



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

**Εφαρμογή & Αξιολόγηση Επίδοσης Συναρτήσεων
Ωφέλειας Για Την Αποτελεσματική Ανάθεση
Πόρων Σε Ασύρματα Δίκτυα Πολυμέσων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΚΩΝΣΤΑΝΤΙΝΟΥ ΚΑΣΤΡΙΝΟΓΙΑΝΝΗ

Επιβλέπων : Συμεών Παπαβασιλείου

Επίκουρος Καθηγητής

Αθήνα, Ιούλιος 2007



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

**Εφαρμογή & Αξιολόγηση Επίδοσης Συναρτήσεων
Ωφέλειας Για Την Αποτελεσματική Ανάθεση
Πόρων Σε Ασύρματα Δίκτυα Πολυμέσων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΚΩΝΣΤΑΝΤΙΝΟΥ ΚΑΣΤΡΙΝΟΓΙΑΝΝΗ

Επιβλέπων : Συμεών Παπαβασιλείου

Επίκουρος Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 10^η Ιουλίου 2007.

.....

Συμεών Παπαβασιλείου

Επίκουρος Καθηγητής Ε.Μ.Π.

.....

Βασίλειος Μάγκλαρης

Καθηγητής Ε.Μ.Π.

.....

Τίμος Σελλής

Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2007

.....
ΚΑΣΤΡΙΝΟΓΙΑΝΝΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών
Ε.Μ.Π.

© 2007 – All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Στον Δημήτρη Σούλιο,
για την δυνατότητα του να πλαισιώνει με μαθηματικά
οποιαδήποτε ιδέα
Κωνσταντίνος Καστρινογιάννης

Ευχαριστήριο σημείωμα

Τον καθηγητή μου Συμεών Παπαβασιλείου, για την δυνατότητα που μου έδωσε να ασχοληθώ με ένα τόσο σημαντικό και ενδιαφέρον θέμα. Οι γνώσεις του και οι πολύτιμες συμβουλές του ήταν πάντα διαθέσιμες. Η εμπιστοσύνη του και η διαρκής ενθάρρυνσή μου, αποτέλεσαν κινητήριο δύναμη για εμένα.

Τον αδερφό και φίλο μου Τιμόθεο Καστρινογιάννη, υποψήφιο διδάκτορα Ε.Μ.Π., για τις γνώσεις και τις ιδέες που μου μετέφερε πάνω στις οποίες στηρίχθηκε αυτή η εργασία και για την συμβολή του στην μεταφορά μου από τον χώρο της μάθησης στον χώρο της έρευνας. Μέσα από αυτή την συνεργασία συνειδητοποίησα νέες πτυχές του αδερφού μου, την μεθοδικότητά του, την αστείρευτη έμπνευσή του και την δυναμική του σαν ερευνητής.

Ήταν τιμή μου η συνεργασία μαζί τους.

Περίληψη

Ο στόχος αυτής της διπλωματικής εργασίας είναι η μελέτη του προβλήματος ανάθεσης πόρων για το πρόβλημα του καναλιού καθόδου σε ασύρματα δίκτυα με πολλαπλή πρόσβαση διαίρεσης κώδικα που υποστηρίζουν υπηρεσίες πολυμέσων. Εισάγεται και αναλύεται ένας αλγόριθμος που στηρίζεται στην δυναμική εκχώρηση πόρων ισχύος και ρυθμού μετάδοσης με βάση την έννοια της συνάρτησης χρησιμοποίησης με σκοπό την βελτιστοποίηση της απόδοσης του συστήματος με παράλληλη ικανοποίηση των διάφορων απαιτήσεων ποιότητας υπηρεσίας για διάφορες κλάσεις υπηρεσιών. Για την αποδοτική υποστήριξη χρηστών πραγματικού χρόνου, η προτεινόμενη μεθοδολογία μας λαμβάνει υπόψη τους βραχυπρόθεσμους περιορισμούς ποιότητας των υπηρεσιών τους, που επιτυγχάνεται από την δυναμική μετατροπή των συναρτήσεων χρησιμοποίησης τους με βάση την πληροφορία της ικανοποίησης των βραχυπρόθεσμων περιορισμών τους. Προς αυτήν την κατεύθυνση, ορίζεται ένα μαθηματικό υπόβαθρο που υποστηρίζει και δικαιολογεί τις προτεινόμενες διαδικασίες αλλαγής των συναρτήσεων χρησιμοποίησης των χρηστών πραγματικού χρόνου και επίσης προτείνει αλγορίθμους χαμηλής πολυπλοκότητας για την υλοποίηση των αλλαγών αυτών. Τέλος μέσα από μοντελοποίηση και προσομοίωση δείχνεται ότι επιτυγχάνεται σημαντική βελτίωση της ικανοποίησης των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης μέχρι και για μεγάλες τιμές του πλήθους χρηστών πραγματικού χρόνου στο σύστημα χωρίς σημαντική απώλεια σε σχέση με τη βέλτιστη συνολική ρυθμαπόδοση του συστήματος

Λέξεις Κλειδιά: << Βραχυπρόθεσμοι περιορισμοί, δρομολόγηση, ασύρματα δίκτυα , υπηρεσίες πολυμέσων, ανάθεση πόρων. >>

Abstract

The purpose of this thesis is the analysis of the problem of resource allocation for the downlink in code division multiple access wireless networks supporting multimedia services is addressed. A utility based dynamic-pricing power and rate allocation algorithm which aims at optimizing system's performance while satisfying multiple services' diverse Quality of Service (QoS) requirements is introduced and analyzed. For the efficient support of real-time users our proposed methodology takes into account their short-term throughput QoS constraints satisfaction, which is achieved by dynamically adapting their utility functions with respect to their corresponding short-term service performance information. Towards that goal, a mathematical framework is provided that supports and justifies the proposed self-tuning procedures of real-time users' utility functions and also, leads to their implementation through low complexity algorithms. Via modelling and simulation it is demonstrated that significant performance improvements are achieved in terms of the short-term throughput requirement satisfaction even for large number of real-time users, without any considerable loss in the total system's throughput.

Keywords: << Short-term requirements, scheduling, wireless networks, multimedia services, resource allocation. >>



Περιεχόμενα

1	Εισαγωγή	13
1.1	Βασικές αρχές του HDR (High Data Rate) μοντέλου και μελλοντικές επεκτάσεις του	14
1.2	Στόχοι των αλγορίθμων χρονοδρομολόγησης.....	17
1.3	Ταξινόμηση κριτηρίων ποιότητας υπηρεσίας χρηστών.....	18
1.3.1	Εγγυήσεις ελάχιστης επίδοσης μεμονωμένα ανά χρήστη.....	18
1.3.2	Υπηρεσίες πραγματικού και μη πραγματικού χρόνου.....	19
1.3.3	Κριτήρια δικαιοσύνης μεταξύ χρηστών.....	26
1.4	Μελέτη απόδοσης αλγορίθμων υψηλής ρυθμαπόδοσης ως προς πιθανοτικούς βραχυπρόθεσμους περιορισμούς καθυστέρησης και ρυθμαπόδοσης.....	36
2	Αλγόριθμος UPR	44
2.1	Αλγόριθμος UPR για ανάθεση ισχύος	44
2.1.1	Μοντέλο περιγραφής.....	45
2.1.2	Συνάρτηση χρησιμοποίησης.....	47
2.1.3	Βασικό Πρόβλημα και απόδειξη βελτιστότητας	48
2.2	Αλγόριθμος επιλογής κινητών σταθμών.....	53
2.2.1	Αλγόριθμος Επιλογής κινητών σταθμών.....	59
2.3	Αλγόριθμος ανάθεσης ισχύος	61
2.4	Ασυμπτωτική βελτιστότητα και κατώτερο όριο επίδοσης του αλγορίθμου	62
2.5	Επέκταση του αλγορίθμου για συνδυασμένη ανάθεση ισχύος και ρυθμού μετάδοσης.....	67
2.5.1	Αναγωγή στο πρόβλημα απόδοσης ισχύος	68

2.5.2	Μελέτη της φυσικής σημασίας του αλγορίθμου για το πρόβλημα ρυθμαπόδοσης.....	71
2.6	Μελέτη πολιτικής	73
3	Βασικοί στόχοι	78
3.1	Βραχυπρόθεσμοι περιορισμοί ρυθμαπόδοσης (Sort-term throughput constraints).....	80
3.2	Οι προδιαγραφές των νέων συναρτήσεων χρησιμοποίησης	84
3.3	Η πληροφορία που θα εκμεταλλευτούμε.....	86
3.4	Προτεινόμενες τροποποιήσεις στις συναρτήσεις χρησιμοποίησης.....	88
3.5	Ανάλυση των τροποποιήσεων της παραμέτρου b	91
3.5.1	Πεδίο τιμών b	91
3.5.2	Συναρτήσεις προσδιορισμού της τιμής της παραμέτρου $b(t)$	92
3.6	Προτεινόμενος αλγόριθμος χρονοδρομολόγησης – M_UPR	96
4	Αριθμητικά αποτελέσματα	99
4.1	Εισαγωγικές πληροφορίες.	99
4.2	Πρώτο σενάριο (Χρήστες ιδίου μέσου καναλιού – CBR 512 Kbps)	101
4.3	Δεύτερο σενάριο (Χρήστες ιδίου μέσου καναλιού – CBR 728 Kbps).....	109
4.4	Τρίτο σενάριο (Χρήστες διαφορετικών μέσων καναλιών).....	115
5	Επίλογος	121
5.1	Σύνοψη και συμπεράσματα.....	121
5.2	Μελλοντικές επεκτάσεις	123
6	Βιβλιογραφία	125

1 Εισαγωγή

Οι ψηφιακές ασύρματες επικοινωνίες δεν είναι νέα ιδέα. Ήδη από το 1901 ο Ιταλός φιλόσοφος Γουλιέλμος Μαρκόνι επέδειξε στο κοινό έναν ασύρματο τηλεγράφο ανάμεσα στα πλοία και την ξηρά χρησιμοποιώντας κώδικα μορς (οι τελείες και οι παύλες είναι άλλωστε δυαδικό σύστημα). τα σύγχρονα συστήματα έχουν βέβαια καλύτερη απόδοση αλλά η βασική ιδέα παραμένει ίδια.

Το 1970 το πανεπιστήμιο της Χαβάης, υπό την καθοδήγηση του Norman Abramson, ανέπτυξε το πρώτο δίκτυο επικοινωνιών στον κόσμο, ονόματι ALOHAnet. Η τοπολογία αστέρα που χρησιμοποιήθηκε και περιλάμβανε επτά υπολογιστές, εγκαταστάθηκε σε 4 νησιά έτσι ώστε αυτά να επικοινωνούν με έναν κεντρικό υπολογιστή στο νησί Oahu, χωρίς να χρειάζονται γραμμές τηλεφώνου.

Το 1979, οι F.R.Greller και U.Bapst δημοσίευσαν ένα άρθρο στα IEEE Proceedings που περιέγραφε ένα πειραματικό ασύρματο τοπικό δίκτυο που χρησιμοποιούσε υπέρυθρες ακτινοβολίες. Αμέσως μετά, το 1980, ο P.Ferrert περιέγραψε ένα σύστημα ασύρματων επικοινωνιών βασισμένο σε ασύρματο δίαυλο μοναδικού κώδικα απλωμένου φάσματος και η εργασία του δημοσιεύτηκε στο συνέδριο IEEE National Telecommunications Conference. Το 1985 ο M. Kavehrad παρουσίασε ένα πειραματικό ασύρματο δίκτυο βασισμένο σε πολλαπλή πρόσβαση διαίρεσης κώδικα. Όλες αυτές οι εργασίες πυροδότησαν πολλές βιομηχανικές δραστηριότητες πάνω στην ανάπτυξη μιας σειράς γενεών τοπικών ασύρματων δικτύων. Η πρώτη γενιά ασύρματων δικτύων αναπτύχθηκε το 1980 από ερασιτεχνικές ομάδες τηλεπικοινωνιών. Αυτά τα ασύρματα δίκτυα πετύχαιναν ρυθμούς μετάδοσης μικρότερους από 9600bits/sec. Η δεύτερη και τρίτη γενιά ασύρματων δικτύων πέτυχαν ρυθμούς μετάδοσης της τάξης εκατοντάδων Kbit/sec και πάνω από 1Mbit/sec αντίστοιχα.

