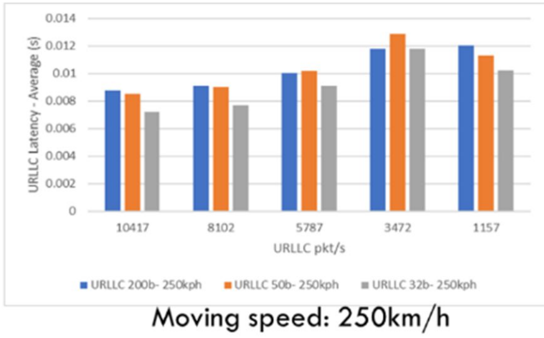
Figure 5-25(a)-(d), illustrate the outcome of the mobility in our scenarios. Success rate and latency is measured for all the cases. The time that a packet trasmittion is considered as successful is proportional to the size of the file. Figure 5-25(a) present the impact to URLLC latency for moving speeds of 70km/h. In general, latency is higher as the URLLC file size increases from 32 bytes to 200 bytes. Similar behaviour is shown in cases (b)-(d) with speeds of 140km/h, 200km/h and 250km/h respectively.





(a) (b)

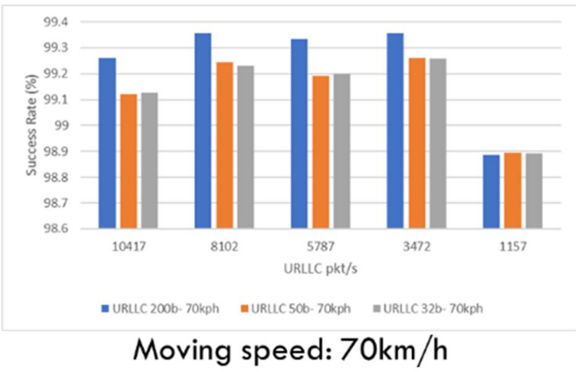


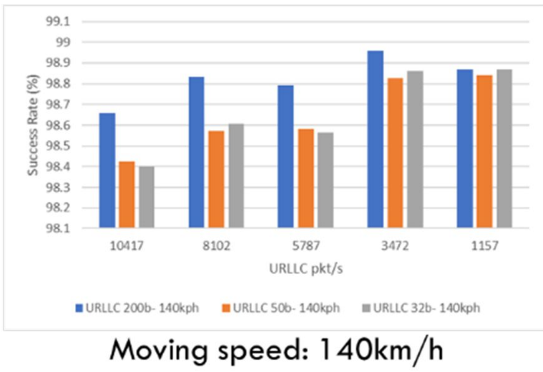


(c) (d)

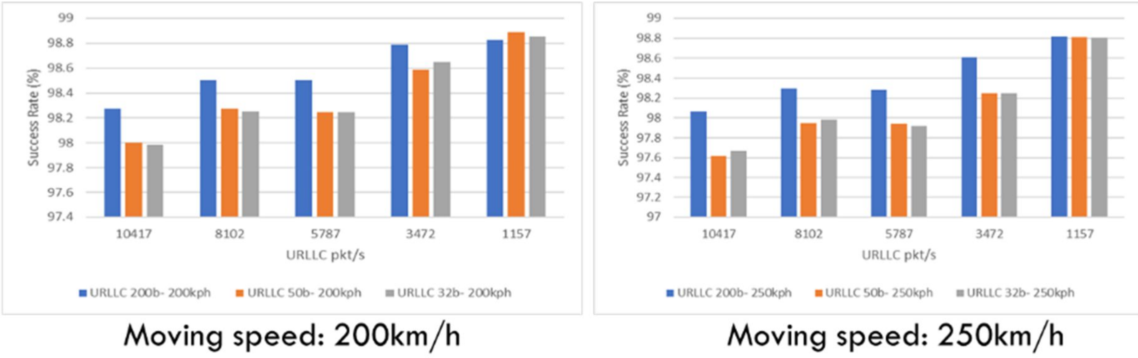
Figure 5-25: Latency for various values of URLLC packets per second and file sizes for different vehicles speed (grouped by moving speeds) (a) 70km/h, (b) 140km/h, (c) 200km/h, (d) 250km/h.

Figure 5-26 (a)-(d) illustrate the success rate for various values of URLLC packets per second and different file sizes which are grouped by speed levels ranging from 70km/h up to 250km/h. Specifically, the success rate is the number of the transmitted packets that have been successfully received from the user or the cell depending on the direction of transmission either downlink or uplink divided by the total transmitted packets (successful and dropped). It is shown that the success rate is better in lower speeds (e.g. Figure 5-26 (a)) compared to higher moving speeds of UEs.





(a) (b)



(c) (d)

Figure 5-26: Success rate for various values of URLLC packets per second and file sizes for different vehicles speed (grouped by moving speed).

Figure 5-27 illustrates the latency of URLLC packets for various values of URLLC packets per second and different file sizes which are grouped by URLLC file sizes ranging from 32 bytes to 200 bytes. It is shown that the latency is better in lower file sizes and lower speeds compared to higher moving speeds of UEs.

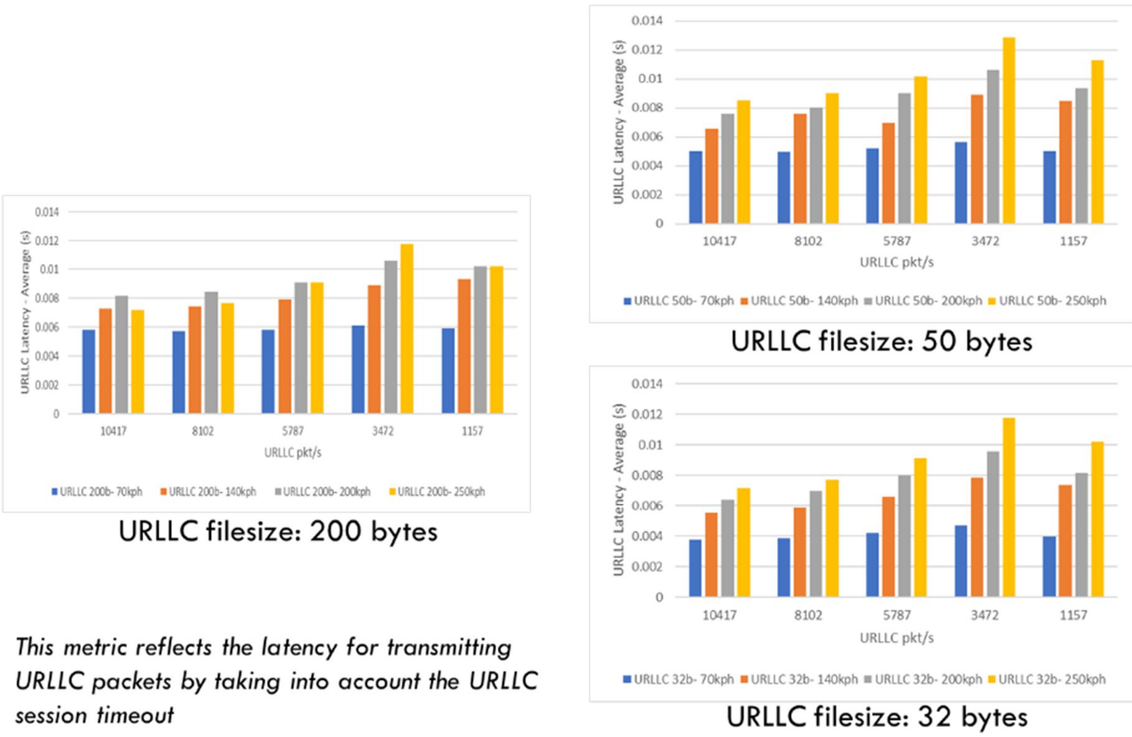
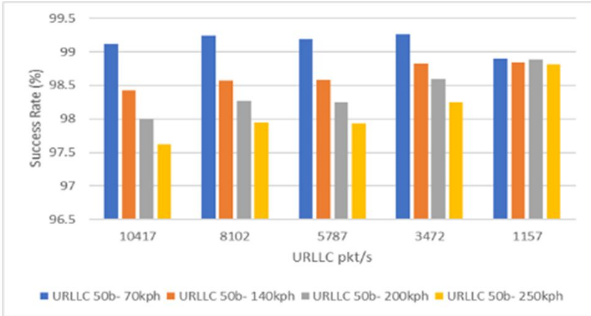
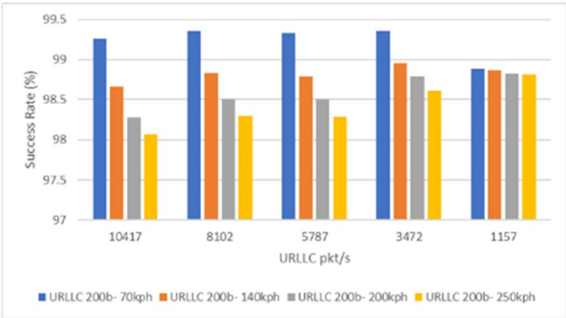


Figure 5-27: Latency for various values of URLLC packets per second and file sizes for different vehicles speed (grouped by file sizes).

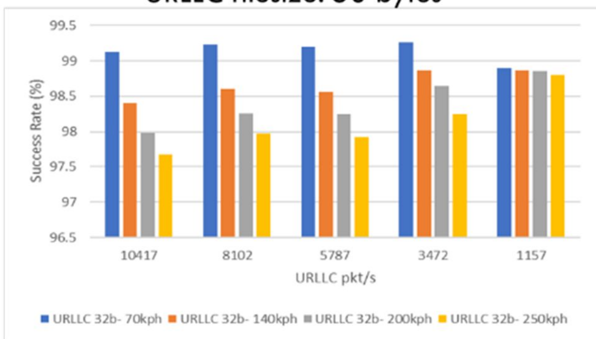
Figure 5-28 illustrates the success rate of URLLC packets for various values of URLLC packets per second and different file sizes which are grouped by URLLC file sizes ranging from 32 bytes to 200 bytes. Overall, it is shown that the success rate is better in lower file sizes and lower speeds compared to higher moving speeds of UEs. However, decrease of success rate is not very large in in higher speeds (e.g. we may see a decrease of success rate from 99% to 97.5%). As a result, the provided services are not degraded a lot, even in a moving environment.



URLLC filesize: 50 bytes



URLLC filesize: 200 bytes



URLLC filesize: 32 bytes

Success rate shows the level of successfully transmitted packets by taking into account the number of URLLC packets per second per UE

Figure 5-28: Success rate for various values of URLLC packets per second and file sizes for different vehicles speed (grouped by file sizes).

5.7 Benefits and Conclusion

The algorithm which is presented in this chapter, tries to solve the problems that are created when various slices with different traffic services are utilized in the same network. For evaluating the performance of the proposed algorithm and all the requirements that have been indicated by 3GPP in order to utilize the network slicing aspects, exhaustive simulations have been conducted.

To efficiently investigate all the situations of making use of the slices in a real environment, multiple case scenarios had to be created. In this work, co-existence of URLLC and eMBB traffic is considered to evaluate the impact of traffic increase (either eMBB or URLLC, and both) to the overall network performance. With this in mind, the network slicing concept has been identified for scenarios where different traffic is demanded simultaneously by users. In addition, various file sizes and user movement that ranges from low to high speeds that the users may have in real life, have been included in simulations. The network slices need to be handled by the system in respect to user latency, throughput and other metrics that are specified in the SLAs. Service quality must be monitored and the algorithm should be able to make all the appropriate changes, such as preemption, reassignment of resources or even to make available extra resources in the problematic area. In this respect, the algorithm that is proposed has been design for creating and deciding on the dynamic resource allocation of network slices which reconfigures and adjusts the slices to provide appropriate QoS levels towards mobile client nodes.

A full range of tests that have been donned in order to evaluate all the scenarios and use cases for the use of two or three network slices were presented. The results shows that prioritizing the URLLC services and thus providing the appropriate resources to the slices that support this service, the system does not misbehave and also the users are not affected from the changes that happen to other slices in the same network. Slice isolation has been achieved even in high mobility situations or extreme URLLC use cases where the traffic demand is increased in higher that normal values for our testing purposes. Overall, in terms of usage of resources between URLLC and eMBB it seems that the impact is not very high, hence critical URLLC traffic can be adequately served by dedicated slices, while the rest of traffic is handled by eMBB slices.

With network slicing, operators can allocate the appropriate amount of network resources to a specific slice providing high availability, a specified latency, data rate and also security. At the same time, a different network slice could be offered to provide high throughput, and good levels of latency without any problems as traffic demand changes to either of the services proving the isolation and QoS levels at the same time.

5.8 References

[1] Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Kostas Tsagkaris, Zwi Altman, Sana Ben Jemaa, Panagiotis Demestichas, Nicolaos Mitrou, "System-level Evaluation and Management of 5G RAN Slicing Concept ", submitted to IEE Vehicular Technology Magazine
[2] 3GPP TS 23.501: URLLC traffic model and QoS parameter" S2-178901 SA WG2 Meeting #124
[3] 3GPP TR 37.868 "RAN Improvements for Machine-type Communications"
[4] 3GPP TS 23.501 "URLLC traffic model and QoS parameter", S2-178901, SA WG2 Meeting #124
[5] 3GPP TR 36.942 "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios"
[6] NGMN, "Description of Network Slicing Concept", Jan. 2016
[7] 5G-PPP, "Architecture White Paper", Jul. 2017
[8] IETF, "Heterogeneous Network Slicing", Oct. 2017
[9] 5G-PPP, "Vision on Software Networks and 5G", 2017
[10] 3GPP TS 23.799, "Study on Architecture for Next Generation System (Release 14)"
[11] 3GPP TR 22.885, "Study on LTE support for Vehicle to Everything (V2X) services (Release 14)"
[12] 3GPP TS 22.261, "Service requirements for the 5G system; Stage 1 (Release 16)"
[13] 3GPP TS 22.891, "Feasibility Study on New Services and Markets Technology Enablers; Stage 1 (Release 14)"
[14] 3GPP TS 38.300, "NR; NR and NG-RAN Overall Description; Stage 2 (Release 15)"
[15] M. R. Sama, S. Beker, W. Kiess and S. Thakolsri, "Service-Based Slice Selection Function for 5G", 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, 2016, pp. 1-6
[16] 3GPP, "3GPP's flexible 5G system architecture", 2016
[17] A. Devlic, A. Hamidian, D. Liang, M. Eriksson, A. Consoli and J. Lundstedt, "NESMO: Network slicing management and orchestration framework", 2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris, 2017, pp. 1202-1208
[18] 3GPP TR 32.899, "Study on charging aspects of 5G system architecture phase 1 (Release 15)"
[19] 3GPP TS 28.530, "Management of 5G networks and network slicing; Concepts, use cases and requirements (Release 15)"
[20] 3GPP TS 28.531, "Provisioning of network slicing for 5G networks and services (Release 15)"
[21] 3GPP TS 23.501, "System Architecture for the 5G System; Stage 2 (Release 15)"
[22] 3GPP TS 23.503, "Policy and Charging Control Framework for the 5G System; Stage 2 (Release 15)"

6 Sharing and allocation of resources for new 5G services (URRLC, eMBB, mMTC)

Heterogeneous networks constitute a promising solution to the emerging challenges of 5G networks. According to the specific network architecture, a macro-cell base station (MBS) shares the same spectral resources with a number of small cell base stations (SBSs), resulting in increased co-channel interference (CCI). The efficient management of CCI has been studied extensively in the literature and various dynamic channel assignment (DCA) schemes have been proposed. However, the majority of these schemes consider a uniform approach for the users without taking into account the different quality requirements of each application, such as ultra reliable communication, extended mobile broadband and massive machine type communications. In this work, we propose an algorithm for enabling Dynamic Channel Assignment in the 5G era that receives information about the interference and QoS levels and dynamically assigns the best channel taking into account the requirements of the users. This algorithm is compared to state-of-the-art channel assignment algorithm. Results show an increase of performance in terms of throughput and air interface latency. Finally, potential challenges and way forward are also discussed.

6.1 Introduction

5G is characterized by the challenges of rapid growth in mobile connections and traffic volume [1], [2]. To address these challenges, the European project SPEED-5G (standing for quality of Service Provision and capacity Expansion through Extended-DSA for 5G) focuses on the efficient exploitation of wireless technologies so as to provide higher capacity along with the ultra-densification of cellular technology [3]. Under the framework of SPEED-5G, novel techniques for optimizing spectrum utilization will be developed, following three main dimensions: i) ultra-densification through small cells, ii) load balancing across available spectrum and iii) exploitation of resources across different technologies. Considering the specific three dimensional model, which is referred to as extended-Dynamic Spectrum Allocation (eDSA), different spectrum bands and technologies can be jointly managed so as to improve the users' Quality of Experience (QoE). Hence, the ultimate goal of SPEED-5G boils down to the development of a dynamic radio resource management framework, including mechanisms for interference control, coexistence of

Figure 6-1: Broadband wireless scenario of SPEED-5G.

heterogeneous networks, management of spectral resources in lightly-licensed bands and other smart resource allocation schemes. It is worth mentioning that this work is an extended version of the work published by the authors in [17]-[18].

One of the main scenarios addressed in SPEED-5G is the case of heterogeneous networks where a massive deployment of small cells is put into place to deliver a uniform broadband experience to the users, considering applications with different QoS requirements, such as high resolution multimedia streaming, gaming, video calling, and cloud services. A significant challenge in these networks is the efficient management of co-channel interference (CCI) that occurs due to proximity among the SBSs. Hence, given that the same channels are reused among SBSs due to the scarce spectral resources, CCI constitutes an important restrictive factor for the network performance.