Σήμερα, τα ασύρματα δίκτυα έχουν μπει στην καθημερινότητα των ανθρώπων των σύγχρονων κοινωνιών. Η μεγάλη ανάπτυξη που παρουσιάζουν οδηγεί στην

σταδιακή αντικατάσταση των ενσύρματων επικοινωνιών. Αυτό αποτελεί πρόκληση για του πολλούς ανά τον κόσμο επιστήμονες ασύρματων δικτύων.

Στην συνέχεια της εισαγωγής θα περιγραφούν οι βασικές αρχές του μοντέλου HDR, θα παρουσιαστούν τα βασικές ανάγκες των υπηρεσιών που χρησιμοποιούνται όπως και οι αλγόριθμοι χρονοδρομολόγησης που προσπαθούν να τις καλύψουν.

1.1 Βασικές αρχές του HDR (High Data Rate) μοντέλου και μελλοντικές επεκτάσεις του

Κατά την εξέταση των ιδιοτήτων και πολιτικών που αναπτύσσονται στα πλαίσια του HDR ([1],[2]) πρέπει να ληφθούν υπόψη δύο θεμελιώδη χαρακτηριστικά των δεδομένων υψηλής ταχύτητας (High-speed data), σε αντίθεση παραδείγματος χάριν με τις απαιτήσεις μετάδοσης φωνής. Αυτά είναι :

- Η εν γένει ασύμμετρη και εκρηκτική κίνηση των δεδομένων
- Η ανοχή των δεδομένων στην καθυστέρηση

Τα χαρακτηριστικά αυτά μας δίνουν την ευχέρεια να εφαρμόσουμε σχήματα χρονοδρομολόγησης τέτοια ώστε να διαταράξουμε την ακολουθία εξυπηρέτησης των χρηστών προς όφελος του συστήματος. Από τεχνικής σκοπιάς, το HDR υλοποιείται με πολλαπλή πρόσβαση διαίρεσης κώδικα κάτω από πρόσβαση διαίρεσης χρόνου (CDMA/TDMA). Τα δεδομένα πολυπλέκονται χρονικά(TDM) σε χρονισχισμές(time slots) μήκους 1.67msec. Σε κάθε χρονοσχισμή θεωρούμε ότι το κανάλι κάθε χρήση παραμένει αμετάβλητο και κατά συνέπεια, η ποιότητα της υπηρεσίας που του παρέχεται θεωρείται σταθερή.

Ως επί το πλείστον, εξετάζεται η περίπτωση μίας κυψέλης (ή ενός τομέα) η οποία περιέχει ένα σταθμό βάσης (ή σημείο πρόσβασης, Access Point) και

πολλαπλά τερματικά (Access Terminals). Αντίθετα είναι επιτακτική η ανάγκη να μελετηθούν οι περιορισμοί που τίθενται στο σύστημα λόγω της συνύπαρξης πολλών σταθμών βάσεις. Οι περιορισμοί αυτοί έχουν να κάνουν με επικάλυψη συχνοτήτων αλλά και με τον διακυψελικός θόρυβο. Τέλος, κάθε σταθμός βάσης αλλά και κάθε χρήστης έχει περιορισμένη ισχύ μετάδοσης και αυτοί οι περιορισμοί πρέπει να ληφθούν υπόψη κατά την χρονοδρομολόγηση.

Σε κάθε χρονοσχισμή, ο λόγος σήματος προς παρεμβολής και θορύβου (SINR) καταγράφεται από τα τερματικά και με βάση αυτόν υπολογίζεται αναλογικά ο αντίστοιχος αξιόπιστος πραγματοποιήσιμος ρυθμός λήψης/εκπομπής δεδομένων κάθε χρήστη, ο οποίος αποστέλλεται στο σημείο πρόσβασης. Η μετάδοση αυτή της πληροφορίας πραγματοποιείται μέσα από κανάλι σηματοδοσίας, το οποίο χρησιμοποιεί φάσμα διαφορετικό από αυτό της μετάδοσης δεδομένων. Στον εκάστοτε σταθμό βάσης, βάσει των προαναφερθέντων αιτήσεων ελέγχου ρυθμού δεδομένων (DRCs) και κάποιας πολιτικής εξυπηρέτησης, ο χρονοδρομολογητής αποφασίζει τη χορήγηση της επόμενης χρονοσχισμής σε κάποιον/κάποιους εκ των υποψηφίων χρηστών. Εφόσον το πρόβλημα της χρονοδρομολόγησης μορφοποιείται ως μεγιστοποίηση ρυθμαπόδοσης και όχι με βάση τυχούσα συνάρτηση χρησιμότητας [2], τότε στην περίπτωση της ζεύξης καθόδου(downlink) το βέλτιστο είναι κάθε χρονοσχισμή να χορηγείται αποκλειστικά σε έναν χρήστη (σε περίπτωση ομογενών χρηστών), ενώ στην περίπτωση της ζεύξης ανόδου το βέλτιστο εν γένει είναι να γίνονται ταυτόχρονες εκπομπές από ένα σύνολο χρηστών.

Όσο όμως αυξάνονται οι δυνατότητες των χρηστών πάνω σε εφαρμογές, παράλληλα και με αναλογικό τρόπο αυξάνονται και οι απαιτήσεις τους από το σύστημα. Συγκεκριμένα, οι νέες απαιτήσεις από ένα σύστημα HDR είναι:

- Υψηλότεροι ρυθμοί μετάδοσης δεδομένων
- Ενοποίηση των υπηρεσιών
- Ευελιξία στην παροχή QoS (ποιότητα υπηρεσίας).

Είναι γεγονός ότι τα ασύρματα δίκτυα τείνουν σε πολλούς τομείς να αντικαταστήσουν τα ενσύρματα. Για να γίνει κάτι τέτοιο εφικτό, επιβάλλεται τα ασύρματα δίκτυα να παρέχουν ανάλογους ρυθμούς μετάδοσης με τα ενσύρματα. Η βασική διαφορά των συστημάτων αυτών έγκειται στην χωρητικότητά τους. Αναλυτικότερα, στα ενσύρματα δίκτυα η χωρητικότητα είναι απεριορίστη αν υποθέσουμε ότι έχουμε την δυνατότητα δίπλα από κάθε υπάρχον ομοαξονικό καλώδιο, ή δίπλα από κάθε υπάρχουσα οπτική ίνα να περάσουμε ένα ακόμα ομοαξονικό καλώδιο ή μια ακόμα οπτική ίνα. Αντιθέτως, στα ασύρματα δίκτυα που το μέσο μεταφοράς δεν είναι κάποιο καλώδιο αλλά ο αέρας, κάτι ανάλογο με τα παραπάνω δεν είναι εφικτό και κατά συνέπεια η μέγιστη χωρητικότητά τους είναι περιορισμένη (και σχετίζεται με τις δυνατότητες διαχωρισμού ενός σταθμού βάσης σε τομείς ή την εξολοκλήρου διάσπασή του σε πολλούς σταθμούς βάσης). Οι περιορισμοί της χωρητικότητας των ασύρματων δικτύων, όπως αυτοί περιγράφηκαν παραπάνω, κάνουν επιτακτική την ανάγκη δημιουργίας χρονοδρομολογητών αρκετά «έξυπνων» ώστε να μπορούν να διαχειριστούν με τον αποδοτικότερο τρόπο την όποια πληροφορία υπάρχει σχετικά με τις συνθήκες επικοινωνίας αλλά και τις ανάγκες για εξυπηρέτηση κάθε χρήστη.

Επιπλέον, πρέπει κάτω από τα ασύρματα δίκτυα του μέλλοντος, να ενοποιούνται όλα τα σχήματα μετάδοσης (transmission schemes) που χρησιμοποιούνται σήμερα και έχουν να κάνουν με την κωδικοποίηση της πληροφορίας που πρέπει να μεταδοθεί. Κάθε σχήμα μετάδοσης οδηγεί σε διαφορετικούς ρυθμούς σφάλματος και κατά συνέπεια πρέπει να αντιμετωπισθεί με τον κατάλληλο τρόπο από τον ίδιο χρονοδρομολογητή.

Τέλος οι χρονοδρομολογητές των επόμενων γενεών πρέπει να έχουν γνώση του είδους της υπηρεσίας που απαιτεί ένας χρήστης και τους περιορισμούς που συνοδεύουν κάθε υπηρεσία. Η πληροφορία αυτή πρέπει να χρησιμοποιηθεί από

τον χρονοδρομολογητή έτσι ώστε να μεγιστοποιηθεί ταυτόχρονα η «ευχαρίστηση» κάθε χρήστη και η χρησιμοποίηση του συστήματος.

Όπως γίνεται εμφανές από τα παραπάνω η δημιουργία ενός αποδοτικού και έξυπνου χρονοδρομολογητή κάτω από τον οποίο θα ενοποιούνται όλα τα είδη μετάδοσης αλλά και όλες οι υπηρεσίες που απαιτούν οι χρήστες αποτελεί μεγάλη πρόκληση για κάθε ερευνητή που ασχολείται με τα ασύρματα δίκτυα και της τηλεπικοινωνίες.

1.2 Στόχοι των αλγορίθμων χρονοδρομολόγησης

Μία από τις βασικότερες παρατηρήσεις που προηγείται της συζήτησης περί χρονοδρομολόγησης σε ασύρματα δίκτυα είναι ενδεχομένως ότι ο ασύρματος διάυλος διαφέρει θεμελιωδώς από τον ενσύρματο με αποτέλεσμα το εν λόγω πρόβλημα να πρέπει να εξεταστεί υπό διαφορετικό πρίσμα. Η κύρια διαφορά συνίσταται στο γεγονός ότι η επίδοση του τυχόντος χρήστη ενός ασυρμάτου συστήματος εξαρτάται άμεσα από την κατάσταση του διαύλου, η οποία εξαρτάται από την θέση του χρήστη και την κατάσταση του περιβάλλοντος γύρω του και κατά συνέπεια είναι χρονικά εξαρτώμενη.

Ο πρώτος στόχος των υπό συζήτηση αλγορίθμων δεν είναι άλλος από τη μεγιστοποίηση της ρυθμαπόδοσης (throughput) του συστήματος και την εξασφάλιση των απαιτήσεων των υπηρεσιών των χρηστών. Όπως συζητήθηκε προηγουμένως, η ανοχή των δεδομένων υψηλής ταχύτητας στην καθυστέρηση μας επιτρέπει να διαχειριστούμε με ελαστικό τρόπο τη σειρά εξυπηρέτησης των χρηστών. Αυτό, σε συνδυασμό με τη χρονικά μεταβαλλόμενη κατάσταση του διαύλου μας επιτρέπει να χρονοδρομολογούμε ευκαιριακά (opportunistic scheduling), δηλαδή να επιλέγουμε εν γένει χρήστες με ευνοϊκές συνθήκες διαύλου, με αποτέλεσμα να αυξάνεται η φασματική απόδοση του συστήματος.

Φυσικά, ενδεχόμενη εφαρμογή της ανωτέρω πολιτικής άνευ περιορισμών οδηγεί εν γένει σε αποκλίνοντα επίπεδα εξυπηρέτησης των χρηστών του συστήματος, ακόμα και σε αποκλεισμό χρηστών σε οριακές περιπτώσεις. Επομένως, όπως διαφαίνεται, ο δεύτερος στόχος των αλγορίθμων επιπλέον της μεγιστοποίησης της ρυθμαπόδοσης του συστήματος θα πρέπει να είναι ο έλεγχος της ποιότητας της παρεχόμενης υπηρεσίας(QoS).

1.3 Ταξινόμηση κριτηρίων ποιότητας υπηρεσίας χρηστών

Ο προαναφερθείς έλεγχος διεξάγεται με την επιβολή κάποιων περιορισμών ποιότητας υπηρεσίας (QoS constraints). Οι περισσότεροι από τους περιορισμούς που έχουν προταθεί κατά καιρούς, είναι δυνατό να ταξινομηθούν σε τρεις κύριες κατηγορίες:

- εγγυήσεις ελάχιστης επίδοσης μεμονωμένα ανά χρήστη
- ζητήματα εξυπηρέτησης υπηρεσιών πραγματικού ή μη πραγματικού χρόνου (Real time – non real time)
- ζητήματα δικαιοσύνης μεταξύ χρηστών(fairness)

1.3.1 Εγγυήσεις ελάχιστης επίδοσης μεμονωμένα ανά χρήστη

Μέσα από περιορισμούς σχετικά με την ελάχιστη επίδοση κάθε χρήστη από το σύστημα μπορούμε να εξασφαλίζουμε ένα κάτω φράγμα στην ποιότητα της υπηρεσίας που του παρέχεται. Η ελάχιστη επίδοση μπορεί να μεταφραστεί είτε σε χρόνο πρόσβασης στο δίκτυο είτε σε ποσότητα δεδομένων είτε σε επίπεδα

μέσης ρυθμαπόδοσης. Η κατηγορία αυτών των περιορισμών δεν είναι ορθογώνια με τις επόμενες δύο και κατά συνέπεια είναι προτιμότερο να μελετηθεί ταυτόχρονα με αυτές στα επόμενα κεφάλια.

1.3.2 Υπηρεσίες πραγματικού και μη πραγματικού χρόνου.

Παρακολουθώντας τον τρόπο με τον οποίο ένας χρήστης χρησιμοποιεί το διαδίκτυο μπορούμε πολύ εύκολα να εντοπίσουμε έντονες διαφοροποιήσεις ανάμεσα στις ανάγκες που έχει σε σχέση με την εκάστοτε υπηρεσία. Αυτές οι διαφοροποιήσεις έχουν να κάνουν με τις ανάγκες κάθε υπηρεσίας ως προς τον ρυθμό με τον οποίο μεταδίδεται η πληροφορία σε βραχυπρόθεσμο πλαίσιο. Με βάση τα παραπάνω μπορούμε να χωρίσουμε τις υπηρεσίες (και κατ' αναλογία και τους χρήστες τους) σε υπηρεσίες πραγματικού χρόνου και υπηρεσίες μη πραγματικού χρόνου.

1.3.2.1 Υπηρεσίες μη πραγματικού χρόνου

Σε αυτή την κατηγορία υπηρεσιών μπορούμε να εντάξουμε τις διαδομένες εφαρμογές , ftp και http, και γενικότερα τις εφαρμογές που εμπεριέχουν λήψη μεγάλων αρχείων (κειμένου, μουσικής, βίντεο, φωτογραφιών, προγραμμάτων και τα λοιπά).

Οι παραπάνω υπηρεσίες έχουν τα εξής χαρακτηριστικά:

- **Μακροπρόθεσμη ικανοποίηση των απαιτήσεων QoS.**

Οι χρήστες υπηρεσιών μη πραγματικού χρόνου είναι ανεκτικοί ως προς το χρονικό πλαίσιο ικανοποίησης των απαιτήσεών τους. Αυτές οι απαιτήσεις έχουν να κάνουν κυρίως με μεγάλους ρυθμούς μετάδοσης σε μακροπρόθεσμο επίπεδο. Σε ένα σύστημα HDR όπου κάθε χρονοσχισμή είναι της τάξης του χιλιοστού του δευτερολέπτου, μπορούμε να εκφράσουμε την έννοια του μακροπρόθεσμου μέσα από χρονικά διαστήματα της τάξης του λεπτού. Κατά συνέπεια ένας χρήστης υπηρεσιών μη πραγματικού χρόνου δεν ενδιαφέρεται τόσο για τους στιγμιαίους ρυθμούς μετάδοσης που επιτυγχάνει αλλά, όσο για την απόκτηση των δεδομένων που θέλει μέσα σε εύλογο (σε σχέση με το μέγεθός τους) χρονικά διαστήματα.

- **Απαιτείται ήπια ανταπόκριση του χρονοδρομολογητή κατά την παραβίαση των QoS. Απαίτηση που ταιριάζει απόλυτα με την λειτουργία ενός ευκαιριακού χρονοδρομολογητή**

Λαμβάνοντας υπόψη ότι οι αποφάσεις του χρονοδρομολογητή καθορίζουν την λειτουργία του συστήματος σε κάθε χρονοσχισμή, συμπεραίνουμε ότι σε περίπτωση που δεν ικανοποιούνται οι απαιτήσεις του χρήστη ή σε περίπτωση που προβλέπεται υποβάθμιση των υπηρεσιών ενός χρήστη, τότε ο χρονοδρομολογητής πρέπει να έχει ήπια ανταπόκριση. Μια ήπια ανταπόκριση θα αρκούσε, από την στιγμή που μακροπρόθεσμα θα οδηγούσε στην ικανοποίηση ή στην αναβάθμιση των υπηρεσιών πραγματικού χρόνου.

Με σεβασμό στα παραπάνω, ένας χρονοδρομολογητής που απευθύνεται σε χρήστες μη πραγματικού χρόνου θα πρέπει να κινείται με αργά βήματα ως προς την εκπλήρωση των κριτηρίων ποιότητας των υπηρεσιών των χρηστών δίνοντας προτεραιότητα στην μέγιστη χρησιμοποίηση του συστήματος που όπως θα αναλυθεί και στα επόμενα κεφάλαια εξασφαλίζεται από την ευκαιριακή χρονοδρομολόγηση.

- **Υψηλή ρυθμαπόδοση**

Τέλος, οι υπηρεσίες αυτές απαιτούν υψηλούς μέσους ρυθμούς μετάδοσης δεδομένων. Αυτή είναι μια απαίτηση που μπορεί να εξασφαλιστεί μόνο αν οι πόροι του συστήματος ανά πάσα στιγμή μοιράζονται με τον καλύτερο δυνατό τρόπο, δηλαδή σε χρήστες με καλά κανάλια. Όπως θα εξηγηθεί στην συνέχεια αυτός ο τρόπος μακροπρόθεσμα θα οδηγήσει στην υψηλή ρυθμαπόδοση όλων των χρηστών.

1.3.2.2 Υπηρεσίες πραγματικού χρόνου

Σε αντίθεση με τις υπηρεσίες του προηγούμενο κεφαλαίου, στα σύγχρονα δίκτυα έχουν δημιουργηθεί υπηρεσίες που απαιτούν σταθερή στιγμιαία ρυθμαπόδοση. Μια πολύ σημαντική επέκταση του διαδικτύου τα τελευταία χρόνια είναι η ανάπτυξη πρωτοκόλλων που υποστηρίζουν τηλεφωνία. Το διαδίκτυο πρέπει να είσαι σε θέση να παρέχει υπηρεσία τηλεφωνίας που είναι ανάλογη των επηρειών των χρηστών από τα ενσύρματα δίκτυα τηλεφωνίας που έχουν σχεδιαστεί αποκλειστικά για αυτό το λόγο. Ακόμα πιο απαιτητικές είναι οι υπηρεσίες που παρέχουν βίντεο κλήσεις και τηλεσυνεδριάσεις όπου σε πραγματικό χρόνο πραγματοποιείται μεταφορά ήχου και εικόνας από δύο ή περισσότερους χρήστες. Τέλος οι δυνατότητες του διαδικτύου έχει φέρει επανάσταση και στην βιομηχανία των παιχνιδιών. Τα περισσότερα νέα

παιχνίδια υπολογιστών υποστηρίζουν την συνύπαρξη πολλών παιχτών στον ίδιο χρόνο ή στην ίδια παρτίδα εκμεταλλευόμενα το διαδίκτυο. Για να είναι επιτυχής αυτή η συνύπαρξη των παιχτών και σε αυτή την περίπτωση είναι απαραίτητη η μετάδοση πληροφορίας με σταθερούς ρυθμούς.

Συγκεντρώνοντας τα χαρακτηριστικά των υπηρεσιών πραγματικού χρόνου έχουμε:

- **Ικανοποίηση των QoS σε βραχυπρόθεσμα χρονικά διαστήματα.**

Για να εξασφαλισθεί η συνέχεια στην μετάδοση της φωνής ή βίντεο το σύστημα πρέπει να μεταφέρει επαρκή ποσότητα πληροφορίας σε βραχυπρόθεσμο πλαίσιο παρατήρησης. Σε αντίθετη περίπτωση μια κλήση θα είχε διακοπές, μια μετάδοση βίντεο θα είχε χαμένα στιγμιότυπα (frames) ή ένα δικτυακό παιχνίδι θα έχανε τον συγχρονισμό του. Κατά συνέπεια ένας χρονοδρομολογητής πρέπει να έχει γνώση της υπηρεσίας που εκτελεί ο κάθε χρήστης και κατά περίπτωση να θυσιάζει την μέγιστη ρυθμαπόδοση συστήματος προς όφελος της ποιότητας της παρεχόμενης υπηρεσίας.

- **Απαιτείται άμεση ανταπόκριση του χρονοδρομολογητή κατά την παραβίαση των QoS.**

Κάθε φορά που παραβιάζονται οι απαιτήσεις ενός χρήστη υπηρεσίας πραγματικού χρόνου, ο χρονοδρομολογητής πρέπει να ανταποκρίνεται άμεσα με διορθωτικές κινήσεις έτσι ώστε να αποκατασταθεί η ποιότητα της εκάστοτε επικοινωνίας. Η εμπειρία έχει δείξει ότι ο χρήστης μιας υπηρεσίας πραγματικού χρόνου δεν είναι ανεκτικός ως προς την παραβίαση της ποιότητας της υπηρεσίας

του και προτιμά να διακόψει την υπηρεσία που χρησιμοποιεί αν δεν εξασφαλίζονται οι περιορισμοί που θέτει.

Κατά συνέπεια ένας έξυπνος χρονοδρομολογητής πρέπει όχι μόνο να προβλέπει αλλά και να διορθώνει άμεσα κάθε παραβίαση των περιορισμών μια τέτοιας υπηρεσίας. Έναν τέτοιο χρονοδρομολογητή θα προσπαθήσουμε να προδιαγράψουμε μέχρι το τέλος αυτής της εργασίας.

- **Άνω και κάτω φραγμένες απαιτήσεις ρυθμαπόδοσης.**

Μια υπηρεσία πραγματικού χρόνου έχει απαιτήσεις ρυθμαπόδοσης που είναι φραγμένες από πάνω και από κάτω. Αυτό θα γίνει πολύ πιο προφανές μέσα από το παρακάτω παράδειγμα. Αν θεωρήσουμε μια τηλεφωνική κλήση, τότε γνωρίζουμε ότι η μεταφορά της φωνής γίνεται με αποδεκτό τρόπο αν εξασφαλισθεί ρυθμαπόδοση 64kbps. Αν επιτευχθεί μικρότερος από αυτόν τον ρυθμό τότε οι δύο χρήστες δεν θα μπορούν να επικοινωνήσουν. Αντίθετα αν επιτευχθεί διπλάσια ρυθμαπόδοση, τότε δεν θα βελτιωθεί η ποιότητα της υπηρεσίας. Κατά συνέπεια σε ένα χρήστη πραγματικού χρόνου πρέπει αφενός να του δίνουμε πάντα στιγμιαία ρυθμαπόδοση μεγαλύτερη ή ίση της επιθυμητής, αφετέρου να περιορίζουμε τις δυνατότητες της για περαιτέρω αύξηση.

1.3.2.3 Μαθηματική μοντελοποίηση των κριτηρίων ποιότητας υπηρεσίας υπηρεσιών μη πραγματικού χρόνου

Ένας τρόπος για περιγράψουμε με μαθηματικό τρόπο της ανάγκες ενός χρήστη υπηρεσίας μη πραγματικού χρόνου, είναι θέσουμε ένα κατώφλι που να προσδιορίζει την ελάχιστη επιτρεπτή εξυπηρέτησή του από το σύστημα.

$$E[r(t)] \geq R_i, \forall i \in N \quad (1.1)$$

στον παραπάνω τύπο με N εκφράζεται το σύνολο των χρηστών υπηρεσιών μη πραγματικού χρόνου στο σύστημα.

Με $E[r(t)]$ περιγράφεται η μέση εξυπηρέτηση του χρήστη την χρονική στιγμή t , όπου σαν εξυπηρέτηση μπορούμε να θεωρήσουμε την πρόσβασή του στο σύστημα ή την ποσότητα των δεδομένων που έχει λάβει.

Αντίστοιχα με R_i περιγράφεται το ελάχιστο κατώφλι εξυπηρέτησης, εκφρασμένο σε πρόσβαση στο σύστημα ή σε ποσότητα δεδομένων.

Κατά συνέπεια, με βάση τον παραπάνω περιορισμό, κάθε χρήστης υπηρεσίας μη πραγματικού χρόνου στο σύστημα πρέπει να εξυπηρετείται κατά μέσο όρο περισσότερο από R_i .

Ο παραπάνω περιορισμός μπορεί να συνδυαστεί με περιορισμούς δικαιοσύνης όπως θα δούμε στην συνέχεια.

1.3.2.4 Μοντελοποίηση υπηρεσιών πραγματικού χρόνου

Για να μοντελοποιήσουμε υπηρεσίες που είναι ευαίσθητες στην καθυστέρηση, πρέπει να εισάγουμε περιορισμούς που να συσχετίζονται με την πρόσβαση ενός χρήστη στο σύστημα, σε βραχυπρόθεσμο επίπεδο.