To confront this challenge, dynamic channel assignment (DCA) techniques have been proposed in the literature, either considering a centralized approach [4] or a distributed one [5]. In particular, in [4], a centralized DCA technique considering a heterogeneous network that consists of small cells and macro cells is investigated based on the graph approach. It should be noted that the centralized approaches have several advantages in terms of performance. Nevertheless, the high computational complexity renders them inappropriate for the case of a heterogeneous network with a massive number of small cells. Therefore, distributed DCA techniques have gained the interest of many researchers as a solution that can be applied in future wireless networks. An interesting approach of a distributed adaptive channel allocation scheme known as channel segregation has been proposed in [6], to improve the spectrum efficiency in cellular networks. According to this approach, each cell creates a priority table with the available channels and tries to use the channels with the highest priority. Using this technique, an efficient stable channel re-use pattern is formed and the system performance is ameliorated. Due to the inherent advantages of this method, various DCA mechanisms based on channel segregation have been proposed by the research community [7]-[9]. However, the majority of the DCA schemes in the literature consider that the SBSs do not differentiate between traffic requests from user equipment (UE) applications, even if the applications do not have the same priority from the user point of view. Considering that in 5G networks, the traffic will range from high data rates to machine type traffic, covering a variety of different applications, there is an emerging need for DCA schemes that provide differentiated QoS to each user, coping with the changing network conditions and the time-varying CCI. Based on this remark and the work presented in [7], we study a modified distributed channel segregation mechanism that takes into account the CCI and the QoS characteristics of the users. The proposed Interference and QoS aware channel segregation based DCA (IQ-CS-DCA) can be employed in order to use the spectral resources efficiently and at the same time prioritize the users with delay-constrained applications (such as video streaming).

The rest of the chapter is organized as follows. In Section II, a brief description of the scenarios considered is given, focusing on the scenario of interest. Section III summarizes the previous work in this research area and Section IV presents an algorithmic description of the proposed IQ-CS-DCA mechanism. Finally, Section V discusses the challenges that should be addressed in the mechanism and some future work whereas Section V concludes the chapter.

6.2 Scenarios description

Mainly investigate scenarios where capacity demands are the highest and where eDSA will exploit efficiently the co-existence of different technologies. More specifically, the selected scenarios are the following:

- Massive IoT (Internet of Things): this scenario refers to the “low-end IoT” and covers devices with sporadic and delay-tolerant traffic, mainly composed of short packets. Among others, this category typically includes wearable devices, smart meters, home automation devices, healthcare, non-critical smart cities sensors and wireless sensor networks for environmental monitoring.
- Ultra-reliable communications: this scenario refers to a network that supports services with extreme requirements on availability and reliability. Particularly, it is envisioned to have new applications based on M2M (machine-to-machine) and IoT communication with real-time constraints, enabling

new functionalities for traffic safety, traffic efficiency or mission-critical control for industrial and military applications.

- Broadband wireless: this use case constitutes the scenario of interest and it focuses on a mixture of domestic, enterprise and public access outdoor and indoor environments located in a densely populated urban area (see Figure 1). In this case, a large number of small cells co-exist within a macro-cell offering an improved communication experience to the users.

In order to meet the 5G requirements, which characterize the specific use case, we propose a DCA mechanism for the efficient usage of the available spectrum, driven by the coordination of the CCI and the users' QoS requirements.

6.3 Related Work

The concept of heterogeneous networks focuses on the improvement of spectral efficiency per unit area using a diverse set of base stations (BS), in a mix of macro cells and small cells. As highlighted in the introduction section, one of the main problems in these networks is the efficient management of CCI between the different cells to enhance the network performance. Towards this direction, one promising category of DCA mechanisms, which has been recently studied in the literature, refers to the channel segregation based DCA (CS-DCA) mechanisms [8].

One of the first approaches of channel segregation appears in [6]. In this work, each BS acquires its favorite channel independently, through learning from statistical data. As a result, the process of channel re-use is self-organized, leading to an amelioration of spectrum efficiency. In the simulation results, the proposed mechanism is compared with a system without segregation and the performance improvement in terms of blocking probability and channel utilization is demonstrated.

In [7], the authors modify the previous CS-DCA mechanism for application to Distributed Antenna Networks (DANs). According to the modified scheme, known as interference-aware CS-DCA (IACS-DCA), the average CCI is computed for the available channels and a CCI table attached to each antenna is created with the first channel to have the lowest CCI value. Hence, before a transmission attempt, the user selects the closest distributed antenna and the first channel of the corresponding table is assigned to him. In the simulation analysis, the performance improvement due to the existence of DAN is confirmed. Furthermore, the superiority of the proposed algorithm compared to a fixed channel allocation (FCA) mechanism is proven. Based on this mechanism, the authors in [9] examine several metrics such as the autocorrelation function, the fairness index of channel reuse pattern and the minimum co-channel distance among the BSs. From their analysis, it can be seen that the proposed scheme forms a CCI minimized channel reuse pattern that ameliorates the signal-to-interference ratio (SIR) compared to other channel assignments schemes.

Taking into account the increasing number of wireless terminals, the authors in [9] propose a modification of the IACS-DCA, named as 'multi-group IACS-DCA'. According to this mechanism, the available channels are divided into multiple groups and the initial IACS-DCA is applied to each channel group. As they highlight in their simulation analysis, the specific mechanism results not only in a more stable reuse pattern in case of multiple users but also in amelioration of the SIR compared to the single-group IACS-DCA scheme.

An energy- efficient approach of the IACS-DCA scheme is presented in [11]. In this work, the use of a transmission power control scheme in combination with IACS-DCA is studied and the stability of the channel reuse patter is verified. From the simulations analysis, some indicative positive results for the transmission power reduction are presented. A similar energy-efficient approach is also proposed by the authors in [12]. In their work, they propose a modified approach of IACS–DCA mechanism in combination with a learning game-theoretic algorithm for the BS sleep process. The performance of the proposed mechanism is investigated through simulation and its superiority in terms of energy and spectral efficiency is proven.

6.4 Algorithmic Approach for RRM/MAC

In this section, we describe the proposed IQ-CS-DCA mechanism that is based on the IACS–DCA mechanism presented in [7]. At first, we present an abstract formulation of the considered optimization problem whereas in the second subsection a more algorithmic approach of the proposed mechanism is given.

6.4.1 Abstract form of the Optimization Problem

In our approach, we consider the scenario of a heterogeneous network with one macro-cell and multiple small-cells, similar to Figure 6-2.

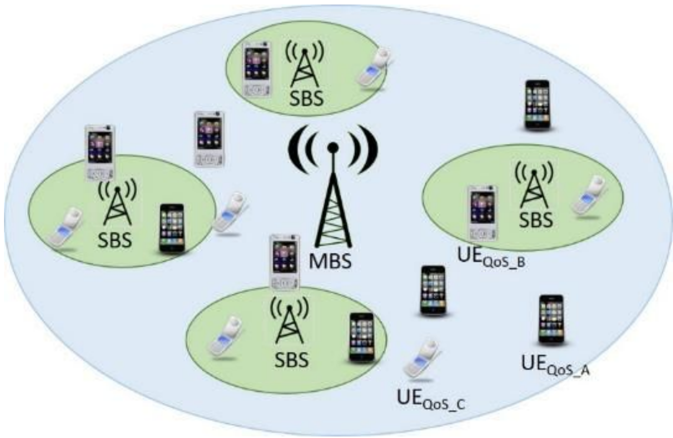


Figure 6-2: Heterogeneous Network scenario.

Considering the network elements, an abstract formulation of the studied resource allocation problem can be given as follows:

Given:

- The large number of SBSs
- The diverse QoS requirements of the UEs
- The time-varying network conditions (due to various traffic characteristics, changing propagation environment, power control etc.)
- The limited number of spectrum channels

Find:

- An efficient association of UEs to SBSs
- An efficient channel assignment to UEs

So as:

- To maximize the spectral-efficiency (via an adequate re-use channel pattern)
- To satisfy the communication quality of the UEs

6.4.2 High level Description of IQ-CS-DCA Mechanism

The proposed mechanism can be divided in five main steps. The following flowchart in Figure 6-3 summarizes the *IQ-CS-DCA* mechanism and each phase is briefly described.

- Initialization Phase: During this phase, each SBS chooses randomly a channel from the pool of available channels and broadcasts a beacon signal on this channel.
- Measurement Phase: Each SBS measures periodically the instantaneous beacon signal power on each of the available channels for a specific time duration. The received power can be computed considering both path loss and fading phenomena for a more complete analysis of the radio propagation environment.
- Creation of channel priority table: Each SBS creates the channel priority table based on average CCI power levels. In this step, the average CCI power can be computed either by using the first-ordering filtering similar to [7] or by using other learning/average mechanisms that use past CCI measurements and result in a stable assignment. The channel with the lowest CCI appears first in the priority table and the other channels with descending order of CCI follow.
- UEs-SBS Association: During this phase, each UE associates with a SBS depending on various metrics (e.g. the highest receive signal strength indicator (RSSI), the load due to other UEs associated with this SBS etc.).
- Collection of requests: Each SBS collects the channel requests from the UEs.
- Creation of User QoS requirement priority tables: UEs are prioritized depending on their application priority and the SBSs divides its priority table into multiple tables (depending on the number of UEs/applications). The first channel of each table is assigned to each UE depending on its application priority and the channel quality given by the CCI power level (better channels are given to UEs with stricter QoS requirements).
- Channel Assignment: Each SBS assigns the channels to the users as follows based on the QoS priority tables.

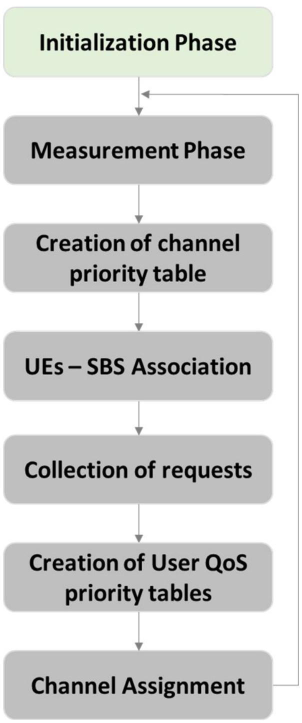


Figure 6-3: Flowchart of IQ-CS-DCA mechanism.

6.5 Implementation approach of the proposed Algorithm

To evaluate our proposed solution, we have implement two algorithms as a first stage. In order to provide results as a "proof of concept" we have introduce a state-of-the-art algorithm and our proposed algorithm which uses the interference levels acquired from the network as well as the QoS requirements from each UE.

The different QoS requirements are related to the 5G scenarios as explained at the SPEED-5G project. For example IoT users will have low priority in our algorithm, Ultra reliable communications will be at highest priority, meaning that they will be assigned to channels with better SINR value. Finally the broadband communications where the users will either have a medium priority with respect to the quality of the channel assigned to or even a high priority given that could belong to a category of users that need to have low latencies and high throughputs.

In advance those two solutions are compared with an algorithm from the presented state-of-the-art. The algorithm that was developed is the dynamic channel assignment scheme for distributed antenna networks found at [7] that our solution was based on. Figure 6-4 illustrates the procedure of the algorithm on the assignment of the channels for the multitude of the antennas as presented.

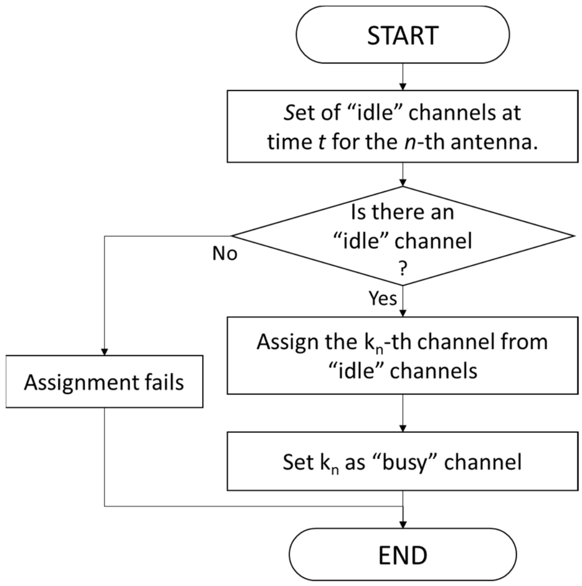


Figure 6-4: Flowchart of channel assignment at the n-th antenna [7].

Specifically, the first algorithm that has been investigated is called "Random-based Channel Assignment (CA) Algorithm" (Figure 6-5). This solution does not have a certain logic for the assignment of channels that's why it's called 'random', it arbitrarily allocates users to different channels without any knowledge of the channels current status or the whole system. As input we have a set of UEs U that want to transmit, a set of macro BS M, a set of channels C, a set of available channels $Ca \subseteq C$. Then a certain procedure is followed in order to allocate certain channels to UEs as the flowchart illustrates.

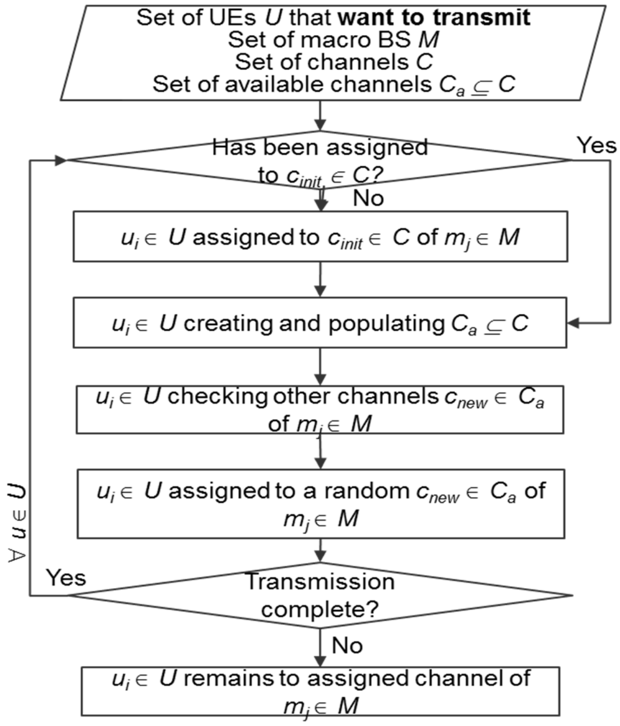


Figure 6-5: Flowchart of the Random-based CA Algorithm (R-CA).

The Random-based Channel Assignment (R-CA) algorithm is used as a baseline in order to compare, evaluate and optimize the effectiveness of the next algorithm that we propose (Figure 6-6). As input we have again a set of UEs U that want to transmit, a set of macro BS M, a set of channels C, a set of available channels Ca $\subseteq$ C. The selection procedure differs from the random channel assignment algorithm since here we introduce a control point for checking the best available channels in order to select these (if available). The best channel is identified according to the SINR and if the SINR of a new channel is better than the currently utilized one, then the UE will switch to the better channel. This algorithm enables context-aware RRM as each base station can collect the interference levels for each user that is connected to a specific channel in order to deduce the radio environment status and exploit it appropriately. In general, it is expected that through this algorithm, will be possible to achieve better quality (e.g., higher throughput, less latency).

In order to calculate the SINR levels, every SBS at every cycle, creates the average interference that it calculates from the input of the UEs in the area for each other SBS. Instead of recalculating the signal strength at the location of the SBS, we are using the feedback from the SBS' served UE devices, to make the measurements more realistic and efficient. The UE devices store the signal strength per SBS in their physical layer variables. Therefore, we are averaging out this signal strength to calculate the average per rat interference. After this calculation, we then turn to the history of our RRM model transmission. Since it is not possible to know beforehand what collisions will occur on an air frame, we are using the previous transmission (history) to calculate these collisions and assume that statistically the impact will be the same. We create the resource utilization mask for each SBS, based on the RBs utilization data that were stored in the history variable.