Μέχρι στιγμής έχον προταθεί δύο βασικές μέθοδοι για την περιγραφή των περιορισμών που πρέπει να συνοδεύουν υπηρεσίες πραγματικού χρόνου.

- **Αυστηροί περιορισμοί καθυστέρησης**

Με βάση τους περιορισμούς αυτούς κάθε χρήστης i εξυπηρετείται σε $\phi_i M$ χρονοσχιμές σε κάθε μη επικαλυπτόμενο παράθυρο M διαδοχικών χρονοσχιμών. Στην περίπτωση αυτή με ϕ_i προσδιορίζουμε ένα προκαθορισμένο βάρος που προσδιορίζει την εκάστοτε υπηρεσία

- **Πιθανοτικοί βραχυπρόθεσμοι περιορισμοί καθυστέρησης**

Οι περιορισμοί αυτοί φράσσουν την πιθανότητα ένας χρήστης υπηρεσίας πραγματικού χρόνου να μην αποκτήσει πρόσβαση στο δίκτυο μέσα σε χρονικό διάστημα μεγαλύτερο από αυτό που εκφράζεται από τις απαιτήσεις της υπηρεσίας του. Όπως φαίνεται στον παρακάτω τύπο

$$\Pr[d_i \geq T_i] < g_{STD,i} \quad (1.2)$$

όπου το d_i αναπαριστά το σύνολο των διαδοχικών χρονοσχιμών που ο χρήστης i δεν έχει πρόσβαση στο σύστημα,

το T_i αναπαριστά το κατώφλι ανοχής στην καθυστέρηση του χρήστη i ,

το $g_{STD,i}$ αναπαριστά το πιθανοτικό βραχυπρόθεσμο κατώφλι ανοχής στην καθυστέρηση του χρήστη i ,

η ποιότητα μιας υπηρεσίας θεωρείται επαρκής μόνο στην περίπτωση που ο χρήστης της εξυπηρετείται σε διαστήματα που απέχουν μεταξύ τους λιγότερο από T_i με πιθανότητα μεγαλύτερη από $1 - g_{STD,i}$.

1.3.3 Κριτήρια δικαιοσύνης μεταξύ χρηστών

1.3.3.1 Αναλογική Δικαιοσύνη (Proportional Fairness, [2],[3])

Στόχος αυτού του αλγορίθμου είναι η μεγιστοποίηση της ρυθμαπόδοσης του συστήματος λαμβάνοντας υπόψιν τόσο την τρέχουσα κατάσταση διαύλου του χρήστη όσο και το ιστορικό της εξυπηρέτησής του. Η τρέχουσα κατάσταση του διαύλου αντιπροσωπεύεται όπως εξηγήσαμε νωρίτερα από την αίτηση ρυθμού δεδομένων του χρήστη (δηλαδή το DRC) ενώ το ιστορικό της εξυπηρέτησης αντιπροσωπεύεται από το μέσο όρο της ρυθμαπόδοσης του χρήστη μέχρι την παρούσα χρονοσχισμή. Η επιλογή βασίζεται στη μεγιστοποίηση του λόγου που μόλις περιγράψαμε και μοντελοποιείται μαθηματικά με τον εξής κανόνα:

$$\operatorname{argmax}_i \frac{\text{DRC}_i(t)}{R_i(t)}, 1 \leq i \leq N \quad (1.3)$$

όπου :

N : ο συνολικός αριθμός των χρηστών στο σύστημα

$\text{DRC}_i(t)$: η αίτηση ρυθμού δεδομένων του χρήστη (data rate request) i τη χρονική στιγμή t

$R_i(t)$: ο εκθετικά εξομαλυμένος μέσος όρος της ρυθμαπόδοσης του χρήστη i μέχρι τη χρονική στιγμή t , ο οποίος δίνεται από τον αναδρομικό τύπο

$$R_i(t+1) = \left(1 - \frac{1}{t_c}\right) \cdot R_i(t) + \frac{1}{t_c} \cdot \text{Current_Transmission_Rate_of_user_}i, \text{ όπου}$$

t_c : μία σταθερά η οποία μετράται σε χρονοσχιμές

$C_T_R_i$: ο τρέχων ρυθμός μετάδοσης του χρήστη i , ο οποίος θεωρείται μηδενικός για όλους τους μη επιλεγέντες χρήστες.

Σημειώσεις:

- i) Η σταθερά t_c στη διαδικασία ενημέρωσης του ιστορικού εξυπηρέτησης είναι κρίσιμη καθότι επηρεάζει το ρυθμό με τον οποίο ενσωματώνονται στο ιστορικό οι αλλαγές στην κατάσταση του διαύλου. Μικρές τιμές της t_c ναί μεν οδηγούν σε γρήγορη ανίχνευση των αλλαγών αλλά προκαλούν έντονη διακύμανση των παρατηρήσεων ενώ μεγάλες τιμές της t_c οδηγούν μεν σε ασφαλέστερη εκτίμηση του R_i αλλά μπορεί να προκαλέσουν τον αποκλεισμό χρηστών από την υπηρεσία για αρκετό χρόνο.
- ii) Το t_c έχει άλλον έναν εξαιρετικής σημασίας ρόλο. Καθότι εν γένει δεν είναι σταθερά αλλά αντιπροσωπεύει την πάροδο του χρόνου, ουσιαστικά είναι ο όρος που εξαναγκάζει τον αλγόριθμο να συγκλίνει στις λύσεις μίας συνήθους διαφορικής εξίσωσης με μοναδικό σημείο ισορροπίας. Γενικά, ο όρος που περιέχει το $\frac{1}{t_c}$ και αντιπροσωπεύει την επίδραση του παρόντος θα τείνει στο μηδέν καθώς $\lim_{t \rightarrow \infty} \frac{1}{t} = 0$, ενώ ο όρος που περιέχει το $\left(1 - \frac{1}{t_c}\right)$ και αντιπροσωπεύει την επίδραση του παρελθόντος θα τείνει στο ένα καθώς $\lim_{t \rightarrow \infty} \left(1 - \frac{1}{t}\right) = 1$.
- iii) Ο αλγόριθμος είναι βέλτιστος υπό την έννοια ότι αν κάποιος άλλος αλγόριθμος χρονοδρομολόγησης χρησιμοποιηθεί για να αυξήσει τη ρυθμαπόδοση ενός χρήστη κατά $x\%$ σε σχέση με αυτή που λαμβάνει

υπό το PFS(Proportional Fair Scheduling), τότε το άθροισμα των μειώσεων των ρυθμαποδόσεων όλων των άλλων χρηστών υπό το νέο αλγόριθμο θα υπερβαίνει το x% (proportional fairness criteria).

1.3.3.2 Μακροπρόθεσμη Δικαιοσύνη Χρόνου Πρόσβασης (Long Term Access Fairness)

Δεδομένου ότι πρακτικά αυτό που μοιράζονται οι χρήστες κατά την εξυπηρέτησή τους από το σύστημα είναι οι χρονοσχισμές, δηλαδή ο χρόνος πρόσβασης, φαντάζει λογικό ο επιβαλλόμενος περιορισμός ποιότητας υπηρεσίας να αφορά τον χρόνο πρόσβασης του κάθε χρήστη. Εδώ λοιπόν, στόχος είναι η μεγιστοποίηση της ρυθμαπόδοσης του συστήματος υπό τον περιορισμό κάποιου δεδομένου ελάχιστου χρόνου πρόσβασης του κάθε χρήστη σε μακροπρόθεσμο ορίζοντα. Σε αυτό το σημείο παραθέτουμε τους ακόλουθους δύο βασικούς αλγορίθμους που επιτελούν την προαναφερθείσα λειτουργία.

1.3.3.2.1 Temporal Fairness ([4],[5])

Σε αυτή την περίπτωση το πρόβλημα διατυπώνεται ποσοτικά ως εξής:

$$\begin{aligned} & \text{maximize}_{Q \in \Theta} E(U_{Q(\bar{U})}) \\ & \text{subject to } P\{Q(\bar{U}) = i\} \geq r_i, \quad i=1,2,\dots,N \end{aligned} \quad (1.4)$$

όπου :

Q : η πολιτική που ακολουθεί ο χρονοδρομολογητής

Θ : το σύνολο όλων των στατικών πολιτικών

E(): αναμενόμενη τιμή

\bar{U} : το διάνυσμα για όλους τους χρήστες της χρησιμότητας που αντλεί ο κάθε χρήστης i , $1 \leq i \leq N$ από την ανάθεση της χρονοσχισμής σε αυτόν

r_i : το ελάχιστο κλάσμα χρονοσχισμών (ως προς τις συνολικές) που πρέπει να ανατεθεί στον χρήστη i . Φυσικά, $r_i \geq 0$ και $\sum_{i=1}^N r_i \leq 1$.

Αποδεικνύεται ότι η βέλτιστη λύση στο πρόβλημα που μόλις διατυπώσαμε επιτυγχάνεται με την ακόλουθη πολιτική:

$$Q^*(\bar{U}) = \arg \max_i (U_i + v_i^*)$$

όπου τα v_i^* είναι πραγματικές παράμετροι τέτοιες ώστε :

- (a) $\min_i (v_i^*) = 0$
- (b) For all i , $P\{Q^*(\bar{U}) = i\} \geq r_i$
- (c) For all i , if $P\{Q^*(\bar{U}) = i\} > r_i$, then $v_i^* = 0$

Στην πράξη, οι ρυθμοί r_i είναι προαποφασισμένοι στο σύστημα και το μόνο που χρήζει υπολογισμού είναι οι παράμετροι v_i^* . Για τον προσδιορισμό τους, δεδομένης της ικανοποίησης μιας σειράς προϋποθέσεων όπως αναλύεται στο [5], χρησιμοποιούμε έναν αλγόριθμο στοχαστικής σύγκλισης της μορφής :

$$v_i^{k+1} = v_i^k - \delta^k \cdot g_i^k \quad (1.5),$$

όπου:

$$g_i^k = (v_i^k - \min_j (v_j^k)) \cdot \left(\mathbf{1}_{\{Q^k(\bar{U})=i\}} - r_i \right), \quad i = 1, 2, \dots, N-1$$

και

$$\text{π.χ. } \delta^k = 1/k$$

Σημείωση:

Τα u_i^* ουσιαστικά αποτελούν συνιστώσες οι οποίες προμοδοτούν τους χρήστες που βιώνουν «σχετικά δυσμενείς» συνθήκες διαύλου ($u_i^* > 0$) εις βάρος των χρηστών που βιώνουν λιγότερο δυσμενείς ή ευμενείς συνθήκες (π.χ. $u_i^* = 0$). Παρόλα αυτά, η εν λόγω προμοδότηση συνεισφέρει μέχρι το σημείο της ελάχιστης κάλυψης της απαίτησης του χρήστη.

1.3.3.2 Αλγόριθμος συσχετιστικά καλύτερου χρήστη (Relatively Best ([6]))

Εδώ, η βέλτιστη πολιτική που προτείνεται είναι η ακόλουθη:

$$i^*(\xi) = \arg \max_i \frac{\xi_i - \bar{\xi}_i}{c_i}, \quad 1 \leq i \leq N \quad \text{όπου :}$$

ξ_i : στιγμιαία κατάσταση του διαύλου του χρήστη i , δηλαδή στιγμιαία τιμή του λόγου σήματος προς παρεμβολής και θορύβου (SINR)

$\bar{\xi}_i$: η μέση τιμή των ξ_i , δηλαδή $\bar{\xi}_i = E[\xi_i(t)]$

c_i : θετικές παράμετροι ελέγχου του χρόνου πρόσβασης κάθε χρήστη

Σημείωση:

i) Όπως διαφαίνεται, η εφαρμογή της ανωτέρω πολιτικής προϋποθέτει πλήρη γνώση της μέσης τιμής των στιγμιαίων καταστάσεων των διαύλων των χρηστών, γεγονός που τη διαφοροποιεί ουσιαστικά από τον προηγούμενο αλγόριθμο (temporal fairness B.1)

ii) Οι παράμετροι c_i ουσιαστικά αποτελούν το ρυθμιστικό μηχανισμό για την ανάθεση εν γένει διαφορετικών κλασμάτων των χρονοσχισμών στους χρήστες. Επομένως, η απόφαση της χρονοδρομολόγησης λαμβάνεται με βάση μία κανονικοποιημένη τιμή της κατάστασης του καναλιού κάθε χρήστη, η οποία αναφέρεται ως η «συγκριτικά καλύτερη». Κατά συνέπεια, κάθε χρήστης εξυπηρετείται μόνο σε «ευνοϊκές» στιγμές αλλά με ένα συγκεκριμένο ασυμπτωτικό χρόνο πρόσβασης, ελεγχόμενο από τις παραμέτρους c_i .

iii) Αξιοσημείωτη είναι η περίπτωση κατά την οποία $c_i = \bar{\xi}_i$, οπότε το ποσοστό ϕ_i των χρονοσχισμών που ανατίθενται στον χρήστη i μεταπίπτει σε $\frac{1}{N}$ ενώ ταυτόχρονα η ασυμπτωτική δικαιοσύνη του Relatively Best(RB) αλγορίθμου εκφυλίζεται σε αυτή του Round Robin(RR).