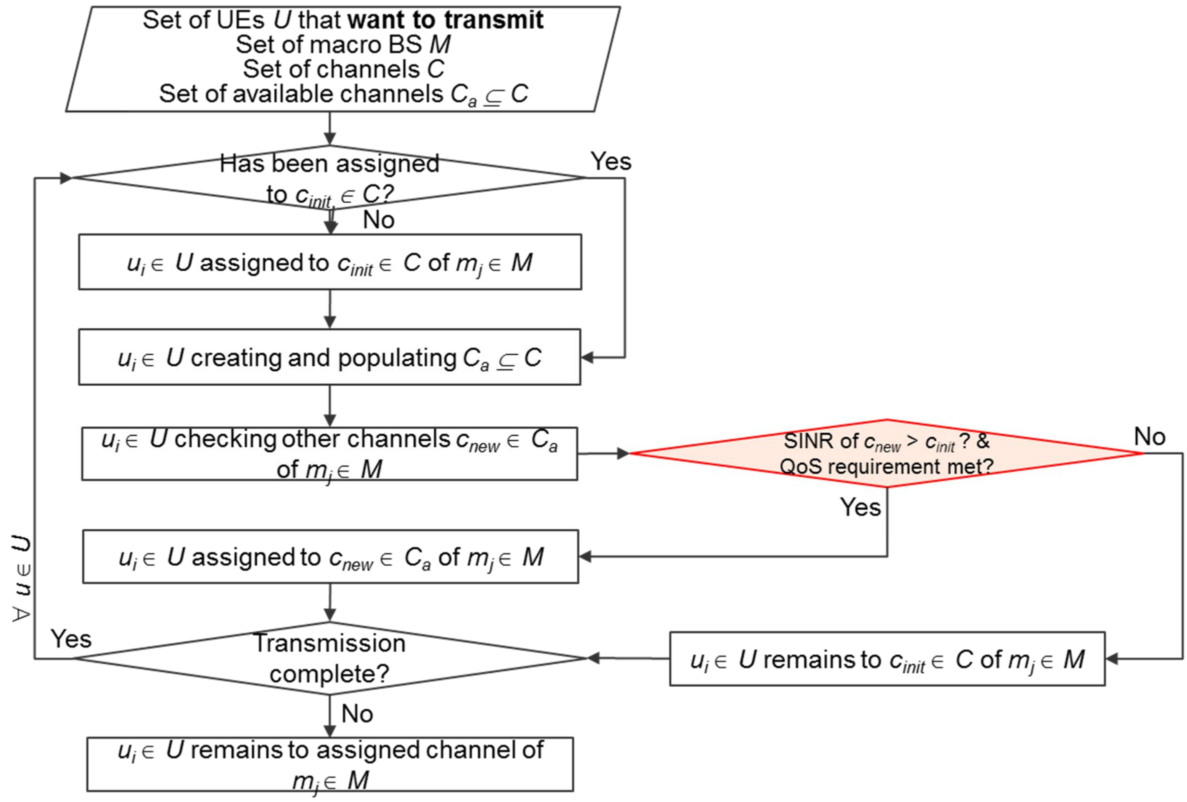


Figure 6-6: Flowchart of our proposed algorithm.

6.6 Evaluation aspects of the proposed algorithms

6.6.1 Simulation Tool

For the evaluation of such concepts, extensive system level simulations are conducted. The implementation of our suggested solution was performed under a proprietary system-level simulation tool which is fully developed in Java with various capabilities and has been calibrated according to the 3GPP specifications. The simulator takes into account various parameters such as traffic level, available infrastructure elements, available channels and evaluates the various test cases. The calibration state of the proprietary simulator has been checked against the reference results of the 3GPP LTE calibration campaign [36.814] [16]. As a result, the Cumulative Distribution Function (CDF) of coupling loss and downlink SINR have been checked in order to calibrate the tool with leading operators and vendors such as Nokia, Ericsson, Docomo, Huawei, Telecom Italia etc.

The configuration is fully customizable so as to include various types of cells (i.e., macro and small cells). Specifically, it is possible to customize the following: the size of playground; the area type (e.g., Dense Urban etc.); the number and position of macro base stations and their inter-site distances (ISDs); the number and position of small cells per macro base station; the number and position of end-user devices in the playground; the number of available channels. In addition, the pathloss models for macrocells at 2 GHz band is set to L = 128.1 + 37.6log10(R), R in km, and for small cells is set to L= 140.7+36.7log10(R), R in km.

6.6.2 Simulation Parameters

The parameters imported to the simulator are, 19 macro base stations (BS) each with three cells and also nine small-cells per (BS) giving us a total of 228 cells (in which 171 are small cells uniformly distributed in the simulation playground) throughout the network. In addition, we have utilized 4 channels at 20MHz bandwidth for every cell. The topology of the network created by the simulation tool is illustrated at (Figure 6-7) and the differing shapes and symbols can be interpreted as the UEs and the wireless communication links. Specifically users are shown as small circles with four different colors (red, green, blue and light blue) that represent the four channels that have been utilized for our scenarios. Furthermore the green arrows illustrate the transmission process and the connection topology between UEs and BSs of each user to a specific cell of our network. The small cells are located close to the center of the macro cells and are working at the 3.5Ghz band in contrary to the macro cells that are working on 2Ghz, giving us a heterogeneous environment for our scenarios.

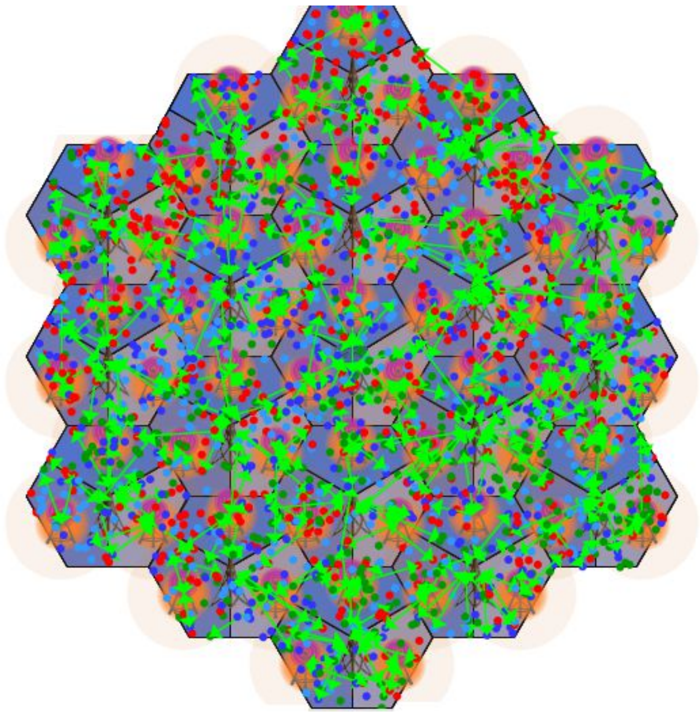


Figure 6-7: Network topology created by the simulation tool.

In addition Table 6-1 presents the configuration of the base stations used in each case and their values that have been introduced to our simulator for the development and evaluation of our solution.

Table 6-1: Configuration of BSs.

BS	MIMO mode	Bandwidth
Macro	2x2	20Mhz
Small	Omni-antenna	20Mhz

6.6.3 Experimentation scenarios and test cases

In order to analyze these two algorithms with the use of the simulation tool we had to introduce different scenarios and test cases that are summarized in our experiments. We have experimented with five values of sessions that every user requests per day ranging from 2,880 up to 14,400 sessions. The file size requested from each user is 2Mbytes, which means that a user can request from 4MB up to 20MB per minute. Those values could provide us with a broader knowledge of the algorithms capabilities for the specific network topology implemented to the simulator.

For each of the cases three algorithms have been evaluated. Algorithm (A) is our proposed algorithm which builds on state-of-the-art and adds the notion of QoS prioritization. Algorithm (B) is a state-of-the-art algorithm which sorts available channels based on SINR values and Algorithm (C) uses a random allocation of channels (not necessarily the best one).

Moreover, high, medium and low priority services are considered where our proposed algorithm tries to allocate the best possible channels firstly to the high priority services, then to the medium and finally to the low priority services.

Table 6-2: Tested scenarios cases.

Test cases	Users	Sessions/Day/User	Packet size
1	5000	14400	2Mbytes
2	5000	11520	2Mbytes
3	5000	8640	2Mbytes
4	5000	5670	2Mbytes
5	5000	2280	2MBytes

6.6.4 Evaluation Results

Figure 6-8 indicates the results from our experimentation. The barchart illustrates the cases of Table 6-2 and specifically the average air interface latency. On average it is shown that our proposed algorithm outperforms the other two algorithms (up to 50%) especially in high and medium priority services by giving them a performance boost. On the contrary, low priority services seem that they don't benefit as much as the other two.

Figure 6-8: Average air-interface latency for each test case (s).

In Figure 6-9 the results are sorted by priority levels and we see in a clearer way the large benefit of our algorithm for high priority which is not the case for low priority services.

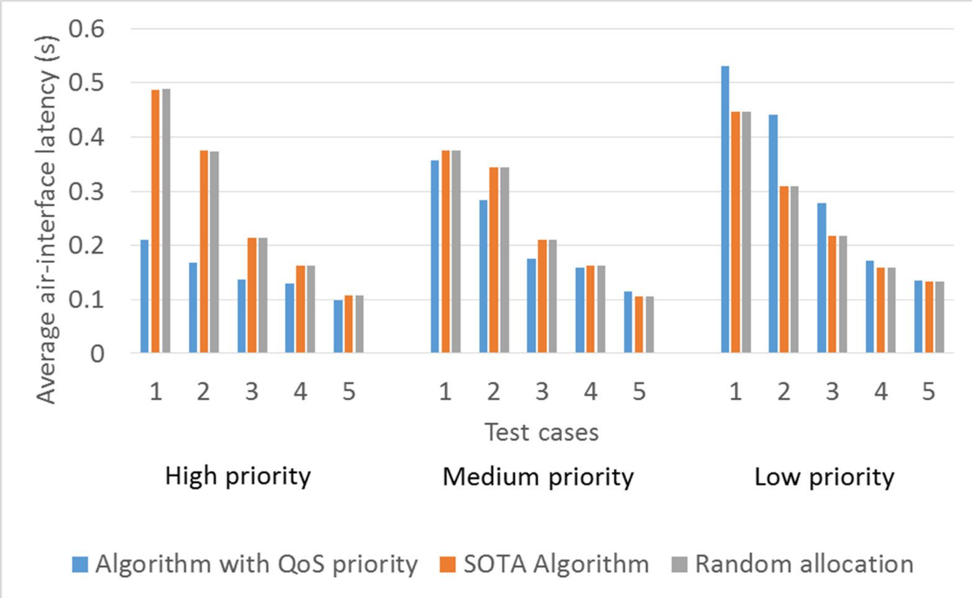


Figure 6-9: Average air-interface latency for each service priority level.

Furthermore, Figure 6-10 illustrates the normalized throughput for each of the test cases and compared among each algorithm. It is evident that our algorithm performs better in almost every test case and especially in cases with higher loads (compared to less-loaded simulations).

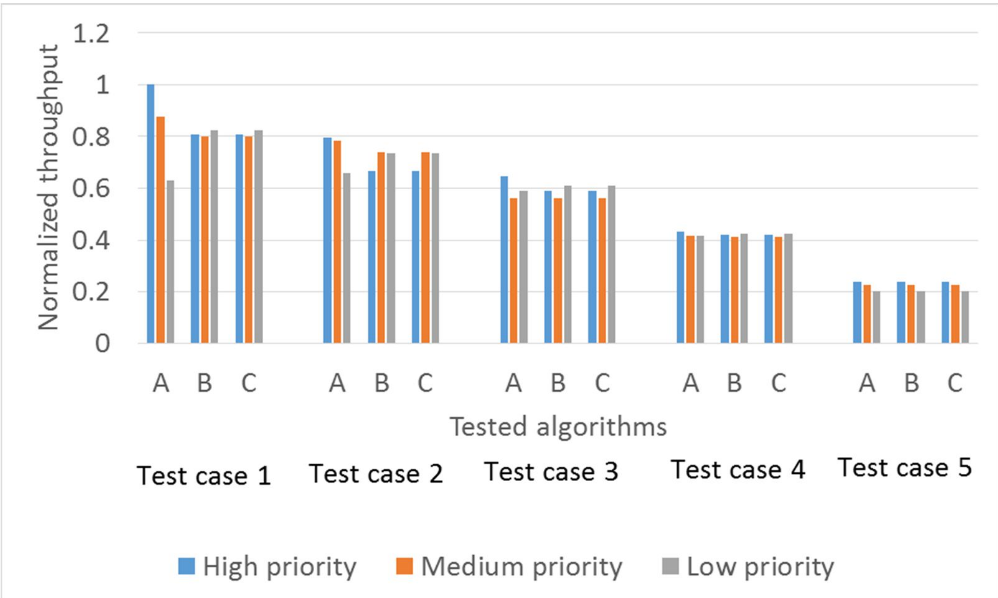


Figure 6-10: Normalized throughput for each test case (s).

Similarly, Figure 6-11 illustrates the normalized throughput as of service priority levels and also here (as shown in latency charts), our solution seems to perform better especially in higher and medium priority services compared to low priority services. Here we investigated how a radio resource management algorithm utilizing contextual information acquired from the network and taking into account QoS requirement can cope on some specific scenarios. The test cases introduced here where designed in order to investigate the performance difference between the state-of-the-art and our proposed solution for an environment of almost one user per square meter at the case of 5000 users and files for 2MB size and variable number of requests per min per user (ranging from 2 to 10). Our proposed solution was able to dynamically choose the best channel based on interference of the current position and thus allow each user to connect with higher speed and receive the file faster with less air interface latency.

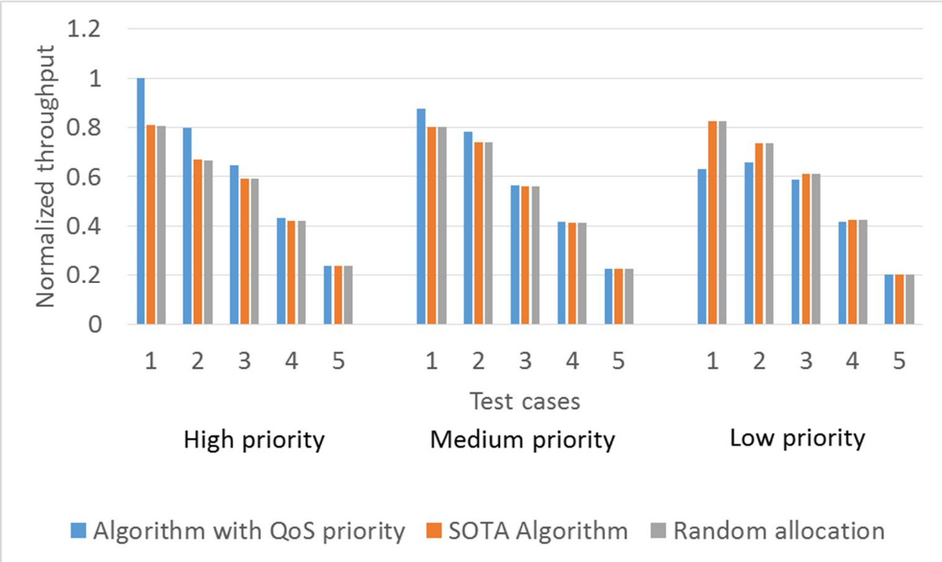


Figure 6-11: Normalized throughput for each service priority level.

On the contrary, the algorithms that used for comparison on average were making the less optimal selection of the channels (without giving priority based on QoS requirements), hence the users were not able to download at full speed and with higher loss packet ration, creating a continuously loop of poor selection of channels without being able to overcome this situation.

6.7 Challenges and future work

In the following, we discuss some challenges that we would like to address in the development of IQ-CS-DCA as well as some of the possible improvements left for future work.

The ultra-densification of networks that is currently envisioned for 5G will bring new challenges to the radio access, especially related to interference management. In fact, ultra-dense networks are characterized by interference patterns that change quickly in time and that strongly depend on the realistic network deployment [13]. Therefore, interference mitigation techniques must be sufficiently dynamic and operate on a sufficiently small time-scale to ensure that the fluctuations in the interference are captured. Clearly, the same challenges also apply to other mechanisms that take into account interference measurements to operate, as in the case of the proposed IQ-CS-DCA.

Additionally, the aggregation of multiple radio access technologies (RATs), possibly operating on very diverse frequency bands with different characteristics, brings new challenges related to efficient ways to ensure the provision of the end-to-end QoS, as some RATs may not supply an explicit way to perform QoS management. This is currently a very active area of research in both academia and industry [15]. Examples of current technologies that will still play an important role in the decades to come are the Long Term Evolution (LTE) and the IEEE 802.11 (WiFi) family of standards. Table 6-3 gathers some of the key differences between the two, exemplifying the challenge due to the differences in the physical layer and the radio channel access.

Table 6-3 Comparison between LTE and WiFi.

	LTE	WiFi
Spectrum access	Licensed	Unlicensed
Channel bandwidths	1.4, 3 5, 10, 15, 20 MHz	5, 10, 20 MHz
Channel access method	Centralized	Contention-based
Physical layer	OFDMA/SC-FDMA	OFDM-CSMA
Optimized for mobility	yes	no

As it has been referred in the Introduction section, a purely centralized approach to DCA might not be feasible in practice due to the large amount of feedback that must be exchanged between the BSs and the strict delay requirements that must be met by the network infrastructure. On the other hand, purely decentralized techniques might not be able to address all the aforementioned challenges. As a future work, we would like to investigate hybrid or semi-decentralized approaches, for instance cluster-based algorithms that take decisions with a minimum amount of feedback to be exchanged between the SBSs. Clearly, investigations should be performed to analyse the tradeoff between the complexity introduced due to the feedback channel required by these techniques versus the achieved benefits.

Hybrid approaches to DCA can also bring benefits against the bursts of interference that can occur in heterogeneous networks, since they may avoid the strong interference factor generated by closely located BSs, during the initial measurement phase of the IQ-CS-DCA algorithm. It is worth saying that the assumption of a clustered architecture of SBSs is a viable hypothesis, since other functional entities of the network may already require this structure (e.g. to perform inter-cell interference coordination or soft handovers between SBSs). These approaches could leverage the X2 interface currently defined in LTE-Advanced, as well as the Xw interface newly introduced in LTE Release 13 between a 3GPP BS and a WiFi access point [14].

The approach presented in this chapter is directly applicable to the case of heterogeneous RATs. Compared to an algorithm that deals with the underlying available channels agnostically, improvements are expected by providing additional information related to the different available technologies in the channel assignment step. As an example, a fast moving UE might better be assigned to a licensed band using LTE, rather than scheduled to an unlicensed band operating on WiFi. In other cases, though, it may be beneficial to off-load broadband static users to unlicensed WiFi in order to reduce the CCI generated in the licensed band. Additional information is needed by such evolved mechanisms, including the UE's capabilities (i.e. the supported bands and technologies), as well as a characterization of the user's mobility pattern, among others.