1.3.3.3 Μακροπρόθεσμη Δικαιοσύνη Ρυθμαπόδοσης (Long-Term Throughput Fairness)

Η ανάγκη επιβολής περιορισμών ρυθμαπόδοσης πηγάζει από την εξής θεμελιώδη διαφορά μεταξύ ενσύρματου και ασυρμάτου διαύλου. Στον ενσύρματο δίαυλο, η χορήγηση ενός συγκεκριμένου ποσοστού του χρόνου πρόσβασης είναι ισοδύναμη με τη χορήγηση ενός συγκεκριμένου ποσοστού της ρυθμαπόδοσης του συστήματος. Αντιθέτως, στον ασύρματο δίαυλο, τα δύο αυτά μεγέθη είναι μεν συσχετισμένα αλλά όχι ευθέως. Επομένως, οι αλγόριθμοι που ακολουθούν είναι απαραίτητοι καθότι εξυπηρετούν το στόχο της μεγιστοποίησης της ρυθμαπόδοσης του συστήματος ενώ ταυτόχρονα εξασφαλίζουν ένα ελάχιστο δεδομένο ποσοστό ρυθμαπόδοσης για κάθε χρήστη.

1.3.3.3.1 Utilitarian Fairness(downlink scenario, [5])

Σε αυτήν την περίπτωση το πρόβλημα λαμβάνει την εξής μορφή:

$$\begin{aligned} & \text{maximize}_{Q \in \Theta} E\left(U_{Q(\bar{v})}\right) \\ & \text{subject to } E\left(U_i \cdot 1_{\{Q(\bar{v})=i\}}\right) \geq a_i \cdot E\left(U_{Q(\bar{v})}\right), \quad i=1,2,\dots,N \end{aligned} \quad (1.6)$$

όπου:

Q : η πολιτική που ακολουθεί ο χρονοδρομολογητής

Θ : το σύνολο όλων των στατικών πολιτικών

$E()$: αναμενόμενη τιμή. Προφανώς $E\left(U_{Q(\bar{v})}\right) = \sum_{i=1}^N E\left(U_i \cdot 1_{\{Q(\bar{v})=i\}}\right)$

\bar{U} : το διάνυσμα για όλους τους χρήστες της χρησιμότητας που αντλεί ο κάθε χρήστης i , $1 \leq i \leq N$ από την ανάθεση της χρονοσχισμής σε αυτόν

a_i : το ποσοστό της ρυθμαπόδοσης που λαμβάνει μακροπρόθεσμα ο χρήστης i .

Φυσικά $a_i \geq 0$ και $\sum_{i=1}^N a_i \leq 1$ και τέλος

$$1_A = \begin{cases} 1, & \text{αν αληθεύει το } A \\ 0, & \text{αλλιώς} \end{cases}$$

Αποδεικνύεται ότι η βέλτιστη πολιτική για την επίλυση του προβλήματος που μόλις διατυπώσαμε είναι η εξής:

$$Q^*(\vec{U}) = \arg \max_i \left[(\kappa + v_i^*) \cdot U_i \right]$$

όπου:

$$\kappa = 1 - \sum_{i=1}^N a_i \cdot v_i^*$$

και τα v_i^* είναι τέτοια ώστε :

(a) $\min_i v_i^* = 0$

(b) Για όλα τα i , $E\left(U_i \cdot 1_{\{Q(\vec{u})=i\}}\right) \geq a_i \cdot E\left(U_{Q(\vec{u})}\right)$, $i=1,2,\dots,N$

(c) Για όλα τα i , αν $E\left(U_i \cdot 1_{\{Q(\vec{u})=i\}}\right) > a_i \cdot E\left(U_{Q(\vec{u})}\right)$, $i=1,2,\dots,N$, τότε $v_i^* = 0$

Όμοια με την περίπτωση B.1, τα ποσοστά επί της ρυθμαπόδοσης που λαμβάνουν οι χρήστες είναι προκαθορισμένα, οπότε πρακτικά αυτό που μένει να υπολογιστεί είναι οι παράμετροι v_i^* . Η διαδικασία που ακολουθείται για την επίτευξη στοχαστικής σύγκλισης είναι ίδια με αυτήν της περίπτωσης B.1.

Σημείωση:

Η χρήση του προαναφερθέντος αλγορίθμου, εξασφαλίζει μεν ένα ελάχιστο ποσοστό ρυθμαπόδοσης για κάθε χρήστη αλλά ενέχει τον κίνδυνο ένας χρήστης με προκαθορισμένο υψηλό ποσοστό ρυθμαπόδοσης ο οποίος τυχαίνει να βιώνει «δυσμενείς» συνθήκες διαύλου να αποτελεί στενωπό στη ρυθμαπόδοση του συστήματος.

1.3.3.2 Μακροπρόθεσμη Δικαιοσύνη Ρυθμαπόδοσης (uplink scenario, [8])

Στην περίπτωση της ζεύξης ανόδου το πρόβλημα της μεγιστοποίησης της ρυθμαπόδοσης υπό περιορισμούς λαμβάνει αρκετά πιο περίπλοκη μορφή καθώς

πλην του χρόνου πρόσβασης και των συνθηκών του διαύλου, υπεισέρχονται στη διαδικασία λήψης απόφασης του χρονοδρομολογητή παράμετροι όπως η ισχύς εκπομπής και ο αριθμός των ταυτόχρονα εξυπηρετούμενων χρηστών. Εδώ, το πρόβλημα μοντελοποιείται αρχικά ως εξής:

$$\begin{aligned} \max \quad & E \left(\sum_{i=1}^{B(k)} r_i \right) \\ \text{s.t.} \quad & \frac{h_i \cdot p_i \cdot G_i}{\sum_{j=1, j \neq i}^{B(k)} h_j \cdot p_j + W \cdot \eta_0} \geq \gamma_i, \quad \text{για } i = 1, 2, \dots, B(k) \quad (1.7) \\ & p_i \leq p_i^{\max}, \quad \text{για } i = 1, 2, \dots, B(k) \\ & \frac{\bar{r}_i}{\phi_i} = \frac{\bar{r}_j}{\phi_j} \quad \text{για } 1 \leq i, j \leq B(k) \end{aligned}$$

όπου :

k : η υπό εξέταση χρονοσχισμή

$B(k)$: ο αριθμός των χρηστών για τους οποίους υπάρχουν δεδομένα προς μετάδοση τη χρονοσχισμή k

r_i : ο ρυθμός μετάδοσης δεδομένων για τον χρήστη i

γ_i : ο σηματοθορυβικός λόγος κατωφλίου του χρήστη i

h_i : το αντιστοιχούν στο γ_i κέρδος διαύλου

G_i : κέρδος εξάπλωσης, το οποίο ορίζεται ως $G_i = \frac{W}{r_i}$

W : το chip rate, το οποίο λαμβάνεται σταθερό

p_i : η ισχύς εκπομπής του χρήστη i

\bar{r}_i : η μέση τιμή του ρυθμού μετάδοσης του χρήστη i

ϕ_i : ένα προκαθορισμένο βάρος συσχετισμένο με την απαίτηση ποιότητας υπηρεσίας του χρήστη.

η_0 : η μονόπλευρη φασματική πυκνότητα ισχύος του προσθετικού, λευκού Gaussian θορύβου(AWGN)

Τελικά, το πρόβλημα της μεγιστοποίησης της ρυθμαπόδοσης υπό περιορισμούς δικαιοσύνης μοντελοποιείται ως πρόβλημα μεγιστοποίησης ρυθμαπόδοσης με βάρη. Δηλαδή, η αντικειμενική συνάρτηση λαμβάνει τη μορφή $\sum_{i=1}^{B(k)} w_i r_i$ όπου η επιλογή των βαρών w_i αντιπροσωπεύει τον προηγούμενο περιορισμό δικαιοσύνης $\frac{\bar{r}_i}{\phi_i} = \frac{\bar{r}_j}{\phi_j}$.

Όμοια με προηγούμενες περιπτώσεις, ο προσδιορισμός των βέλτιστων βαρών w_i^* είναι απαραίτητος και επιτυγχάνεται με τον εξής αλγόριθμο στοχαστικής σύγκλισης:

$$w_i(n+1) = w_i(n) - \frac{Y_n}{n} \quad (1.8)$$

όπου:

$$Y_n = \frac{r_i(n)}{E\left[\sum_j r_j(n)\right]} - \frac{\phi_i}{\sum_j \phi_j}$$

Και για την εκτίμηση της μέσης ρυθμαπόδοσης $E\left[\sum_j r_j(n)\right]$ χρησιμοποιούμε την εξομαλυμένη τιμή του $\bar{R}(n)$:

$$\bar{R}(n) = \bar{R}(n-1) \cdot \beta + (1-\beta) \cdot \sum_j r_j(n-1)$$

όπου β είναι ο παράγων εξομάλυνσης που καθορίζει πως το εκτιμώμενο $\bar{R}(n)$ ακολουθεί τη μεταβολή της πραγματικής επιτευχθείσας ρυθμαπόδοσης του συστήματος.

Σημείωση:

- i) Η βέλτιστη λύση του προαναφερθέντος αρχικού προβλήματος, η οποία προκύπτει μέσω της μετατροπής του σε δυαδικό πρόβλημα της μορφής knapsack έχει την ακόλουθη θεμελιώδη ιδιότητα:

$$p_i(k) \in \{0, p_i^{\max}\}, i = 1, 2, \dots, B(k)$$

Δηλαδή, η βέλτιστη λύση επιτυγχάνεται όταν ένας χρήστης είτε εκπέμπει με μέγιστη ισχύ είτε δεν εκπέμπει καθόλου.

1.4 Μελέτη απόδοσης αλγορίθμων υψηλής ρυθμαπόδοσης ως προς πιθανοτικούς βραχυπρόθεσμους περιορισμούς καθυστέρησης και ρυθμαπόδοσης

Σε αυτό το κεφάλαιο θα μετρήσουμε την συμπεριφορά ευκαιριακών αλγορίθμων χρονοδρομολόγησης ως προς βραχυπρόθεσμους περιορισμούς που εκφράζουν τις ανάγκες υπηρεσιών πραγματικού χρόνου [9].

Ας θεωρήσουμε ότι S είναι το σύνολο όλων των ενεργών χρηστών ενός συστήματος, με $i_Q^*(t) = i$ την πολιτική χρονοδρομολόγησης Q που επιλέγει την εξυπηρέτηση του χρήστη i την χρονοσχισμή t , και με $\Pr[i_Q^*(t) = i]$ την πιθανότητα του χρήστη i να επιλεγεί κάτω από την πολιτική Q κάποια χρονική στιγμή.

Θεωρώντας ότι ο χρονοδρομολογητής έχει γνώση των καναλιών των χρηστών κάθε χρονική στιγμή, έχει αποδειχθεί ότι η πιθανότητα ένας χρήστης να

επιλεγεί κάποια χρονική στιγμή t είναι ίση με την πιθανότητα να επιλεγεί από το σύστημα οποιαδήποτε χρονική στιγμή. Επιπλέον έχει αποδειχθεί ότι υπό σταθερές πιθανότητες επιλογής κάθε χρήστη από στο σύστημα, τότε η καθυστέρηση ανάμεσα σε διαδοχικές προσβάσεις του στο σύστημα ακολουθεί γεωμετρική συνάρτηση κατανομής πιθανότητας που δίνεται από τον τύπο

$$\Pr[d_i \geq T] = [1 - \Pr[i_Q^*(t) = i]]^T \quad (1.8)$$

όπου d_i είναι μια τυχαία μεταβλητή που αναπαριστά την καθυστέρηση ενός χρήστη, εκφρασμένη σαν αριθμός διαδοχικών χρονοσχισμών που αυτός δεν έχει αποκτήσει πρόσβαση στο σύστημα και T είναι το κατώφλι της μέγιστης καθυστέρησης.

Η εξυπηρέτηση ενός χρήστη υπηρεσιών πραγματικού χρόνου κάνει επιτακτική την ανάγκη σταθερών πιθανοτήτων παραβίασης του κατωφλιού καθυστέρησης του.

Για να μπορέσουμε να μελετήσουμε καλύτερα την ποιότητα της υπηρεσίας των χρηστών πραγματικού χρόνου, πρέπει να εξετάσουμε την πιθανότητα αυτός να πάρει ποσότητα δεδομένων, μικρότερη από ένα προκαθορισμένο κατώφλι, μέσα σε ένα παράθυρο παρατήρησης. Αν ορίσουμε με b_i την τυχαία μεταβλητή που εκφράζει την ποσότητα των δεδομένων που έχει λάβει ένας χρήσης μέσα σε παράθυρο T_{b_i} χρονοσχισμών, τότε η πιθανότητα η ποσότητα b_i να είναι μικρότερη από ένα προκαθορισμένο κατώφλι B_i δίνεται από τον τύπο $\Pr[b_i \leq B_i]_{T_{b_i}}$ και για να εξασφαλίζεται η ποιότητα της υπηρεσίας πρέπει να είναι άνω φραγμένη.

Θεωρώντας ότι ο ρυθμός μετάδοσης δεδομένων ενός χρήστη διατηρείται σταθερός και ίσος με τον μέσο ρυθμό μετάδοσης κάτω από την πολιτική χρονοδρομολόγησης Q τότε ως

$$\Pr[b_i < B_i(k, \bar{R}_{Q,i})]_{T_s} = \sum_{k'=0}^{k-1} \binom{T_s}{k'} (\Pr[i_Q^*(t) = i])^{k'} (1 - \Pr[i_Q^*(t) = i])^{T_s - k'} \quad (1.9)$$

όπου $k=0,1,\dots, T_{Bi}$, $B_i(k, \bar{R}_{Q,i}) = k \cdot \bar{R}_{Q,i} \cdot T_s$ και T_s είναι η διάρκεια μιας χρονοσχιμής

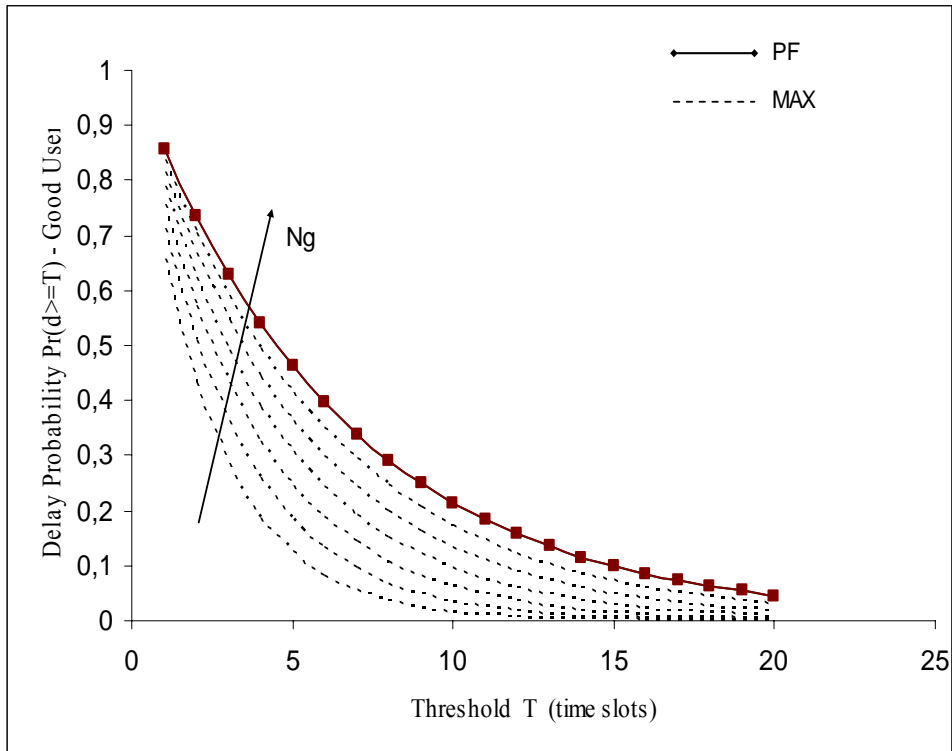
Υπό αυτό το πρίσμα θα μελετηθούν τρεις πολιτικές χρονοδρομολόγησης

1. άπλειστη ευκαιριακή χρονοδρομολόγηση (MAX)
2. αλγόριθμος αναλογικής δικαιοσύνης [2]
3. αλγόριθμος συσχετιστικά καλύτερου χρήστη (Relatively Best algorithm)

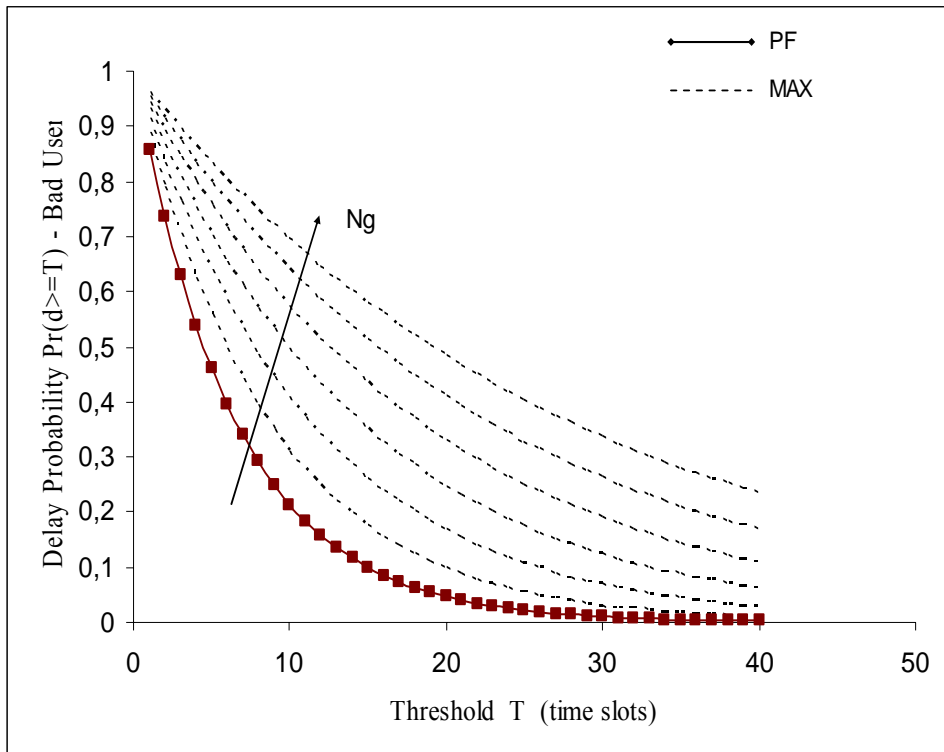
Τα παρακάτω αποτελέσματα έχουν προκύψει από ένα σύστημα που έχει 7 μόνιμους χρήστες (continuously backlogged users) που χωρίζονται σε καλούς και κακούς χρήστες (ως προς το μέσο κανάλι τους).

Στο πρώτο σενάριο έχουμε $\bar{\rho}_{g,sc1} = 3dB$ and $\bar{\rho}_{b,sc1} = 0dB$.

Στο δεύτερο σενάριο έχουμε $\bar{\rho}_{g,sc1} = 6dB$ and $\bar{\rho}_{b,sc1} = 3dB$.



Σχήμα 1.1

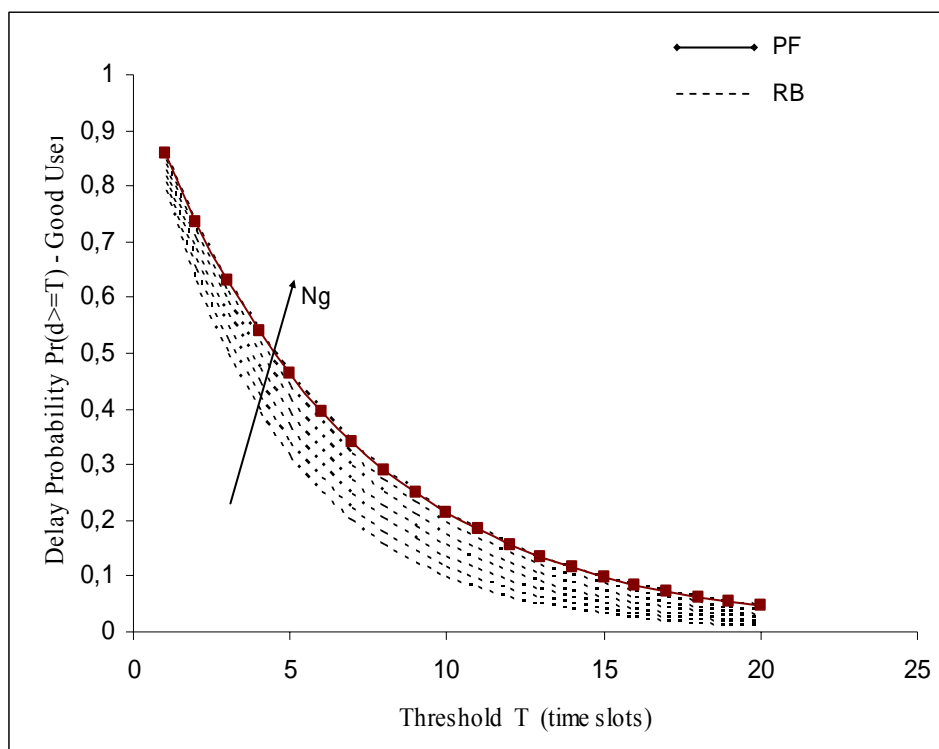


Σχήμα 1.2

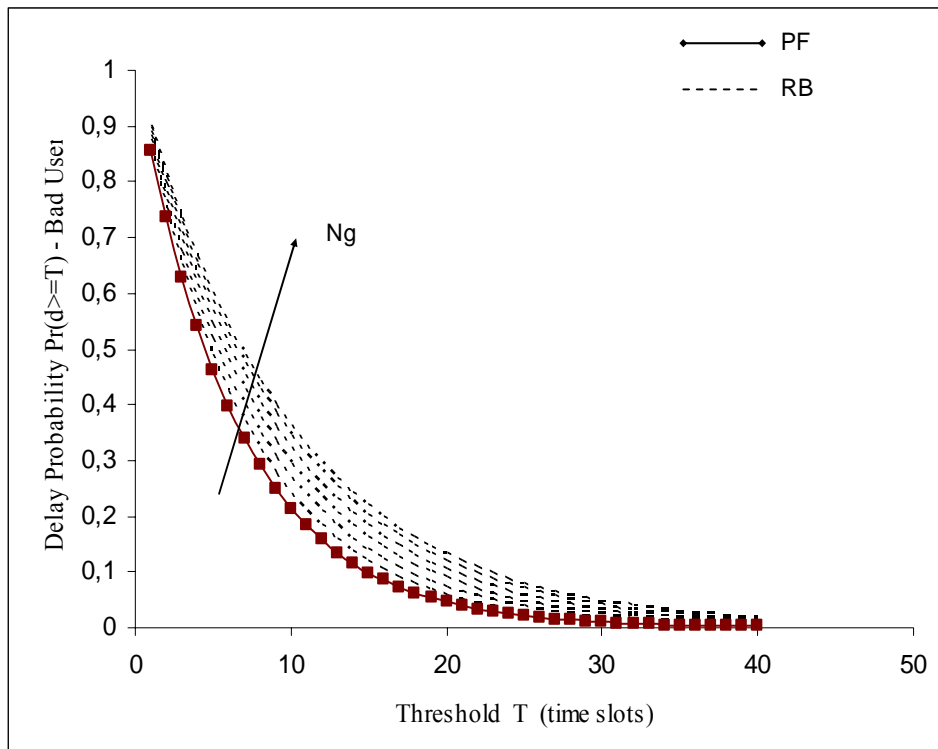
Αρχικά δίνουμε βάση στις πιθανότητες καθυστέρησης των χρηστών. Στις παραπάνω εικόνες (1,2) παρατηρούμε την πιθανότητα καθυστέρησης όπως αυτή περιγράφηκε στον τύπο 1.8 για καλούς και κακούς χρήστες αντίστοιχα, καθώς ο αριθμός των καλών χρηστών αυξάνει ($N_g = 1,2,\dots,7$). Οι αλγόριθμοι που εξετάζονται είναι ο MAX και ο PF. Στα παραπάνω σχήματα θεωρούμε μια υπηρεσία αποτυχημένη αν $\Pr[d \geq T] \leq 0.001$.