Finally, the growing importance of techniques for dynamic spectrum access in next-generation wireless systems should also be taken into consideration in the design of efficient DCA algorithms, together with the diverse characteristics of the underlying frequency bands. For instance, TV White Spaces (TVWS) are highly location dependent. TVWS frequencies might be available in a particular geographical area while being completely occupied at another location. Therefore, a centralized geolocation database might be necessary to implement a coexisting LTE-TVWS system. In such system, the information provided by the geolocation database is semi-static and does not change often in time. However the inherent unpredictability of these frequencies might make them ideal only to best-effort applications with no QoS requirements, further highlighting the importance of intelligent RAT-aware algorithms, as pointed out in [15].

6.8 Conclusions

In this chapter, we consider the case of a heterogeneous network scenario with macro-cell and small cells. Due to the inherent network architecture, one major challenge is the efficient management of CCI. In our work, numerous interference aware DCA mechanisms are discussed and the channel segregation approach is presented. A high-level modified interference aware DCA mechanism that considers the differentiated QoS requirements of the users is proposed. Furthermore, the algorithmic approach of the mechanism is presented through some flow charts and two algorithmic approaches for the evaluation and the proof of concept are examined. The algorithm with the interference aware capability that uses the SINR measurements acquired from the network can provide better results. The advantage of our algorithm introduced here can further explored with the utilization of the small cell at the macro cell edge in order to attract problematic traffic due to the great CCI presented in those areas. Finally, possible challenges and points for future work are recognized in order to have a more complete analysis of the resource allocation problem.

6.9 References

[1] J. G. Andrews et al. "What will 5G be?." Selected Areas in Communications, IEEE Journal on ,32.6, pp. 1065-1082, 2014.

[2] P. Demestichas et al. "5G on the horizon: key challenges for the radio-access network." Vehicular Techn. Magazine, IEEE 8.3, 47-53, 2013.

[3] SPEED-5G EU project site: https://speed-5g.eu/

[4] S.-J. Kim, I. Cho, Y.-K. Kim, and C.-H. Cho, “A two-stage dynamic channel assignment scheme with graph approach for dense femtocell networks,” IEICE Trans. Commun., vol.E97-B, no.10, pp.2222–2229, Oct. 2014.

[5] D. Goodman, S.A. Grandhi, and R. Vijayan, “Distributed dynamic channel assignment schemes,” Proc. IEEE 43rd Vehicular Technology Conference (VTC1993-Spring), May 1993.

[6] Y. Furuya, and Y. Akaiwa. "Channel segregation, a distributed adaptive channel allocation scheme for mobile communication systems." IEICE Transactions on Communications 74.6, pp. 1531-1537, 1991.

[7] R. Matsukawa, T. Obara, and F. Adachi. "A dynamic channel assignment scheme for distributed antenna networks." Vehicular Technology Conference (VTC Spring), 75th. IEEE, 2012.

[8] S. Glisic, and B. Lorenzo. Advanced wireless networks: cognitive, cooperative & opportunistic 4G technology. John Wiley & Sons, 2009.

[9] Y. Matsumura, et al. "Interference-Aware Channel Segregation Based Dynamic Channel Assignment for Wireless Networks." IEICE Transactions on Communications 98.5, pp. 854-860, 2015.

[10] T. Katsuhiro, and F. Adachi. "Multi-group interference-aware channel segregation based dynamic channel assignment." Network Infrastructure and Digital Content (IC-NIDC), 2014 4th IEEE International Conference on. IEEE, 2014.

[11] Y.Matsumura, et al. "Interference-aware channel segregation based dynamic channel assignment using SNR-based transmit power control." Intelligent Signal Processing and Communications Systems (ISPACS), 2013 International Symposium on. IEEE, 2013.

[12] A. Mehbodniya, et al. "Energy-efficient dynamic spectrum access in wireless heterogeneous networks." Communication Workshop (ICCW), 2015 IEEE International Conference on. IEEE, 2015.

[13] B. Soret, et al., "Interference coordination for dense wireless networks", IEEE Communications Magazine, Vol.53, No.1, 102-109, 2015.

[14] 3GPP TR 37.870 V13.0.0, “Study on multiple radio access technology (Multi-RAT) joint coordination”, June 2015.

[15] U. Herzog et al., “Quality of service provision and capacity expansion through extended-DSA for 5G”, European Conference on Networks and Communications (EuCNC 2016), June 2016, in press.

[16] 3GPP TR 36.814, “Further advancements for E-UTRA physical layer aspects”, March 2010.

[17] S. Vassaki, A. Georgakopoulos, F. Miatton, K. Tsagkaris and P. Demestichas, "Interference and QoS aware channel segregation for heterogeneous networks: A preliminary study," 2016 European Conference on Networks and Communications (EuCNC), Athens, 2016, pp. 195-199

[18] A. Georgakopoulos, A. Margaris, K. Tsagkaris and P. Demestichas, "Resource Sharing in 5G Contexts: Achieving Sustainability with Energy and Resource Efficiency," in IEEE Vehicular Technology Magazine, vol. 11, no. 1, pp. 40-49, March 2016.

7 Radio Frequency Resource Management (RRM) in a multi-connection environment

Leveraging multiple simultaneous connections is a solution to enhance the throughput performance per user. Also can be a promising technique for utilizing small cells and millimeter-wave (mmWave) cellular systems that suffer from frequent link interruptions due to blockage in ultra-dense urban deployments. There are performance benefits of multi-connectivity strategies but remain an open research question. Carrier aggregation is one technique used in LTE-Advanced to increase the bandwidth, and thereby increase the data rate per user, whereby multiple frequency blocks (called component carriers) utilizing multi-connectivity. Each aggregated carrier (component carrier or CC) can have a bandwidth of 1.4, 3, 5, 10, 15 or 20 MHz and a maximum of five component carriers can be aggregated, hence the maximum aggregated bandwidth is 100 MHz. For 5G the aggregated bandwidth should surpass the LTE's and go up to 400MHz when

To show the potential of multi-connectivity a framework for better management has been integrated into a simulation tool and assessed in a load-imbalanced scenario. System level simulations in the 5G era, consider demanding use cases with high load and very limited latency in order to cover services such as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC) and ultra-reliable low-latency communications (URLLC). As such, appropriate configuration, environment and network models need to be defined in order to proceed to performance evaluation. The system-level simulation platform is a discrete event simulation environment for the simulation of heterogeneous networks which is extended with new features to support the new functionalities of 5G.

7.1 Introduction to system level simulation

The system-level simulation platform for 5G is a Discrete Event Simulation (DES) environment for the simulation of heterogeneous networks. Also, the platform is extended with new features to support the new functionalities of 5G. The main modules supported are macro cells, small cells and UEs. Based on the DES approach, created events are the basic Third Generation Partnership Project (3GPP) signaling events, mobility events, application layer events and system level events that enable the collection of measurements and the control of auxiliary artifacts (graphics, controls etc.) as described in the sections that follow. The tool has the potential of simulating various scenarios under different assumptions and conditions. Through the flexibility of available modules, it is possible to customize various parameters.

One of the functionalities that was developed for this simulation tool is to be able to deal with is multi-connectivity [2], which aims at being one the enablers for the wide multi-service behavior of the upcoming 5G networks. With multi-connectivity, a user may aggregate radio resources from more than two network nodes, thus allowing throughput and reliability to be noticeably increased.

7.2 Overall Functional Architecture

For the system level evaluation through simulations we need to consider certain aspects related to configuration, environment models, network (simulated system) models, analytics, event management and all of these in a user-friendly graphical user interface (GUI).

Environment models and configuration: An important aspect of system-level simulations is to specify the simulated system, designate the environments and select analytics. The "Environment models" concern aspects related to traffic (e.g. proper modeling of enhanced mobile broadband (eMBB), massive machine-type communications (mMTC) etc., anticipated load, mobility and radio conditions (e.g. propagation models).

Network (Simulated System) models: System aspects include considerations relevant to network deployment (e.g. small cells and macro cells etc.). Also, spectrum aspects are considered for utilization of bands. Abstraction of the physical (PHY) and medium access control (MAC) layers is taken into account. Radio resource management (RRM) algorithms are also considered.

Analytics: Results from the 5G system level simulator are evaluated against certain KPI targets (e.g. in terms of throughput or latency). The results are analyzed and visualized. KPIs and target values are well defined in various standardization bodies and organizations e.g. 3GPP, Next-Generation Mobile Networks (NGMN) etc. in order to reflect the requirements of different services.

Event Management: An event may be distinguished by time, location, type (e.g., session set up, call request, packet transmission), services, devices, users and supplementary info. Details on event management are provided later in this chapter.

Graphical User Interface (GUI): A user-friendly GUI is essential for easy handling of simulations and demonstrations. The GUI consists of user-friendly tabs, text boxes and input fields in order to create an easy-to-use environment for data input as well as extraction of results by visualizing results in graphs and charts.

7.3 Overview of the Environment Models

5G is poised to support a set of ambitious use cases as mentioned in [3]. For instance, use case families in NGMN include broadband access in dense areas (eMBB), massive Internet of Things (IoT) and mMTC as well as ultra-reliable low-latency communications (URLLC). For the needed representation/ modeling of such aspects, environment models shall take into account area aspects, traffic, mobility and propagation models based on the classification, which is depicted in Figure 7-1.

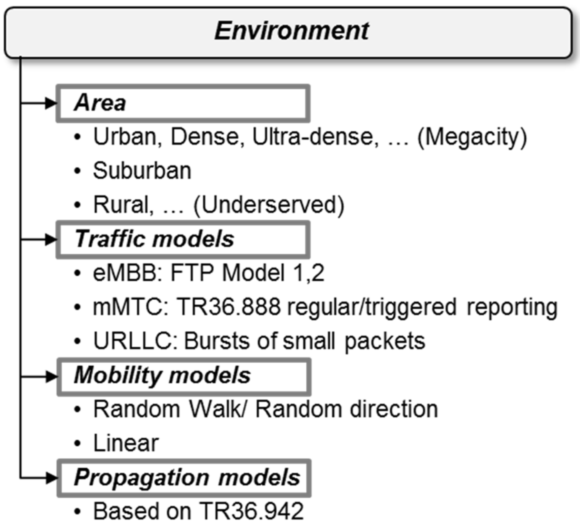


Figure 7-1: Environment aspects.

7.3.1 Area

An area can be characterized by its type, for instance of being Urban, dense or ultra-dense (e.g. capturing megacity requirements), suburban or even rural (e.g. capturing underserved areas). Different user and traffic densities are considered depending on the area type.

7.3.2 Traffic models

- *mMTC:* 3GPP has defined specific traffic models which consider bursty traffic with regular or triggered reporting. The system-level simulator considers three models of mMTC based on TR36.888 [4] and TR37.868 [5].
- *eMBB:* 3GPP has also defined file transfer protocol (FTP) models with inter-arrival rates following a Poisson process. The simulator considers FTP Model 1, 2, 3 for simulating eMBB-related traffic based on 3GPP TR 36.814 [8].
- *URLLC:* Bursts of small packets following beta distribution are considered according to [5][6].

7.3.3 Mobility models

The following mobility models where supported:

- *Random Walk:* according to this model each UE changes its speed and direction at each time interval. The default value of time interval is 1s, while these values can be configured from the simulator's graphical interface. For every time interval, the direction is chosen from (0, 2π], while speed follows a uniform or Gaussian distribution from [0, v_{max}].
- *Linear Motion:* each UE chooses randomly a direction and moves along it, with a constant speed. The UE continues moving in this direction even if it reaches the cell boundaries due to wrap-around.
- *Random Direction:* each UE chooses randomly a direction and moves along it, with a constant speed, until it reaches the boundary of the cell. Then, the UE chooses another direction to travel and moves with the same speed until it reaches the boundary of the cell again.

7.3.4 Propagation models

Different propagation models are taken into account depending on whether communication is taking place indoor or outdoor and depending also on the frequency of operation. Propagation models have been implemented based on [7].

7.4 Network/ Simulated System Models

7.4.1 Overview

A network/ simulated system model consists of access points, backhaul and core entity model. In this work, we particularly focus on the access part and the precise modeling of base stations, as depicted in Figure 7-2.

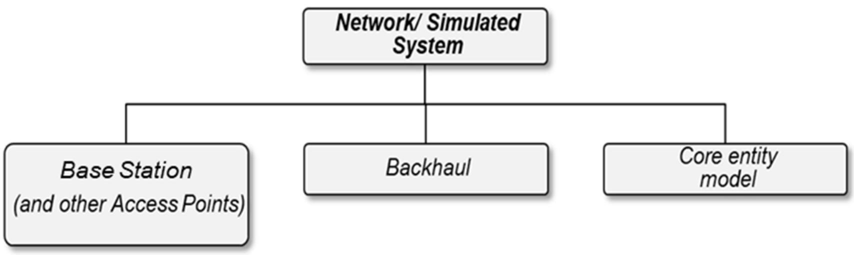


Figure 7-2: Network/ Simulated system models.

A base station can be characterized by various attributes as depicted in Figure 7-3. In particular, sectors have transceivers (TRXs) which can be characterized by spectrum aspects; PHY/MAC abstractions and RRM mechanisms. For instance, typical RRM algorithms such as round-robin, but also research algorithms as mentioned in [8] have been implemented in the system-level simulator.

7.4.2 Spectrum aspects

Spectrum aspects include information and implementations related to bands and carriers, policies (such as allowed carriers) and licensing schemes (e.g. licensed, light-licensed, License Assisted Access (LAA)-like usage of spectrum etc.).

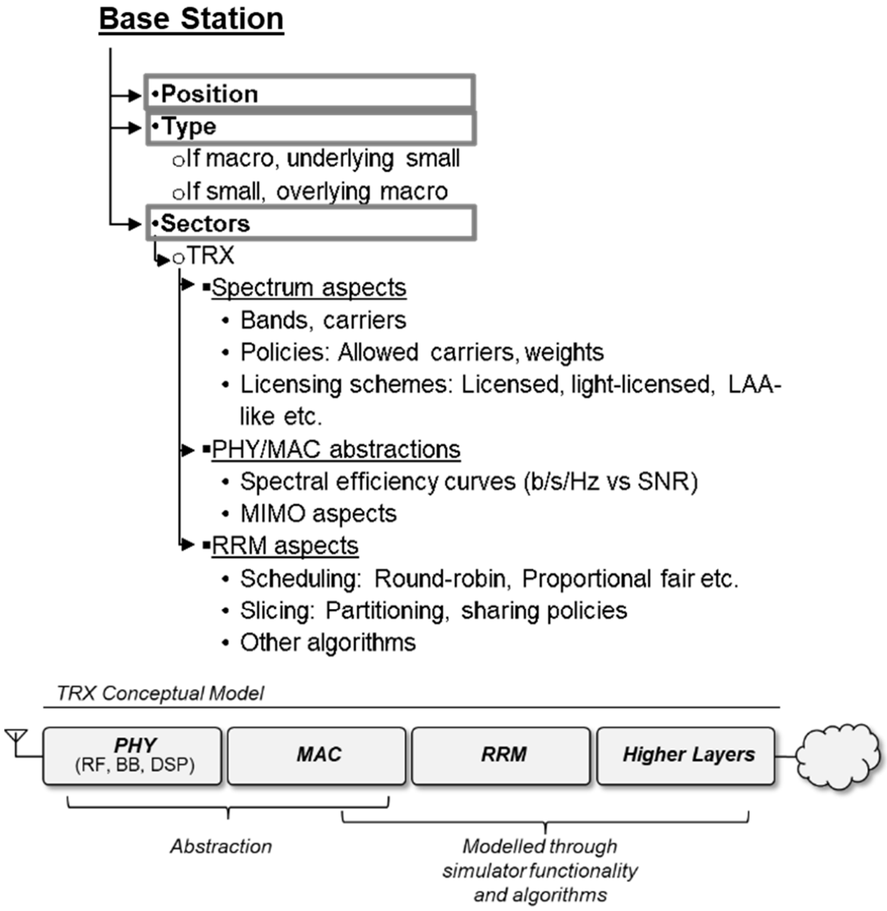


Figure 7-3: Base station main attributes and TRX conceptual model.

7.4.3 Abstractions of PHY/MAC layers

The evaluation of large networks, many devices, or a long time is hardly feasible for MIMO technologies. On the other hand, massive MIMO is a key-technology for 5G and evaluation on system or network level is essential. Therefore, an abstraction is necessary.

In [13] a PHY-layer abstraction model is proposed and adapted to fit in the 5G system-level platform. The information flow chart and principle components of the PHY-layer abstraction model are shown in 0 with a numerical example. First, the PHY layer MIMO simulation is performed for technology component A with parameter configuration according to the use case or scenario requirements. From these simulations, the following two outputs are required for the system level abstraction:

- The number of spatially multiplexed users per time-frequency resource
- The achieved user spectral efficiency over the geometry or SNR

In Figure 7-4 the cumulative distribution function (CDF) of the number of multiplexed user is given for output one. The user spectral efficiency over the geometry is given for output two. Note, that as an alternative to the geometry also the SNR can be used, e.g. if inter-cell interference in the PHY layer simulation is not explicitly simulated and considered as noise. The geometry here means the un-precoded sum power ratio of the serving BS antennas over all other BS antennas for a user.