Παρατηρούμε ότι κάτω από την πολιτική RF όλοι οι χρήστες έχουν σταθερές πιθανότητες να αποκτήσουν πρόσβαση στο σύστημα κάθε χρονική στιγμή και κατά συνέπεια επιτυγχάνεται αφενός μακροπρόθεσμη δικαιοσύνη ως προς τον χρόνο πρόσβασης στο σύστημα, αφετέρου δικαιοσύνη σχετικά με την εξασφάλιση των περιορισμών καθυστέρησης μεταξύ ισοδύναμων χρηστών.

Όπως φαίνεται στις εικόνες 1,2,3,4 οι απαιτήσεις καθυστέρησης των χρηστών υπηρεσιών πραγματικού χρόνου δεν εξασφαλίζονται, ούτε και στις περιπτώσεις χρηστών με καλό κανάλι.



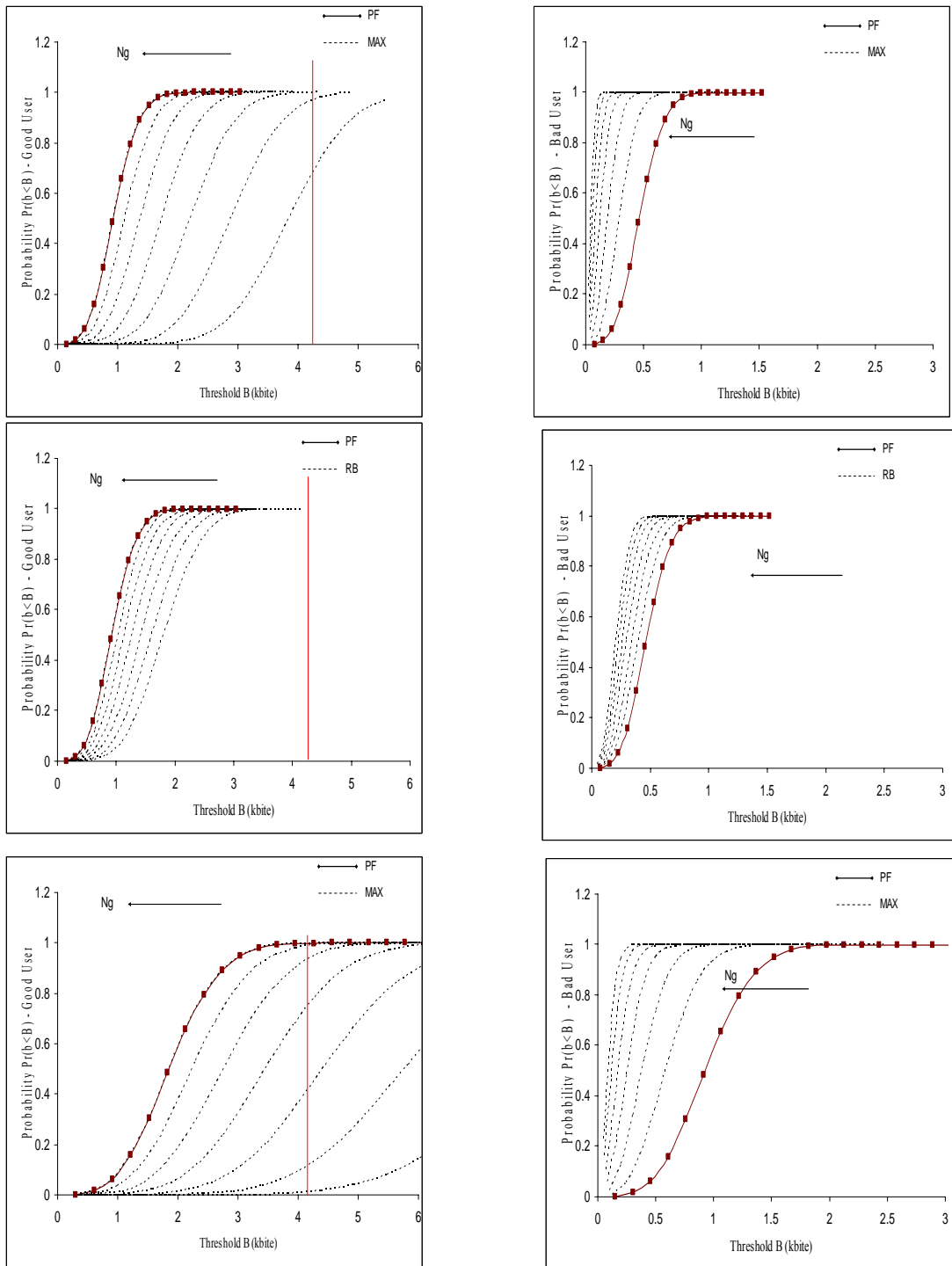
Σχήμα 1.3



Σχήμα 1.4

Στις εικόνες 3 και 4 παρατηρούμε τις πιθανότητες καθυστέρησης για καλούς και κακούς χρήστες κάτω από τις πολιτικές PF και RB.

Παρατηρούμε ότι σε κάθε σενάριο οι πιθανότητες των περιορισμών καθυστέρησης εξαρτώνται από τον αριθμό των χρηστών στο σύστημα και από την μέση ποιότητα του καναλιού του καθενός. Από την στιγμή που δεν μπορούν να ρυθμιστούν οι παραπάνω παράμετροι γίνεται δύσκολο να εξασφαλιστούν οι αντίστοιχοι περιορισμοί.



Σχήμα 1.5

Τέλος στη ομάδα σχημάτων 5 παρουσιάζονται τα αποτελέσματα που προέκυψαν από την εξέταση των παραπάνω αλγορίθμων σε σχέση με την ικανοποίηση των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης υπηρεσιών πραγματικού χρόνου τα οποία εκφράζονται από τις κατακόρυφες κόκκινες γραμμές. Παρατηρείτε ότι

κανένας από τους προς εξέταση αλγορίθμους δεν ικανοποιεί εν τέλει τα προς εξέταση βραχυπρόθεσμα κριτήρια, για κανένα πλήθος καλών και κακών χρηστών στο σύστημα σε σχέση με την ποιότητα καναλιού τους.

2 Αλγόριθμος UPR

Στη συνέχεια θα παρουσιάσουμε τον αλγόριθμο UPR που πρότειναν οι Jang Won Lee, Ravi R. Mazumdar και Ness B. Shroff. Ο αλγόριθμος αυτός μελετά την ανάθεση πόρων σε ετερογενή περιβάλλοντα με σκοπό την βελτιστοποίηση της συνολικής συνάρτησης χρησιμοποίησης (utility function) των χρηστών σε κάθε χρονική στιγμή. Πάνω στον αλγόριθμο αυτό έχουμε στηριχθεί για να μοντελοποιήσουμε τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης (short-term throughput constraints) που μελετάμε.

Ο αλγόριθμος UPR χρησιμοποιείται για την επίλυση δύο προβλημάτων το πρώτο είναι το πρόβλημα απόδοσης ισχύος από την κεραία στους κινητούς σταθμούς κατά την ζεύξη καθόδου και το δεύτερο είναι το συνδυασμένο πρόβλημα ανάθεσης ρυθμού μετάδοσης και ισχύος κατά την ζεύξη καθόδου. Ο ίδιος αλγόριθμος με μικρές μετατροπές χρησιμοποιείται για την ανάλυση και των δύο προβλημάτων. Για την καλύτερη κατανόηση του αλγορίθμου αρχικά θα περιγραφεί ο αλγόριθμος επίλυσης του απλούστερου προβλήματος ανάθεσης ισχύος και στην συνέχεια θα αναφερθούν οι επεκτάσεις του για την επίλυση του δεύτερου προβλήματος.

2.1 Αλγόριθμος UPR για ανάθεση ισχύος

Πριν μελετήσουμε τον αλγόριθμο θα παρουσιάσουμε το μοντέλο περιγραφής πάνω στο οποίο έχει στηριχθεί καθώς και θα αποδείξουμε την βελτιστότητα του αλγορίθμου χρησιμοποιώντας τεχνικές από την θεωρία βελτιστοποίησης.

2.1.1 Μοντέλο περιγραφής

Αρχικά θεωρούμε μόνο μία κυψέλη ενός δικτύου που περιέχει ένα σταθμό βάσης (κεραία) και M κινητούς σταθμούς. Το σύστημα μας λειτουργεί σε προκαθορισμένες χρονοσχισμές τέτοιες που μπορούμε να υποθέσουμε ότι κατά τη διάρκεια μιας χρονοσχισμής ο θόρυβος, η ποιότητα καναλιού και η διακυβελωτή παρεμβολή κάθε χρήστη είναι σταθερή. Το σύστημα μας μπορεί να εξυπηρετήσει χρήστες που απαιτούν διαφορετικές υπηρεσίες (ετερογενές περιβάλλον). Σκοπός είναι η απόδοση ισχύος σε κάθε κινητό σταθμό από την κεραία κατά την επίλυση του προβλήματος downlink. Ο περιορισμός λειτουργίας είναι ότι η κεραία έχει συγκεκριμένο άνω όριο ισχύος P_T που μπορεί να αποδώσει.

Ποιότητα σήματος

Στην συνέχεια, περιγράφουμε τη παράμετρο «ποιότητας σήματος» κάθε χρήστη που θα χρησιμοποιηθεί για τον ορισμό της συνάρτησης χρησιμοποίησης. Η παράμετρος αυτή εξαρτάται από την αποδιδόμενη ισχύς σε κάθε χρήστη και εξαρτάται από το στιγμιαίο κανάλι του και την παρεμβολή από τις ισχύς των υπολοίπων χρηστών που βρίσκονται γύρω του. Η παράμετρος αυτή δίνεται από τον τύπο

$$\begin{aligned} \gamma_i(\bar{P}) &= \frac{N_i G_i P_i}{G_i \theta (\sum_{m=1}^M P_m - P_i) + I_i} \\ &= \frac{N_i P_i}{\theta (\sum_{m=1}^M P_m - P_i) + A_i}, \end{aligned} \quad (2.1)$$

Όπου

- P_i : ισχύς που παραχωρείται στο χρήστη i

- N_i : κέρδος επεξεργασίας (processing gain) για το χρήστη i που ορίζεται σαν W_i/R_i
- W_i : ρυθμός ψηφίδας (chip rate) για το CDMA δίκτυο
- R_i : ρυθμός δεδομένων για το χρήστη i
- G_i : κέρδος μονοπατιού
- I_i : θόρυβος για το χρήστη i
- A_i : ποιότητα περιβάλλοντος εκπομπής που ορίζεται σαν I_i/G_i
- M : πλήθος χρηστών
- θ : παράγοντας ορθογωνικότητας $0 \leq \theta \leq 1$

Λήμμα 1: Αν $\bar{P} = (P_1, P_2, \dots, P_M)$ η βέλτιστη ανάθεση τιμών που μεγιστοποιεί το άθροισμα των συναρτήσεων χρησιμοποίησης και $\sum_{i=1}^M P_i < P_T$ για την ανάθεση αυτή τότε μπορούμε να βρούμε μία διαφορετική ανάθεση $\bar{P}^* = (P_1^*, P_2^*, \dots, P_M^*)$ όπου $\sum_{i=1}^M P_i^* = P_T$ και το άθροισμα των συναρτήσεων χρησιμοποίησης για κάθε χρήστη είναι μεγαλύτερο.

Απόδειξη: Αφού από τον ορισμό της συνάρτησης χρησιμοποίησης είναι αύξουσα ως προς την μεταβλητή των πόρων (στην περίπτωση μας η ισχύς P) τότε αν διαμοιράσουμε ισόποσα την διαφορά $P_T - \sum_{i=1}^M P_i$ σε κάθε κινητό σταθμό

αποδίδοντας ισχύς ίση με $P_i^* = P_i + \frac{(P_T - \sum_{i=1}^M P_i)}{M}$ τότε η τιμή της συνάρτησης χρησιμοποίησης για τη νέα ανάθεση θα είναι μεγαλύτερη και άρα η συνολική τιμή της θα είναι μεγαλύτερη.