In the second step, the system level or network layer simulation is performed assuming MIMO technology component A. Therein, (multiple) users on a time-frequency resource have to be selected according to the mapping table or curve from PHY layer simulation, where simplest case is to use always the median value. Note that the active user selection in the system level simulation may depend on traffic and mobility models. After the user selection, the SNR or geometry of these users is determined and used as input to the mapping from SNR or geometry to user spectral efficiency. Note, that the user spectral efficiency from the PHY layer MIMO simulation can also include the case that users have multiple antennas and maybe receive multiple data streams. With spectral efficiency of the active users, the sum spectral efficiency for the given time-frequency resource is obtained in the system level simulation for the given MIMO technology component A.

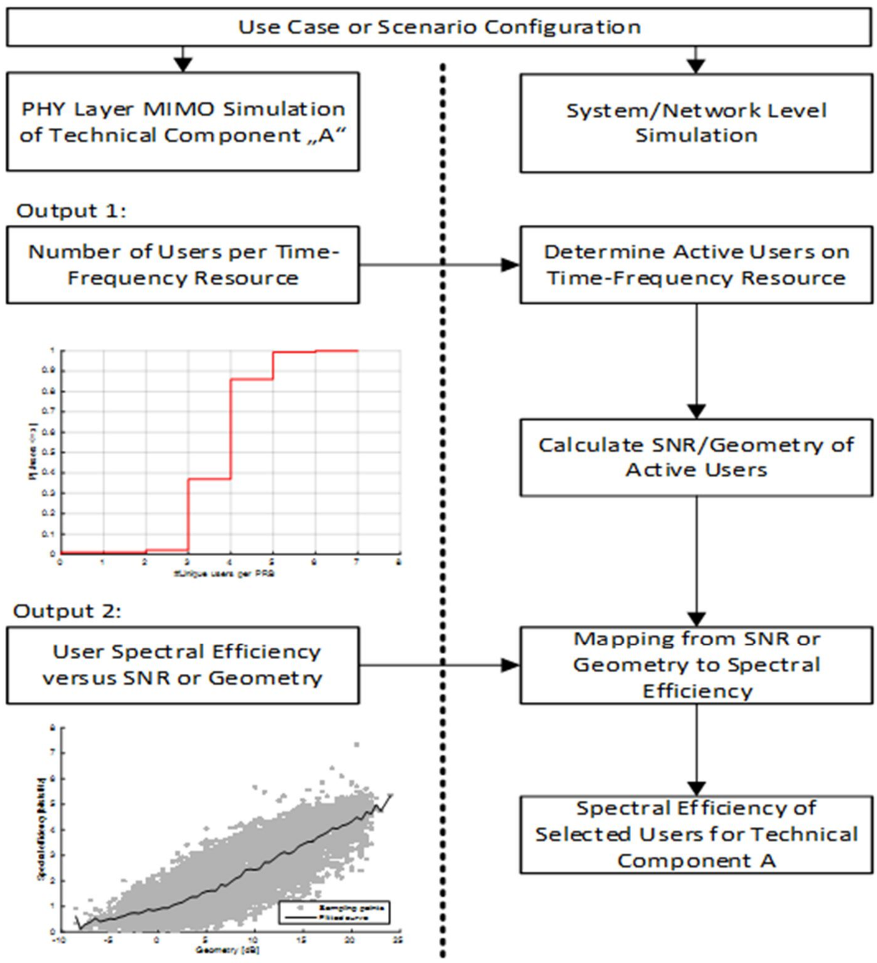


Figure 7-4: Physical Layer Abstraction of MIMO Technology Components for System Level Simulations.

7.4.4 RRM and higher layer mechanisms

In this part, aspects on scheduling, slicing and other RRM algorithms are considered. In terms of scheduling, round-robin scheduler is utilized in the simulator. With respect to slicing; partitioning and sharing policies are considered in order to allocate appropriate resources to slices according to the requested services. Furthermore, various RRM algorithms are implemented in order to evaluate different resource management strategies depending on the considered use cases.

7.4.5 Overview of supported simulator features towards NR

A set of new features coming from 3GPP latest releases needs to be appropriately modeled in order to be supported according to the latest specifications. Such features include among others:

- *Dynamic, slot-based frame structure:* Enables future-proof and ultra-lean design as well as self-contained subframe structure that allows for data transmissions that efficiently support diverse use cases with requirements that include low latency, high peak-rate, and high reliability.
- *Massive multiple input multiple output (MIMO):* control and data channel support for Massive MIMO features based on beam-centric design that improve spectral efficiency and achieve higher data rates, boosting performance for consumers.
- *mmWave communications:* Propagation models and spectral efficiency curves for mmWave bands can be incorporated.
- *Channel codecs:* channel coding schemes based on latest technology in advanced low-density parity-check (LDPC) codes to support large data blocks and extreme peak rates.
- *Support for network slicing:* Monitoring and slice visualization capabilities. Base station elements can have certain capacity and allocated resources (e.g., channels/ bandwidth) for different types of services in order to allocate different resources to eMBB, URLLC slices etc. Each user will make use of the intended slice based on the requested service/traffic type.
- *Support for edge computing:* Traffic can be served from points nearer to the end-user. Latency related to core network for accessing distant servers can be minimized.
- *Network sharing:* Simulator supports multi-operator environment and in case of network sharing, traffic of two or more operators can be served from a single infrastructure. However, increased traffic load is foreseen in the shared infrastructure.
- *Packet-splitting support*: Packets are split to multiple segments in order to be able to send different segments to different cells. As such, throughput may be improved by assuming that shorter packets are sent to better quality links.
- *Multi-connectivity support*: The end-user is connected to more than one network node in order to enable multi-link transmission. This can be used in conjunction or not with packet-splitting support. Reliability may also be improved if multiple links are available.

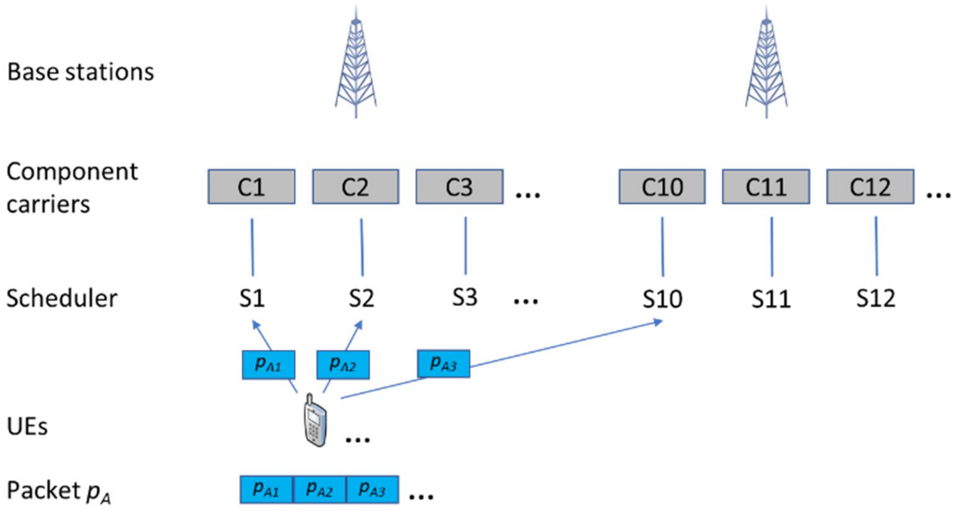


Figure 7-5: Multi-connectivity with packet splitting and carrier aggregation.

- *Carrier Aggregation Enhancement:* Carrier aggregation is used in order to increase the bandwidth, and thereby increase the bitrate. Each aggregated carrier is referred to as a component carrier (CC). The CC can have a bandwidth of 1.4 to 20 MHz and a maximum of five component carriers can be aggregated, hence the maximum aggregated bandwidth is 100 MHz for Long-Term Evolution (LTE). For supporting New Radio (NR) network, the CC can reach up to 400MHz for bands of higher than 6GHz in NR-related bandwidths and up to 16 component carriers.

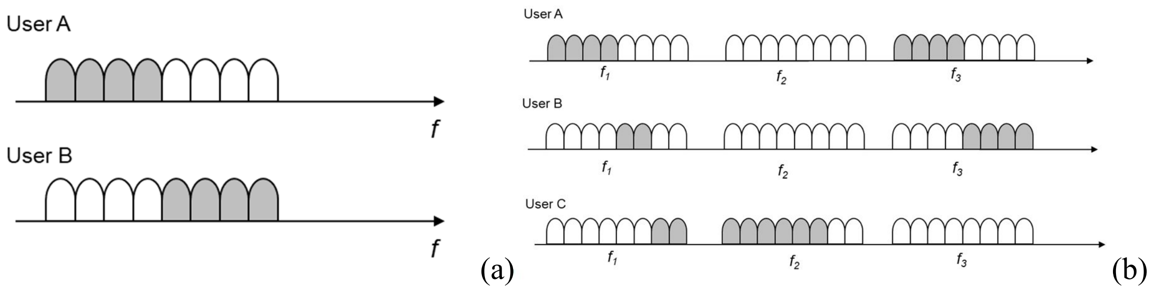


Figure 7-6: Utilization of RBs between users, Intra-band (a), Inter-band (b).

- *Enhancement to Narrowband IoT:* NB-IoT technology is deployed "in-band" in spectrum allocated to LTE, using resource blocks within a normal LTE carrier.
- *CoMP:* Coordinated multipoint (CoMP) is used to send and receive data to and from an end user from several points to ensure the optimum performance is achieved even at cell edges.

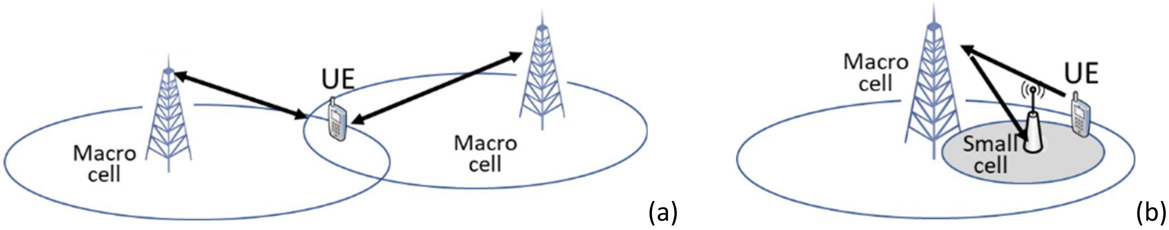


Figure 7-7: Multi-connectivity with CoMP enabled for UEs.

- *NOMA:* The key idea of Non-Orthogonal Multiple Access (NOMA) is to use the power domain for multiple access, whereas the previous generations of mobile networks have been relying on the time/ frequency/ code domain. Increased spectral efficiency with NOMA compared to orthogonal frequency division multiple access (OFDMA) can be achieved through various references in literature.

7.4.6 Event handling

It is important to define an event management mechanism, which is characterized by distinct phases. Such phases can be initialization; event extraction; trigger event handling; trigger event scheduling and consolidation of report regarding performance. The phases are illustrated also in the flowchart that is illustrated in Figure 7-8 and will provide a correct simulation of the mechanism that are developed in order to check performance benefits of multi-connectivity strategies as well as the service quality gains that every user in our simulations will have.

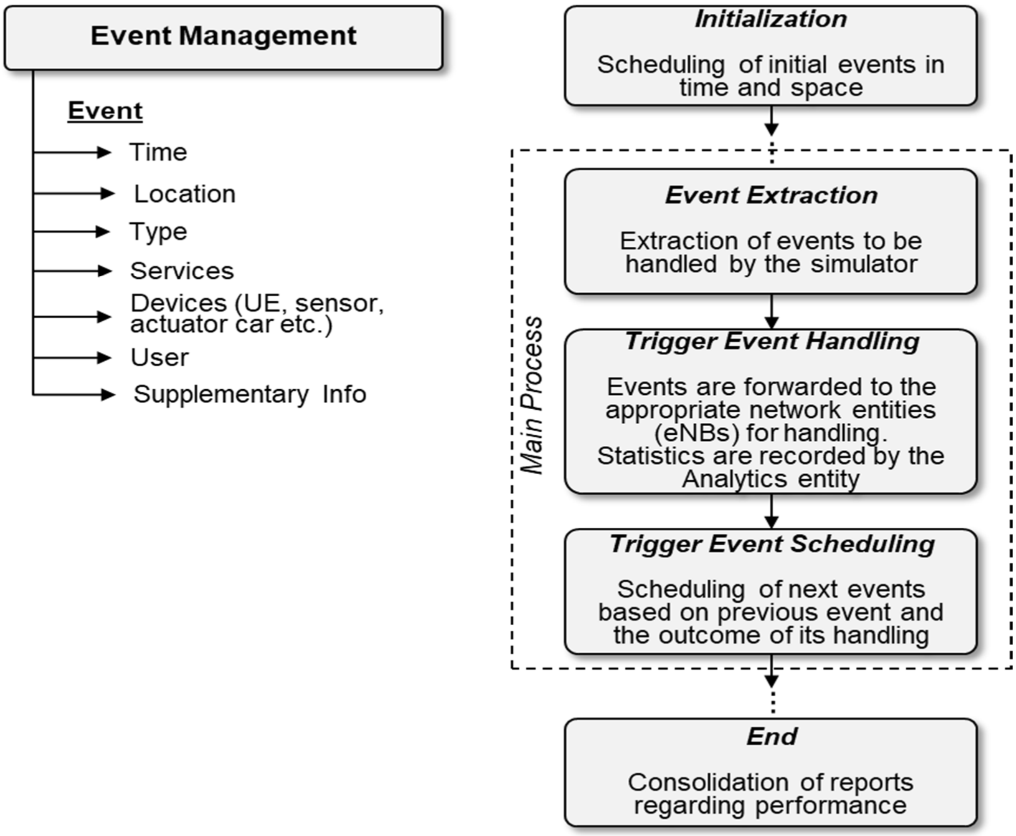


Figure 7-8: Process description for handling of events in the simulator.

Based on NGMN in [3] and the International Telecommunication Union (ITU) in [10] certain KPIs have been defined related to user experience and system performance. User experience KPIs involve mainly data rate in downlink and uplink, latency (associated with the radio access) and mobility levels. These KPIs are covered by the system level simulator. With respect to system performance KPIs, these involve mainly connection density, traffic density, spectrum efficiency, coverage. Spectrum efficiency in the system level simulator is mainly covered by mapping curves which are usually created through link-level simulations and plot bit/s/Hz to signal-to-noise ratio (SNR). With respect to connection density and traffic density attributes, these are related also to the area under investigation (e.g. dense urban, rural/underserved etc.).

7.5 Initial Results

The system-level simulation platform has been calibrated against the reference results of the 3GPP LTE calibration campaign [8]. In this subchapter, we present the Cumulative Distribution Function (CDF) of coupling loss for 3GPP case 1 – 2D and 3D scenarios. The results of the calibration process for the aforementioned scenarios are depicted in Figure 7-9 and Figure 7-10 respectively.

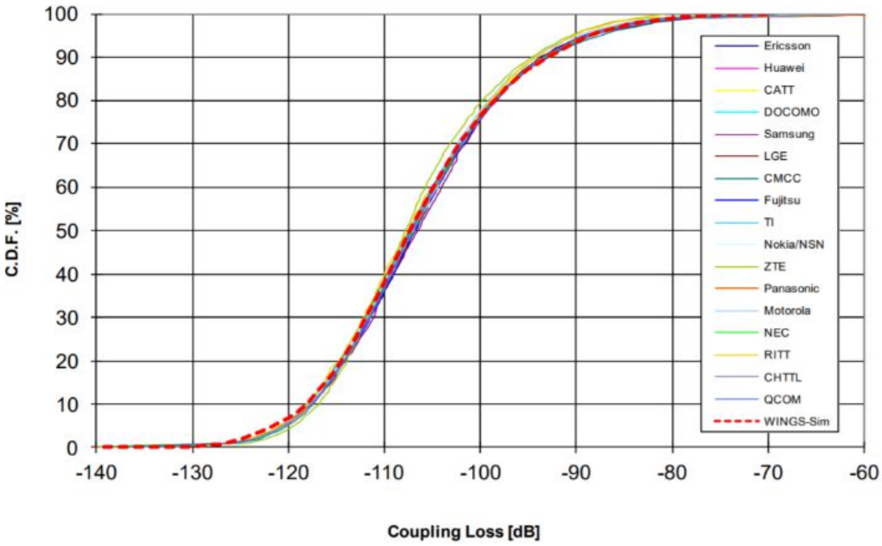


Figure 7-9: Coupling Loss – 3GPP case 1 – 2D scenario.

Moreover, Figure 7-11 illustrates the capability of the simulator to assess different RRM algorithms and test cases (in this case we assume algorithm with QoS priority, a state-of-the-art RRM algorithm and a random RRM algorithm) as mentioned in [9]. Values of normalized throughput were calculated for various test cases.

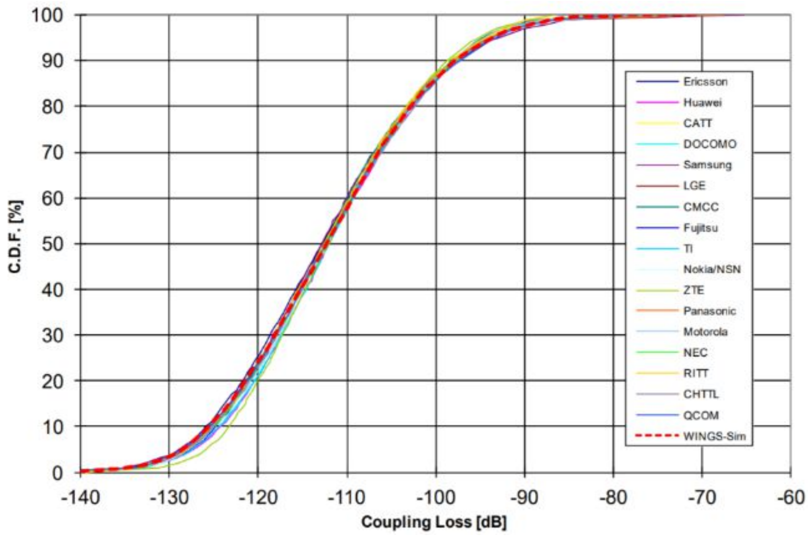


Figure 7-10: Coupling Loss – 3GPP case 1 – 3D scenario.

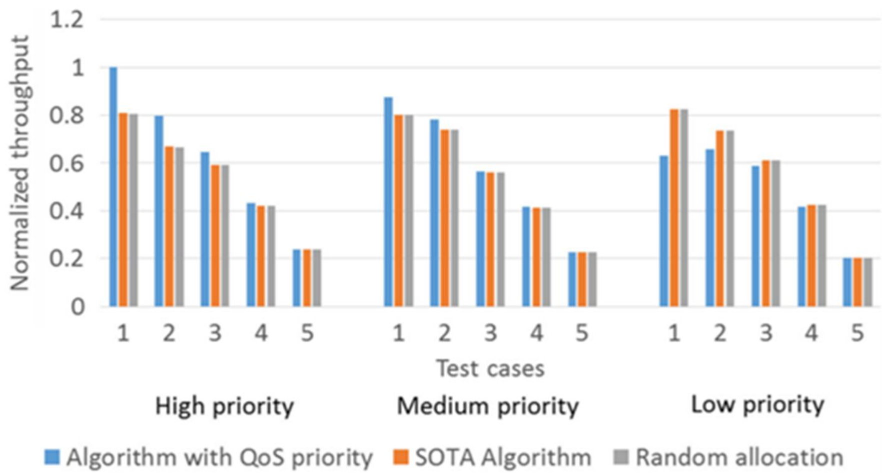


Figure 7-11: Normalized throughput measured for various algorithms and test cases.

Figure 7-12 illustrates the protocol overhead in terms of average number of retransmissions which have been calculated through the simulator. The figure shows that the one-stage pooled protocol has the lower number of retransmissions for medium arrival rates, while the one-stage protocol (20db) has the greatest performance for high arrival rates. In contrast to ARP with many retransmissions and high overhead, all the proposed protocols show low overheads. In most cases, the MTC devices are low power (use of batteries) and therefore their lifetime is highly affected by the transmission phase. Therefore, another important metric is the total number of transmissions per data packets. In the case of one-stage protocols the transmission of preamble and data are realized in one burst, therefore the number of transmissions is minimized. From this figure, it becomes clear that one-stage protocols are more appropriate for low-power devices since they manage to minimize the total transmissions (low number of retransmissions and simultaneously transmission of control and data information) and therefore the total energy consumption. On the other hand, two-stage protocols generate more transmissions, and in order to become comparable to one-stage protocols should retain the retransmission numbers in very low values.

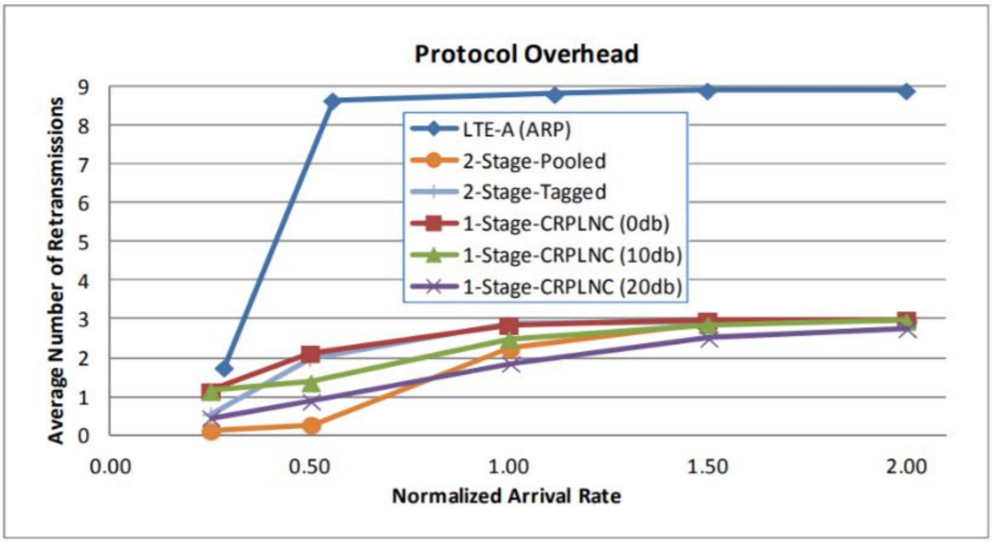


Figure 7-12: Evaluation of massive access protocol overhead.

7.6 Component Carrier Management Aspects

In this work, a novel RRM functionality [10] for CC management is developed, integrated and assessed. This functionality is referred to as the component carrier manager (CCM) and has been devised for the enhancement of performance metrics, such as the UE throughput (for eMBB services) or the connection reliability (for URLLC services) through the management of multiple links in a multi-connectivity environment. Different number of component carriers can be aggregated for the downlink and uplink. This is an important property from a device complexity perspective where aggregation can be supported in the downlink where very high data rates are needed without increasing the uplink complexity. Component carriers do not have to be contiguous in frequency, which enables exploitation of fragmented spectra; operators with a fragmented spectrum can provide high-data-rate services based on the availability of a wide overall bandwidth even though they do not possess a single wideband spectrum allocation. In addition, macro and small cells as illustrated in Figure 7-14 can work together and carriers can be chosen from both cells with different frequencies based on a specific methodology.

Initially a device will be connected to a nearby cell tower. Once that a UE has been assigned a primary cell (PCell), and thus, a master node (MN), according to 3GPP-compliant mobility criteria, an instance of the CCM is run for this UE in this MN. From the CCs that the UE receives with a minimum power, the CCM identifies the subset of CCs to be assigned to this UE as primary secondary cells (PSCells) and secondary cells (SCells). With the former, the CCM eventually performs a UE-to-network node association. The latter (SCells) may be linked either to the MN or to a given SN to extend the available bandwidth between the UE and that network node.

To identify the subset of CCs to be assigned to a UE, the CCM computes a score for every available CC, according to some operator's policy. The top scoring CCs are then assigned to this UE, hosting either a SCell (if they belong to the MN or a previously assigned SN) or a PSCell, if they are the first CC assigned to this UE belonging to a node different from the MN. In [10], the scoring method is implemented a rule-based system.

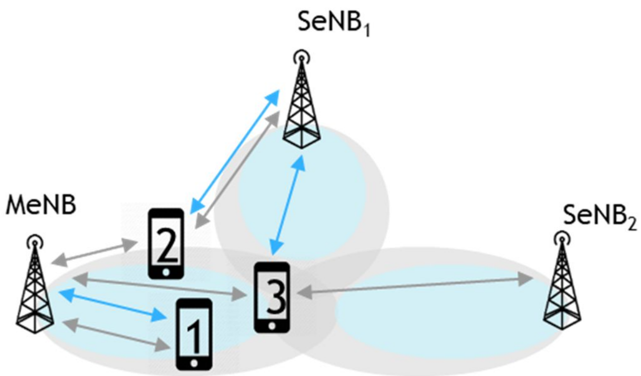


Figure 7-13: Multi-link/Multi-node connectivity.

7.6.1 Evaluation and Results

In this section, a proof of concept is carried out to show the benefits of using the CCM in a multi-connectivity environment using the simulation tool that has been described. First, a typical optimization use case in which to use the CCM is described, a network load imbalance, together with the simulation setup. Then, the simulation results are shown and conclusions are drawn.

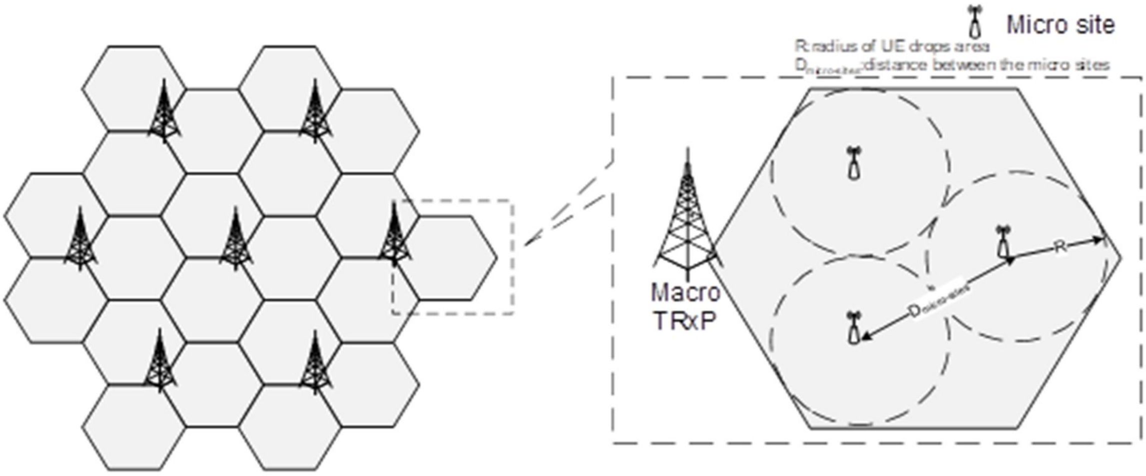


Figure 7-14: Cellular topology (ITU-R M.[IMT-2020.EVAL], “Guidelines for evaluation of radio interface technologies for IMT-2020”).

7.6.2 Experiment setup

Often, users tend to gather around interest points, causing a non-homogenous traffic distribution along the cellular network Figure 7-15. As a consequence, this causes a network load imbalance, in which some network nodes deal with a big number of users, thus lacking in radio resources, and some other nodes remain almost unused. In order to take advantage of the full network capacity, cell edge users are then forced to hand over to non-congested cells, despite these are not the cells providing the best received signal quality/power. This is the optimization use case of load balancing and has been addressed in many different ways along the years. The most common way to tackle this issue has been the mobility load balancing (MLB); that is, the adjustment of certain mobility configuration parameters, such as the handover margins, to manipulate the cell service area [11].

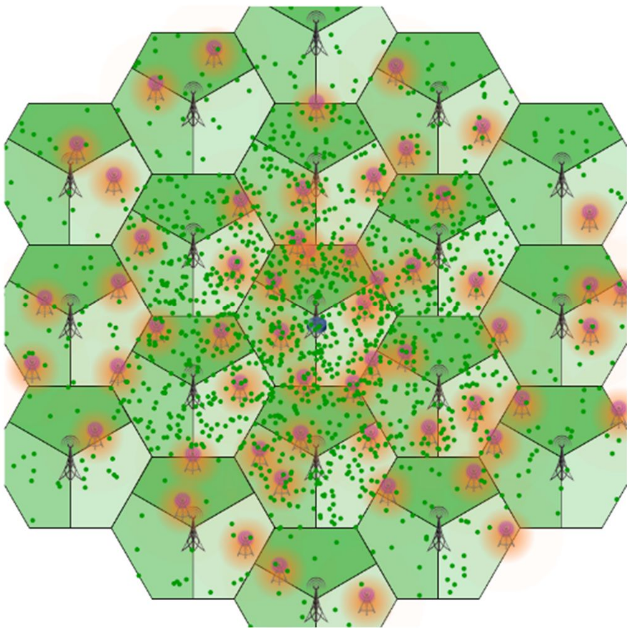


Figure 7-15: Load imbalance topology.

With the CCM, however, this task can be addressed by acting on the network access procedure. To that end, the CCM is designed so that the assigned CCs contribute to maximize every user's throughput. This can be achieved if the score of each CC is computed according to both its current load (e.g., either in terms of used physical resource blocks, number of currently scheduled users, etc.) and the signal quality perceived by the user, according to the well-known Shannon's theorem. The main metric that has been considered is throughput.

This is compared with a multi-connectivity baseline situation in which cells are assigned to UEs only according to their reference signal received power (RSRP), as an extension to multi-connectivity of current 3GPP policies for dual-connectivity and carrier aggregation management.

The scenario described in this subchapter extends that of [10] in two main aspects. First, an FTP service traffic model has been followed over connections governed by the transmission control protocol (TCP). In [11], however, a finite buffer traffic type on the MAC layer was used, lacking in more realistic effects, like retransmissions from higher layers. The second differencing aspect with respect to [10] is that, in this work, a heterogeneous network made up of both macrocells and small cells is considered. This allows further enhancing the benefits of load balancing due to the tradeoff between the coverage of macrocells (usually acting as MNs) and the capacity of small cells (usually acting as SNs). In [10], however, a macrocell scenario is considered.

In order to simulate the load imbalance, more than half users have been placed around certain network nodes, representing most of the total offered traffic. We consider only eMBB traffic and selecting cells with the criterion of Reference Signal Received Power (RSRP). The proposed solution is based on signal quality (RSRQ) and the load of candidate component carriers. Different types of inputs are considered, such as: (a) metrics reported by the user, like the Reference Signal Received Quality (RSRQ) and (b) metrics from the carriers (like their load). Based on these inputs, the CC manager computes a score for each of the available carriers indicating the carrier suitability for a specific user. This score can be computed in different ways depending on the target criterion (e.g., if a load balancing approach is followed, those CC with a lower load will receive a higher score; in the case of a target focused on signal quality, CC with higher RSRQ will be selected). Component Carrier (CC) manager is proposed to determine the number of carriers to be assigned to a user as already discussed. This CC manager could be implemented in the gNodeB and necessary information could be exchange by gNodeB by using Xn interfaces. Additionally, the carrier indices, the source nodes, and flow are also proposed by the CC manager.

Table 7-1 summarizes the main configuration parameters that have been used in the simulations. Two cases have been considered. In case a), 1 MB-files are sent to the UEs through a FTP downlink connection. In case b), the size of these files is 8 MB.

Table 7-1: Simulation parameters.

Parameter	Value
Number of Macro BS	19 macro 3-sector base stations
Number of Small BS	57 small base stations
Number of users	1000 users
Network area	2200x2200 meters
ISD	500 meters for macros
Frequencies	2GHz
Request interarrival time	Poisson
Traffic data generation	1440 files per user per day

File Size	1MB (case a), 8 MB (case b)
Packet Splitting	Enabled
Simulation time	60 sec
Bandwidths	20MHz
Subcarrier spacing	15KHz
Carrier aggregation	Enabled
Component Carriers	1 to 7
Propagation model	L=128.1+37.6log10(R), R in km
FTP direction	Downlink

7.6.3 Results

Figure 7-16 and Figure 7-17 show the UE downlink throughput achieved in case a) and case b), respectively, for an increasing number of CCs eventually assigned to UEs following the baseline case (blue line) and using the CCM (orange line). Both figures show that, given a load imbalance, the proposed approach allows to improve the users' throughput regardless the number of assigned CCs given the joint consideration of each CC RSRQ and load. In particular, a maximum throughput gain of 40% with respect to the baseline is achieved in case a) when 2 CCs are assigned and a 60% gain is achieved when 3 CCs are considered in the 8-MB case, which shows the potential of the proposed framework for multi-connectivity management.

Figure 7-16: Impact of CCs number to throughput for 20MHz bandwidth and 1MB file size (case a).

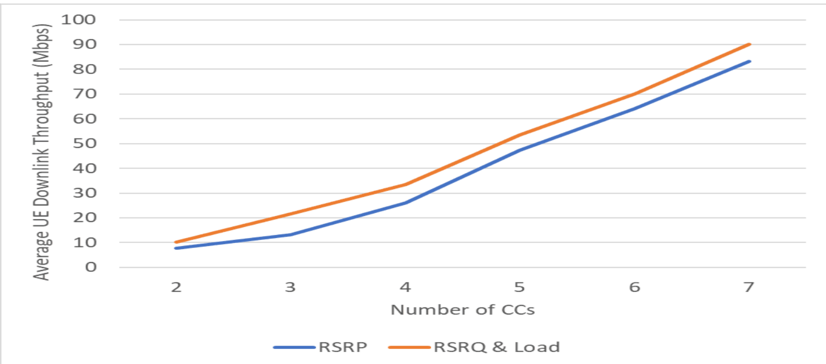


Figure 7-17: Impact of CCs number to throughput for 20MHz bandwidth and 8MB file size (case b).

Finally, Figure 7-18 illustrates the result of throughput by considering different number of component carriers when using higher bandwidths of up to 100MHz (for taking into account 5G assumptions of higher bandwidths) for each CC. Similar to previous results, for all values of CC, the RSRQ & load approach achieves slightly higher throughput compared to the RSRP .

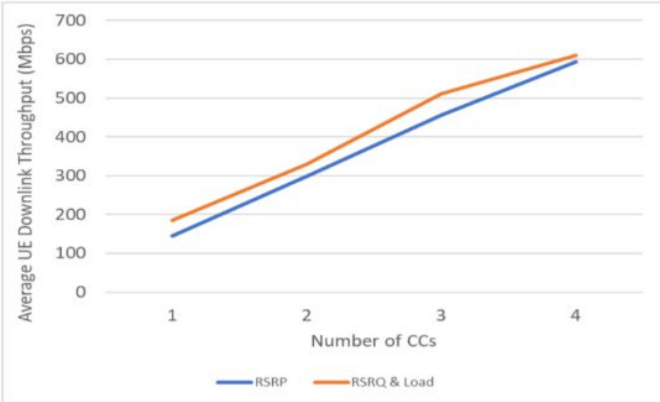


Figure 7-18: Impact of number of CCs to throughput for a bandwidth of 100MHz and file size of 8MB.

7.7 Introduction to connection density

Applications in massively connected sensors or robots require reliability and low latency. Connection density in 5th generation networks can be used in big megacities or factories of the future (Figure 7-19). Real-time feedback, control of the production machinery and process, and diagnosis, error/fault identification and calibration/ recovery of the machinery. In FoF many processes exist and each of them is important for the whole manufacturing process. Thus, sensors should be deployed throughout the processes to monitor, analyze and predict potential problems inside each process and of course to optimize the manufacturing process. Those processes have a large number of equipment that may introduce problems and thus should be monitored through sensors especially in the scope of the factory of the future where automation is an important part of the process in conversion of a factory of our days into a smart factory. The network connects and retrieve information from various processes, sensors and equipment which are active in the production line to analyze network and send control commands to stop/change/fix the process and robots to minimize losses or damages to equipment or even human injuries. Multiple equipment such as robotic arms, conveyor belts, assembly/scraping machine etc., should be controlled. The past years GSM/EDGE had a latency of 150ms, HSPA/HSPA+ had a lower latency of about 50ms while nowadays LTE has reached 20ms of edge-to-edge latency and cannot support a great number of devices in a small area. The next generation 5G communications would be capable to support the expected revolution in manufacturing, considering the stringent industry-specific requirements as presented above.

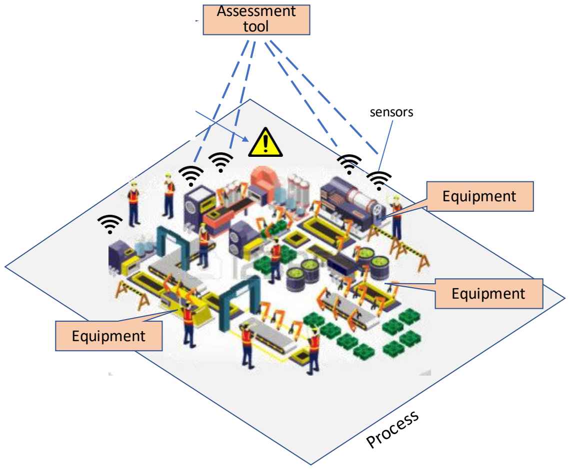


Figure 7-19: Sensors, equipment and robots in FoF

7.8 Evaluation Methodology and KPIs

In mMTC environments, one of the important parameters is the connection density of devices. According to ITU document [14] [15] [16] the connection density is the total number of devices fulfilling a specific quality of service (QoS) per unit area (per km^2). Connection density should be achieved for a limited bandwidth and number of connectivity points. The target QoS is to support delivery of a message of a certain size within a certain time and with a certain success probability. This requirement is defined for the purpose of evaluation in the mMTC usage scenario. According to ITU, the minimum requirement for connection density is 1,000,000 devices per km^2.

Also, ITU has defined the following steps, for the evaluation of connection density:

- Step 1: Set system user number per TRxP as N.
- Step 2: Generate the user packet according to the traffic model.
- Step 3: Run non-full buffer system-level simulation to obtain the packet outage rate. The outage rate is defined as the ratio of the number of packets that failed to be delivered to the destination receiver within a transmission delay of less than or equal to 10s to the total number of packets generated in Step 2.
- Step 4: Change the value of N and repeat Step 2-3 to obtain the system user number per TRxP N' satisfying the packet outage rate of 1%.
- Step 5: Calculate connection density by equation C = N' / A, where the TRxP area A is calculated as A = ISD2 × sqrt(3)/6, and ISD is the inter-site distance.
- The requirement is fulfilled if the connection density C is greater than or equal to 1,000,000 The simulation bandwidth used to fulfill the requirement should be reported. Additionally, it is encouraged to report the connection efficiency (measured as N' divided by simulation bandwidth) for the achieved connection density.

The considered traffic model for such an evaluation is message size of 32 bytes with either 1 message/day/device or 1 message/2 hours/device. Packet arrival follows Poisson arrival process for non-full buffer system-level simulation.

Baseline evaluation configuration parameters and additional parameters for system-level simulation are presented at Table 7-2.
Table 7-2: mMTC parameters for connection density evaluation

Parameters	Values
Carrier frequency for evaluation	700 MHz
BS antenna height	25 m
Total transmit power per TRxP	49 dBm for 20 MHz bandwidth 46 dBm for 10 MHz bandwidth
UE power class	23 dBm
Percentage of high loss and low loss building type	20% high loss, 80% low loss
Inter-site distance	500 m
Number of antenna elements per TRxP	Up to 64 Tx/Rx
Number of UE antenna elements	Up to 2 Tx/Rx
Device deployment	80% indoor, 20% outdoor, randomly and uniformly distributed over the area
UE mobility model	Fixed and identical speed |v| of all UEs of the same mobility class, randomly and uniformly distributed direction.
UE speeds of interest	3 km/h for indoor and outdoor
Inter-site interference modelling	Explicitly modelled
BS noise figure	5 dB
UE noise figure	7 dB
BS antenna element gain	8 dBi
UE antenna element gain	0 dBi
Thermal noise level	-174 dBm/Hz

7.9 Simulation Results

System-level simulations have been conducted for the evaluation of connection density in mMTC environments. Narrowband parameters are taken into account in the simulation. As such, considered bandwidth is from 180KHz up to 1.08MHz. The success rate (i.e. successful transmission of messages) is calculated in order to check the acceptable level of connection density for meeting the threshold of 99% of success (1% of loss). During the evaluation process, the lower number of the considered message generation frequency (e.g. 1 message/day/device) fulfills the requirements of the connection density. The results showed that the 99th percentile of the delay per user was less than 10s for both the 180KHz and 1.08MHz tests. Two configurations for ISD of 500m and 1732m were examined during the evaluation. As a result, the focus was given on the investigation and analysis of the higher message frequency of 1 message/2 hours/device which had a different behavior than the previous.

Figure 7-20 shows the success rate for different number of devices when bandwidth of 180KHz is used. According to the results, it is evident that with such bandwidth, up to 2 million devices per km2 assuming messages of 32 bytes and 1 message/2 hours/device can be served. When 3 million devices per km2 were simulated, the success rate dropped below 99%.

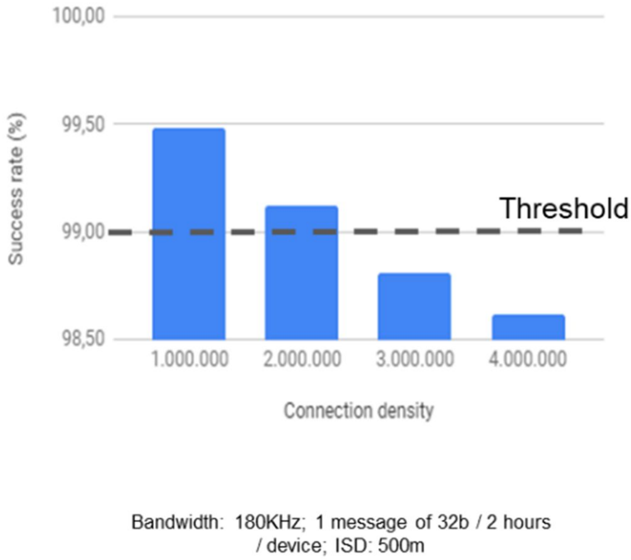


Figure 7-20: Connection density (nr. of devices per km^2)

Figure 7-21 shows how much bandwidth is needed for serving 1 million devices with 1 message of 32 bytes/2 hours/device as we have seen at Figure 1, but this time by examining at which level the success rate will reach the highest level. The results show that even from 180KHz, the success rate of 99% is fulfilled and as the bandwidth increases, the success rate is even higher, reaching almost the 100% at 540KHz.

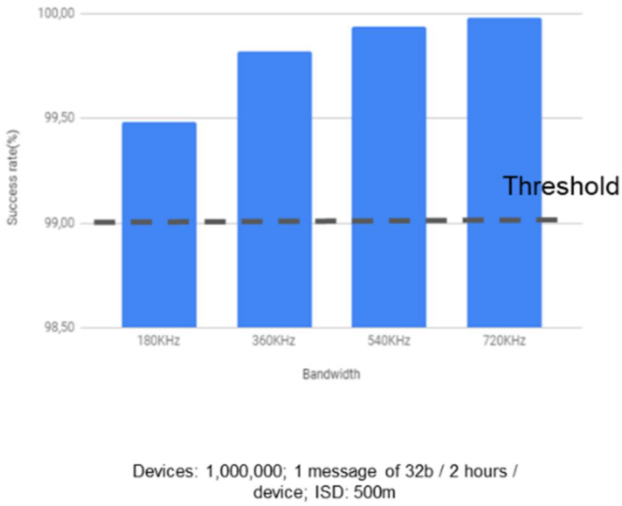


Figure 7-21: Success rate depending on bandwidth (ISD 500m)

As a next step we changed the simulation parameters to higher ISD value of 1732m and run the same evaluation process as before. Figure 7-22 shows the success rate for different number of devices when bandwidth of 1.08MHz is used. According to the results, it is evident that with such bandwidth, up to 40 million devices per km2 can be enabled in the area without serious problems assuming messages of 32 bytes and 1 message/2 hours/device.

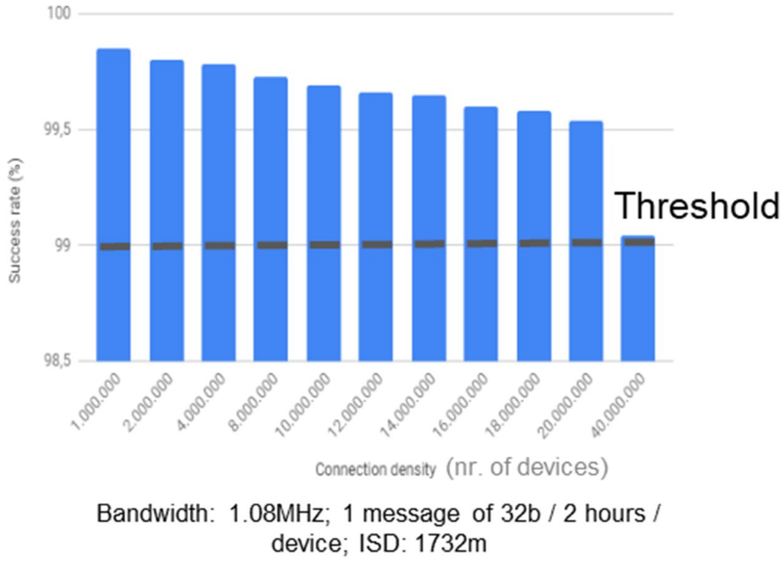


Figure 7-22: Connection density (nr. of devices per km^2)

Figure 7-23 shows how much bandwidth is needed for serving 1 million devices with 1 message of 32 bytes/2 hours/device. The results show that from 500KHz and above, the success rate of 99% is met. However, smaller bandwidths (e.g. 180 or 360KHz) are possible but the success rate is a bit lower than 99%.

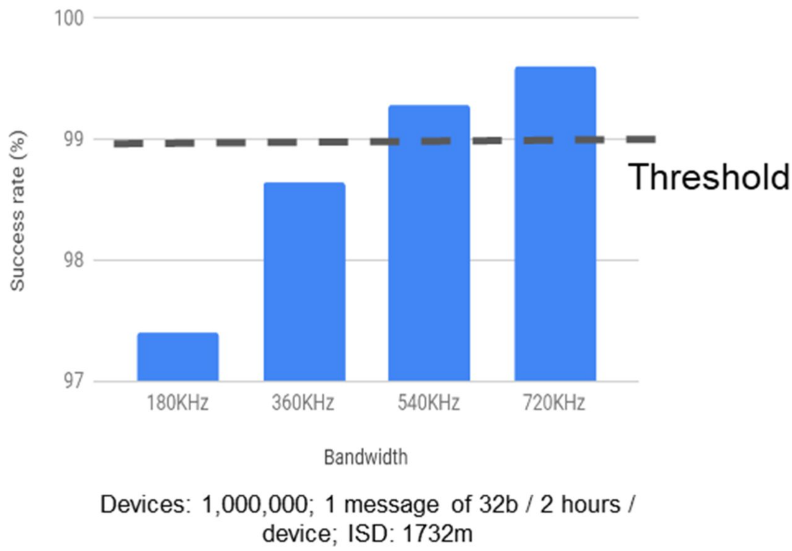


Figure 7-23: Success rate depending on bandwidth (ISD 1732m)

7.10 Architecture, interface and mechanisms

New Radio and 5G networks need to support large numbers of concurrent MTC connections in a specific area as already presented. Figure 7-24 shows the steps that the proposed algorithm takes to ensure that high levels of connection density is achieved. The proposed mechanism could run on each base station and configure its resources so that minimum resources are used when MTC connections are under a specific and recommended limit. If the numbers are beyond the limit (e.g., over 1 million users) then the mechanism can reconfigure each BS in order to satisfy the packet outage of 1% that is acceptable and stabilize the connections to all MTC devices.

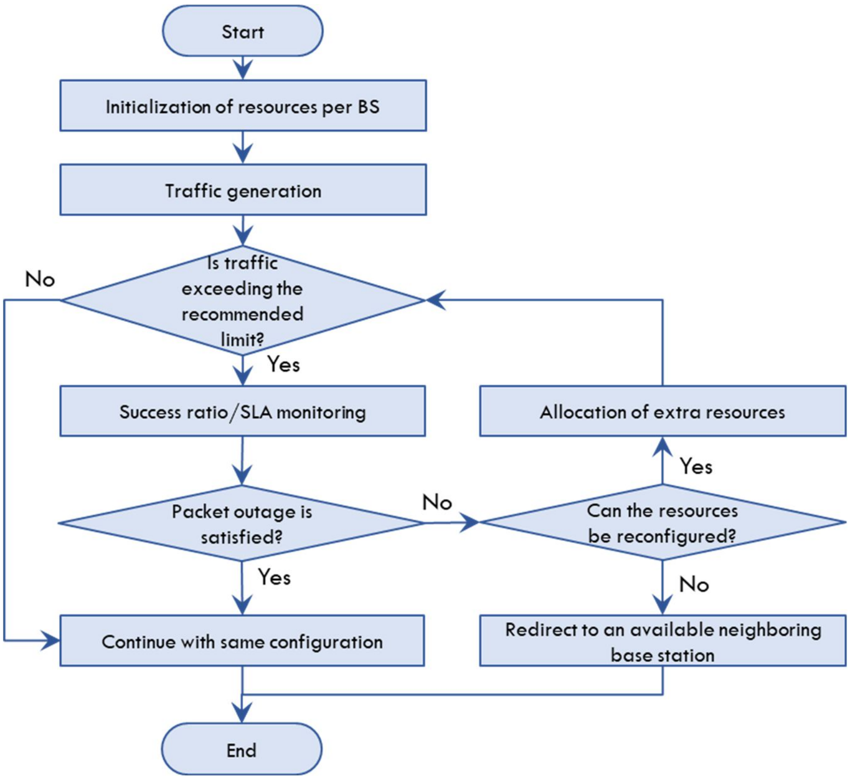


Figure 7-24: Flow chart of proposed algorithm

7.11 Conclusion

This work elaborated on the presentation of the framework for multi-connectivity, being one of the functionalities to be used to deal with the dissimilarity of service requirements of 5G networks. Simulation results show how a proper assignment of component carriers (CCs) in this situation allows increasing the users' throughput by up to a 60% when compared to a simple received power scheme for link management. Connection density also plays an important role on 5G and beyond environments. The usage of narrowband technologies is encouraged, especially for small and frequent transmissions. As a result, the provided evaluations consider these assumptions to show the number of devices that can be supported with a specific QoS.

In cases of 180KHz of bandwidth the scenarios of ISD at 500m showed that there were not any major problems for the device density that was considered. In addition, the results for ISD of 1732m reveal that there is a need of higher bandwidths to meet the requirements and achieve the proposed success rates, which in many cases more than three times the initial bandwidth had to be used.

It should be noted that the results are consistent with results of vendors (such as Huawei and Ericsson) who followed the same evaluation process, utilizing the same parameters at their proprietary simulator. For the bandwidth of 1.08MHz the evaluation process showed that it is possible to handle effectively more than 1 million devices per km2 in every situation

7.12 References

[1] Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos, MIsabel de la Bandera-Cascales, David Palacios, Raquel Barco, Panagiotis Demestichas, “5G Component Carrier Management Evaluation by Means of System Level Simulations”, in Proc. 2019 European Conference on Networks and Communications (EuCNC), Valencia, 18-21 June, 2019

[2] 3GPP, TS 37.340, “NR; Multi-connectivity; Overall description; Stage-2” Rel-15, V15.0.0, Dec. 2017.

[3] NGMN, “5G White Paper”, available online at: https://www.ngmn.org/5g-white-paper/5g-white-paper.html

[4] 3GPP TR 36.888 “Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE”

[5] 3GPP TR 37.868 “RAN Improvements for Machine-type Communications”

[6] 3GPP TS 23.501 “URLLC traffic model and QoS parameter”, S2-178901, SA WG2 Meeting #124

[7] 3GPP TR 36.942 “Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios”

[8] 3GPP TR 36.814 “Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects”

[9] I.-P. Belikaidis, S. Vassaki, A. Georgakopoulos, A. Margaris, F. Miatton, U. Herzog, K. Tsagkaris, P. Demestichas, “Context-aware radio resource management below 6 GHz for enabling dynamic channel assignment in the 5G era”, EURASIP Journal on Wireless Communications and Networking (2017) 2017:162.

[10] ITU, Revision 1 to Document 5/57-E, 17 October 2017

[11] N. H. Mahmood, D. Laselva, D. Palacios, M. Emara, M. C. Filippou, D. Min Kim, I. de-la-Bandera, "Multi-channel access solutions for 5G New Radio" in *IEEE Wireless Communications and Networking Conference* (WCNC), 2019, Marrakech. Accepted for publication.

[12] P. Muñoz, R. Barco, J. M. Ruiz-Avilés, I. de la Bandera and A. Aguilar, "Fuzzy Rule-Based Reinforcement Learning for Load Balancing Techniques in Enterprise LTE Femtocells," in IEEE Transactions on Vehicular Technology, vol. 62, no. 5, pp. 1962-1973, Jun 2013.

[13] Y. Wu et al., “Green transmission technologies for balancing the energy efficiency and spectrum efficiency trade-off,” IEEE Communications Magazine, vol. 52, no. 11, pp. 112–120, Nov. 2014.

[14] ITU, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)" (Document 5/40)

[15] ITU, "Guidelines for evaluation of radio interface technologies for IMT-2020" (Document 5/57)

[16] ITU, "Requirements, evaluation criteria and submission templates for the development of IMT-2020" (Document 5/56)

8 Conclusion

This dissertation provided in depth analysis of solutions for managing networks and services based on artificial intelligence in heterogeneous broadband environments for 5th generation and beyond by considering the current status and looking forward to the emerging challenges. Specifically, mesh networks have been investigated by proposing algorithms for examining the optimum settings (position, density, transmission ranges) in the flexible and dynamic topologies based on moving access points (MAPs) environment. By testing different scenarios we tried to deduce what would be the best case scenario for the utilization of these algorithms. Our simulation showed that the proposed algorithms were able to choose the best possible paths in a short time period (i.e., fewer iterations) in almost all scenarios. An interesting direction resides in devising more sophisticated approaches towards a better selection from the infectors vicinity, e.g., by applying machine learning.

Another important topic that the dissertation dealt with had to do modeling and analysis of management in heterogeneous networks. This chapter elaborated on the status & challenges in hardware/ software development and in 5G wireless communications by focusing on MAC and RRM layers. Also the benefits of machine learning in 5G network management were discussed. By taking into account the diversity of infrastructure, radio resources and services that will be available in 5G, an adaptive network solution framework will become a necessity. Breakthrough developments in several RAN technologies will be required for realizing novel, 5G solutions. Such technologies include among others, multiple access and advanced waveform technologies combined with coding and modulation algorithms, massive access protocols, massive MIMO and virtualized and cloud-based radio access infrastructure. In addition, aspects of radio frequency resource management in a multi-tenant environment were elaborated in order to provide useful insights on the operation of systems in 3.5 GHz. By satisfying the required quality of certain user categories it is possible to increase the overall performance of the system by better allocation of resources to other user categories as well. In future work, in order to enhance the performance of the system, it is planned to investigate how machine learning principles can be utilized.

Furthermore, the topic of network slicing was investigated for evaluating the co-existence of URLLC and eMBB traffic and the impact of traffic increase (either eMBB or URLLC, and both) to the overall network performance. In this respect, an algorithm was proposed for creating and deciding on the dynamic resource allocation of network slices which reconfigures and adjusts the slices so as to provide appropriate QoS levels towards mobile client nodes. Also RRM issues in a multi-connection environment were elaborated through a simulation study for simulating densely connected and highly demanding next-generation networks. Finally, the benefits from using narrowband technologies, especially for small and frequent transmissions for high connection density were presented. As a result, the provided evaluations take into account these assumptions in order to show the number of devices that can be supported with a specific QoS. It was shown that up to 1 million mMTC devices for small packet transmissions (mMTC traffic) can be supported with low bandwidths around 1MHz. The next steps would deal with optimization issues with multi-connectivity, such as link robustness in high mobility scenarios or high reliability cases, under the scope of URLLC communications.

8.1 Future research directions

This research may be expanded in many directions. For instance, applying artificial intelligence (AI) to both the 5G network and the device will lead to more efficient wireless communications, longer battery life, and enhanced user experiences. AI is a powerful tool, and the key to harnessing AI to improve wireless is to focus on important wireless challenges that are both difficult to solve with traditional methods and are also a good fit for machine learning. The low latency and high capacity of 5G will also allow AI processing to be distributed among the device and cloud thus enabling flexible system solutions for a variety of new and enhanced experiences. This wireless edge architecture is adaptable and allows appropriate tradeoffs to be made per use case [1].

8.2 References

[1] Qualcomm, "5G+AI: The ingredients fueling tomorrow's tech innovations", 2020.

9 List of publications

The work described in Chapters 2-7 represents a selection of all the work carried out during the course of this PhD study. Below is listed the complete list of publications to journals, conferences and books resulted from this PhD research work.

9.1 Journal publications

I. Uwe Herzog, Andreas Georgakopoulos, Ioannis-Prodromos Belikaidis, Michael Fitch, Keith Briggs, Salvador Diaz, Óscar Carrasco, Klaus Moessner, Benoit Miscopein, Shahid Mumtaz, Panagiotis Demestichas, "Quality of Service Provision and Capacity Expansion through Extended-DSA for 5G", Transactions on Emerging Telecommunications Technologies, Vol. 27, No. 9, Sept. 2016, pp. 1250–1261.

II. Frank Schaich, Berna Sayrac, Salah-Eddine El-ayoubi, Ioannis-Prodromos Belikaidis, Marco Caretti, Andreas Georgakopoulos, Xitao Gong, Evangelos Kosmatos, Hao Lin, Panagiotis Demestichas, Belkacem Mouhouche, Klaus Pedersen, Nuno Pratas, Malte Schellmann, Martin Schubert, Musbah Shaat, Gerhard Wunder, "FANTASTIC-5G: flexible air interface for scalable service delivery within wireless communication networks of the 5th generation", Transactions on Emerging Telecommunications Technologies, Vol. 27, No. 9, Sept. 2016, pp. 1216–1224.

III. Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Panagiotis Demestichas, Benoit Miscopein, Marcin Filo, Seiamak Vahid, Bismark Okyere, Michael Fitch, "Multi-RAT Dynamic Spectrum Access for 5G Heterogeneous Networks: The SPEED-5G Approach", IEEE Wireless Communications, vol. 24, no. 5, pp. 14-22

IV. Ioannis-Prodromos Belikaidis, Stavroula Vassaki, Andreas Georgakopoulos, Aristotelis Margaris, Federico Miatton, Uwe Herzog, Kostas Tsagkaris, Panagiotis Demestichas, "Context-aware Radio Resource Management Below 6 GHz for Enabling Dynamic Channel Assignment in the 5G era", EURASIP Journal on Wireless Communications and Networking, December 2017, 2017:162.

V. Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos, Valerio Frascolla, Panagiotis Demestichas, "Management of 3.5-GHz Spectrum in 5G Dense Networks: A Hierarchical Radio Resource Management Scheme", IEEE Vehicular Technology Magazine, vol 13, Issue 2, June 2018, pp 57-64

VI. Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Kostas Tsagkaris, Zwi Altman, Sana Ben Jemaa, Panagiotis Demestichas, Nicolaos Mitrou, "5G Radio Access Network Slicing: System-Level Evaluation and Management “, IEEE Vehicular Technology Magazine, vol 14, Issue 4, December 2019, pp. 49-55

9.2 Conference publications

VII. Andreas Georgakopoulos, Ioannis-Prodromos Belikaidis, Kostas Tsagkaris, Vera Stavroulaki, Panagiotis Demestichas, "Wireless Access Infrastructure Expansions Through Opportunistic Networks of Moving Access Points", in Proc. 2016 European Conference on Networks and Communications (EuCNC), Athens, 27-30 June, 2016, pp. 163-167.

VIII. Uwe Herzog, Andreas Georgakopoulos, Ioannis-Prodromos Belikaidis, Panagiotis Demestichas, Salvador Diaz, Oscar Carrasco, Federico Miatton, Klaus Moessner, Valerio Frascolla, “Quality of Service Provision and Capacity Expansion Through Extended-DSA for 5G”, in Proc. 2016 European Conference on Networks and Communications (EuCNC), Athens, 27-30 June, 2016, pp. 200-204.

IX. Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Panagiotis Demestichas, Uwe Herzog, Oscar Carrasco, Valerio Frascolla, Michael Fitch, Benoit Miscopein, Klaus Moessner, Harald Weigold, “Flexible RRM/MAC solutions in a dense small cell environment: The SPEED-5G case”, in Proc. 2016 European Conference on Networks and Communications (EuCNC), Athens, 27-30 June, 2016.

X. Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Panagiotis Demestichas, Uwe Herzog, Klaus Moessner, Seiamak Vahid, Michael Fitch, Keith Briggs, Benoit Miscopein, "Trends and Challenges for Autonomic RRM and MAC Functionality for QoS Provision and Capacity Expansions in the Context of 5G Beyond 6GHz", in Proc. 2017 European Conference on Networks and Communications (EuCNC), Oulu, 12-15 June, 2017, pp. 1-5.

XI. Jose Alcaraz-Calero, Ioannis-Prodromos Belikaidis, Carlos Jesus Bernardos Cano, Pascal Bisson, Didier Bourse, Michael Bredel, Daniel Camps-Mur, Tao Chen, Xavier Costa-Perez, Panagiotis Demestichas, Mark Doll, Salah Eddine El-Ayoubi, Andreas Georgakopoulos, Aarne Mämmelä, Hans-Peter Mayer, Miquel Payaro, Bessem Sayadi, Muhammad Shuaib Siddiqui, Miurel Tercero, Qi Wang, “Leading Innovations Towards 5G: Europe’s Perspective in 5G Infrastructure Public-Private Partnership (5G-PPP)”, IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2017), Montreal, QC, Canada, 2017, pp. 1-5.

XII. Andreas Georgakopoulos, Evangelos Kosmatos, Ioannis-Prodromos Belikaidis, Martin Kurras, Lars Thiele, Panagiotis Demestichas, “Enabling Advanced 5G Component Validations and Optimizations by Means of System Level Simulations Platform, Abstractions, Models, Results and Further Challenges”, in Proc. 2018 European Conference on Networks and Communications (EuCNC), Ljubljana, 18-21 June 2018

XIII. Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos, MIsabel de la Bandera-Cascales, David Palacios, Raquel Barco, Panagiotis Demestichas, “5G Component Carrier Management Evaluation by Means of System Level Simulations”, in Proc. 2019 European Conference on Networks and Communications (EuCNC), Valencia, 18-21 June, 2019

9.3 Book publications

XIV. Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Evangelos Kosmatos, Stavroula Vassaki, Orestis-Andreas Liakopoulos, Vassilis Foteinos, Panagiotis Vlacheas, Panagiotis Demestichas, “Emerging technologies in software, hardware and management aspects towards the 5G era: Trends and Challenges”, 5G Networks: Fundamental Requirements, Enabling Technologies, and Operations Management, Oct. 2018

XV. Michal Maternia, Jose F. Monserrat, David Martín-Sacristán, Yong Wu, Changqing Yang, Mauro Boldi, Yu Bao, Frederic Pujol, Giuseppe Piro, Gennaro Boggia, Alessandro Grassi, Hans-Otto Scheck, Ioannis-Prodromos Belikaidis, Andreas Georgakopoulos, Katerina Demesticha, Panagiotis Demestichas, “Performance, Energy Efficiency and Techno-Economic Assessment”, 5G System Design: Architectural and Functional Considerations and Long Term Research, Apr. 2018

9.4 Contributions

XVI. International Telecommunication Union, ITU-R, "Final evaluation Report from the 5G infrastructure association on IMT-2020 proposal IMT-2020/19 (TSDSI)", IMT 2020 and beyond February 2020

10 List of Acronyms

- **3GPP** 3rd Generation Partnership Project
- **5G EIR** 5G Equipment Indentity Register
- **5G** Fifth Generation Mobile Networks
- **5QI** 5G QoS Indentifier
- **API** Application Program Interface
- **APs** Access Points
- **ARP** Allocation retention priority
- **BF** Beamforming
- **BS** Base station
- **CA** Carrier Aggregation
- **CAPEX** Capital Expenses
- **CC** Compoment Carrier
- **CCI** Co-Channel Interference
- **CCM** Component Carrier Management
- **CDF** Cumulative Distribution Function
- **CDMA** Code Division Multiple Access
- **cMTC** Critical Machine Type Communications
- **CN** Core Network
- **CoMP** Coordinated Multi-Point Transmission and Reception
- **CP** Control Plane
- **CRI** CSI-RS Resource Indicator
- **CS-DCA** Channel segregation based DCA
- **CSI** Channel State Information
- **CU** Control/ User Plane OR Central Unit
- **D2D** Device-to-Device
- **DANs** Distributed Antenna Networks
- **DB** Database/ Dual Band
- **DC** Dual Connectivity
- **DCA** Dinamic Channel Asignment
- **DC-HSPA** Dual Carrier-High Speed Packet Access
- **DCI** Downlink Control Indicator
- **DL** Downlink
- **DN** Data Network
- **EDGE** Enhanced Data Rate for Global Revolution
- **eMBB** Enhanced Mobile Broadband
- **ETSI** European Telecommunications Standards Institute
- **FCC** Federal Communications Commission
- **FoF** Factory of the Future
- **FTP** File Transfer Protocol
- **GAA** General Authorized Access
- **GBR** Guaranteed Bit Rate
- **GSM** Global system for mobile
- **HARQ** Hybrid Automatic Repeat Request

- **HO** Handover
- **HSPA** High Speed Packet Access
- **IEEE** Institute of Electrical and Electronics Engineers
- **Inter-RAT** Inter- Radio Access Technology
- **IoT** Internet of Things
- **IP** Internet Protocol
- **IQ-CS-DCA** Interference and QoS aware channel segregation based DCA
- **ISD** Inter-Site Distance
- **ITU** International Telecommunications Union
- **ITU-T** ITU-Telecommunication Standardization Bureau
- **KPIs** Key Performance Indicators
- **LAA** Licensed Assisted Access
- **LDPC** Low-density parity-check
- **LIR** Location Information Request
- **LMF** Location Management Function
- **LSA** Licensed Shared Access
- **LTE** Long-Term Evolution
- **M2M** Machine-to-Machine
- **MA** Multiple Access
- **MAC** Medium Access Control
- **MANO** Management Nodes
- **MAPS** Mobile access points
- **MCC** Mission Critical Communications
- **MEC** Mobile Edge Computing
- **MIMO** Multiple-Input Multiple-Output
- **MIOT** Massive Internet of Things
- **MLB** Mobility Load Balancing
- **mMTC** Massive Machine Type Communications
- **mmWave** Millimeter Wave
- **MN** Master Node
- **MNO** Mobile Network Operators
- **MTC** Machine Type Communications
- **MTC-M2M** Machine Type Communications/Machineto-Machine
- **MU-MIMO** Multi-User Multiple-Input Multiple-Output
- **N3IWF** Non-3GPP InterWorking Function
- **NB-IoT** Narrowband IoT
- **NFV** Network Function Virtualization
- **NG** Next Generation
- **NMs** Neural Networks
- **NOMA** Non-Orthofonal Multiple Access
- **NR** New Radio
- **NSD** Network Service Descriptors
- **NSI** Network Slice Instance
- **NWDAF** Network Data Analytics Function

- **OFDM** Orthogonal Frequency Division Multiplexing
- **OPEX** Operating Expenses
- **OTA** Over-The-Air
- **PAL** Priority Access Lincenses
- **PDF** Probability Density Function
- **PDU** Protocol Data Unit
- **PSCells** Primary Secondary Cells
- **QCI** QoS class Identifier
- **QoE** Quality of Experience
- **QoS** Quality of Service
- **RACH** Random Access Channel
- **RAN** Radio Access Network
- **RAT** Radio Access Technology
- **RRC** Radio Resource Control
- **RRH/RRU/ RU** Remote Radio Head/Remote Radio Unit/ Radio Unit
- **RRM** Radio Resource Management
- **RSRP** Reference Signal Received Power
- **RSRQ** Reference Signal Received Quality
- **SAS** Spectrume Access System
- **SBSs** Small cell base stations
- **SC** Service Continuity
- **SCells** Secondary Cells
- **SDN** Software-Defined Networking
- **SINR** Signal to Interference and noise ratio
- **SLA** Service License Agreement
- **SNR** Signal-to-noise ratio
- **SON** Self-Optimizing or Self-Organizing Network
- **SRB** Signaling Resource Bearer
- **SSID** Service Set Identification
- **SVM** Support Vector Machine
- **TC** Traffic Class
- **TLS** Transport Layer Security
- **ToS** Type of Service
- **TPC** Transmit Power Control
- **TRP** Transmission Reception Point
- **TTI** Transmit Time Travel
- **UE** User Equipement
- **URLLC** Ultra-Reliable Low Latency Communications
- **V2X** Vehicle-to-Everything
- **VNF** Virtualization of Network Functions
- **WLAN** Wireless Local Area Network