Με βάση το παραπάνω λήμμα έχουμε ότι η βέλτιστη ανάθεση τιμών πρέπει να ικανοποιεί την σχέση $\sum_{i=1}^M P_i = P_T$ και έτσι η συνάρτηση γ έχει την μορφή

$$\begin{aligned} \gamma_i(\bar{P}) &= \frac{N_i P_i}{\theta(\sum_{m=1}^M P_m - P_i) + A_i} \\ &= \frac{N_i P_i}{\theta(P_T - P_i) + A_i} \\ &\triangleq \gamma_i(P_i), \quad i = 1, 2, \dots, M. \end{aligned} \quad (2.2)$$

Που όπως παρατηρούμε πλέον εξαρτάται μόνο από την ισχύ που αποδίδεται σε κάθε χρήστη και όχι από το συνολικό διάνυσμα απόδοσης.

2.1.2 Συνάρτηση χρησιμοποίησης

Η συνάρτηση χρησιμοποίησης που θα χρησιμοποιηθεί για την επίλυση του παραπάνω προβλήματος είναι η εξής.

$$U_i(\gamma_i(P)) = R_i f_i(\gamma_i(P)) \quad (2.3)$$

Όπου η συνάρτηση f ικανοποιεί τα κριτήρια μιας συνάρτησης χρησιμοποίησης που αναφέραμε παραπάνω και μπορεί να έχει έναν από τους διαφορετικούς τύπους (κυρτή, κοίλη, σιγμοειδής) που αναφέρθηκε παραπάνω. Η συνάρτηση αυτή εξαρτάται από την παράμετρο γ που εκφράζει την «ποιότητα σήματος» κάθε χρήστη.

2.1.3 Βασικό Πρόβλημα και απόδειξη βελτιστότητας

Με βάση τα παραπάνω το πρόβλημα βελτιστοποίησης είναι το ακόλουθο

$$\begin{aligned} \max_{\bar{P}} \quad & \sum_{i=1}^M U_i(\gamma_i(P_i)) \\ \text{subject to} \quad & \sum_{i=1}^M P_i \leq P_T, \\ & 0 \leq P_i \leq P_T, \quad i = 1, 2, \dots, M. \end{aligned} \quad (2.4)$$

Που εκφράζει την ανάγκη μας σε κάθε χρονική στιγμή να βρούμε την ανάθεση ισχύος που μεγιστοποιεί το άθροισμα των συναρτήσεων χρησιμοποίησης σε όλους τους χρήστες με περιορισμό ότι η συνολική ανάθεση δε μπορεί να υπερβαίνει τις φυσικές δυνατότητες του σταθμού βάσης.

Στο παραπάνω πρόβλημα μπορούμε με βάση τη θεωρία βελτιστοποίησης να δημιουργήσουμε το δυαδικό πρόβλημα χρησιμοποιώντας της παράμετρο Laplace λ . Το δυαδικό πρόβλημα γίνεται ως εξής

$$\begin{aligned} L(\bar{P}, \lambda) &= \sum_{i=1}^M U_i(\gamma_i(P_i)) + \lambda(P_T - \sum_{i=1}^M P_i), \\ S &= \{\bar{P} | \bar{0} \leq \bar{P} \leq \bar{P}_T\}, \end{aligned} \quad (2.5)$$

$$Y(\lambda) = \{\bar{x} \in S \mid L(\bar{x}, \lambda) = \max_{\bar{P} \in S} \{L(\bar{P}, \lambda)\}\},$$

Και ορίζουμε

ότι

(2.6)

όπου $Y(\lambda)$ είναι το διάνυσμα ανάθεσης ισχύος που μεγιστοποιεί το δυαδικό πρόβλημα L για μία συγκεκριμένη τιμή της παραμέτρου λ . Με βάση τη θεωρία βελτιστοποίησης για κάθε $\lambda \geq 0$ το διάνυσμα ανάθεσης $\bar{P}(\lambda) \in Y(\lambda)$ αποτελεί βέλτιστη λύση του παρακάτω προβλήματος βελτιστοποίησης

$$\begin{aligned} & \max_{\bar{P}} \sum_{i=1}^M U_i(\gamma_i(P_i)) \\ & \text{subject to } \sum_{i=1}^M P_i \leq \sum_{i=1}^M P_i(\lambda) \\ & \quad 0 \leq P_i \leq P_T, \quad i = 1, 2, \dots, M \end{aligned} \quad (2.7)$$

Το παραπάνω πρόβλημα διαφέρει από αυτό που αναφέραμε αρχικά στη παράμετρο $\sum_{i=1}^M P_i \leq \sum_{i=1}^M P_i(\lambda)$. Η λύση του παραπάνω προβλήματος ταυτίζεται με την λύση του αρχικού όταν το λ επιλεγεί με τέτοιο τρόπο ώστε $\sum_{i=1}^M P_i(\lambda) = P_T$. Η ακρίβεια της λύσης που δίνεται από το παραπάνω πρόβλημα εξαρτάται από την τιμή της $\left| \sum_{i=1}^M P_i(\lambda) - P_T \right|$. Όσο πιο κοντά στο μηδέν είναι η τιμή αυτή τόσο πιο κοντά στη βέλτιστη λύση του αρχικού προβλήματος είναι η λύση του αντίστοιχου δυαδικού του. Με βάση αυτή την παρατήρηση μπορούμε να μετασχηματίσουμε το πρόβλημα στο παρακάτω

$$\begin{aligned} & \min_{\lambda} \left\{ P_T - \sum_{i=1}^M P_i(\lambda) \right\} \\ & \text{subject to } \bar{P}(\lambda) = \arg \max_{0 \leq \bar{P} \leq \bar{P}_T} \{L(\bar{P}, \lambda)\} \\ & \quad \sum_{i=1}^M P_i(\lambda) \leq P_T. \end{aligned} \quad (2.8)$$

Η παραπάνω συνθήκη αφού το πρόβλημα είναι διαχωρίσιμο γίνεται

$$P_i(\lambda) \in \{0 \leq q \leq P_T \mid L_i(q, \lambda) = \max_{0 \leq P \leq P_T} L_i(P, \lambda)\}$$

Όπου $L_i(x) = U_i(\gamma_i(x)) - \lambda x$.

Μπορούμε να ερμηνεύσουμε φυσικά την λαπλασιανή παράμετρο λ σαν μια τιμή κόστους ανά μονάδα ισχύος που επιβάλλει ο σταθμός βάσης στην ισχύ που θα διαθέσει. Το νέο αυτό πρόβλημα μπορεί να περιγραφεί σαν την προσπάθεια της βάσης να βρει την μεγαλύτερη τιμή κόστους για να μπορέσει να «πουλήσει» όλη την ισχύ που διαθέτει. Κάθε κινητός σταθμός μπορεί να «αγοράσει» μέχρι ένα συγκεκριμένο αριθμό ισχύος ανάλογα με την τιμή κόστους που επιβάλλει η βάση. Όσο μεγαλώνει η τιμή του κόστους τόσο λιγότερο μπορεί να «αγοράσει» κάθε χρήστης. Η ποσότητα αγοράς για κάθε χρήστη καθορίζεται από την $P_i(\lambda)$. Με βάση την παραπάνω ανάλυση βλέπουμε εύκολα ότι ο αλγόριθμος ανάθεσης πρέπει να χωριστεί σε δύο φάσεις. Η πρώτη φάση περιέχει το καθορισμό της παραμέτρου λ και κατ'επέκταση των κινητών σταθμών που μπορούν να συμμετέχουν στον ανταγωνισμό της ισχύος για αυτό και ονομάζεται αλγόριθμος επιλογής κινητών σταθμών (mobile selection algorithm). Στην συνέχεια αφού καθοριστούν οι σταθμοί που θα συμμετέχουν στον ανταγωνισμό, καθορίζεται το ποσό ισχύος που ανατίθεται σε κάθε χρήστη με σκοπό να αποδοθεί όλη η ποσότητα P_T που μπορεί να διαθέσει η βάση.

Λήμμα 2: $P_i(\lambda) = \arg \max_{0 \leq P \leq P_T} \{U_i(\gamma_i(P)) - \lambda P\}$ Αν τότε είτε $P_i(\lambda) = 0$ είτε $P_i(\lambda) = P_T$ είτε η P βρίσκεται στην κυρτή περιοχή.

Απόδειξη: Παραγωγίζοντας την συνάρτηση προς μεγιστοποίηση έχουμε ότι

$$\left. \frac{\partial U_i(\gamma_i(P))}{\partial P} \right|_{P=P(\lambda)} = \lambda \text{ και } \left. \frac{\partial^2 U_i(\gamma_i(P))}{\partial P^2} \right|_{P=P(\lambda)} < 0 \text{ που αποδεικνύει το λήμμα.}$$

Με βάση το λήμμα 2 κάθε χρήστης που έχει κοίλη συνάρτηση χρησιμοποίησης μπορεί να δεχθεί είτε μηδενική είτε τη μέγιστη ισχύς.

Στην συνέχεια για την περιγραφή του αλγορίθμου θα ορίσουμε την ποσότητα

$$\lambda_i^{max} = \arg \min_{0 \leq \lambda \leq \infty} \left\{ \max_{0 \leq P \leq P_T} \{U_i(\gamma_i(P)) - \lambda P\} = 0 \right\} \quad (2.9)$$

Τέτοια ώστε για κάθε $\lambda > \lambda_i^{max}$ $P_i(\lambda) = 0$. Ο υπολογισμός και η απόδειξη ύπαρξης της ποσότητας αυτής γίνεται με τον παρακάτω τρόπο. Αρχικά ορίζω την παράμετρο

$$P_i^o = \begin{cases} P^*, & \text{if } \frac{\partial^2 U_i(\gamma_i(P))}{\partial P^2} \Big|_{P=P^*} = 0, 0 \leq P^* \leq P_T, \\ 0, & \text{if } \frac{\partial^2 U_i(\gamma_i(P))}{\partial P^2} < 0 \text{ for } 0 \leq P \leq P_T, \\ P_T, & \text{if } \frac{\partial^2 U_i(\gamma_i(P))}{\partial P^2} > 0 \text{ for } 0 \leq P \leq P_T, \end{cases} \quad (2.10)$$

Που είναι το γενικευμένο σημείο καμπής κάθε μίας από τις διαφορετικές συναρτήσεις χρησιμοποίησης (κυρτή, κοίλη και σιγμοειδής). Αρχικά για την σιγμοειδή περίπτωση όπου υπάρχει σημείο καμπής εσωτερικά στο πεδίο ορισμού του P. Από το λήμμα 3 για την σιγμοειδή συνάρτηση έχουμε ότι το $P_i(\lambda)$ πρέπει να βρίσκεται στην κυρτή περιοχή. Από τα παραπάνω μπορούμε να αποδείξουμε ότι η $P_i(\lambda)$ είναι φθίνουσα συνάρτηση του λ και το λ_i^{max} πρέπει να είναι το σημείο όπου $P_i(\lambda_{max}) = 0$ άρα για την περίπτωση της σιγμοειδούς το λ_i^{max} δίνεται από το τύπο

$$\frac{\partial U_i(\gamma_i(P))}{\partial P} \Big|_{P=P^*} = \lambda_{max} \quad \text{με } P^* \text{ να είναι η λύση της εξίσωσης}$$

$$U_i(\gamma_i(P)) - P \frac{\partial U_i(\gamma_i(P))}{\partial P} = 0, \quad P_i^o \leq P \leq P_T \quad (2.11)$$

Στην περίπτωση που εξετάζουμε κυρτή συνάρτηση έχουμε ότι η $P_i(\lambda)$ είναι φθίνουσα συνάρτηση του λ και ισχύει το παρακάτω

$$P_i(\lambda) = \begin{cases} P_T, & \text{if } \lambda < \frac{\partial U_i(\gamma_i(P))}{\partial P} \Big|_{P=P_T}, \\ P^*, & \text{if } \lambda = \frac{\partial U_i(\gamma_i(P))}{\partial P} \Big|_{P=P^*}, \quad 0 \leq P^* \leq P_T, \\ 0, & \text{if } \lambda > \frac{\partial U_i(\gamma_i(P))}{\partial P} \Big|_{P=0}. \end{cases} \quad (2.12)$$

Από το παραπάνω βλέπουμε ότι $\lambda_{\max} = \frac{\partial U_i(\gamma_i(P))}{\partial P} \Big|_{P=0}$ αφού ισχύει ότι για λ μεγαλύτερο της ποσότητας αυτής ισχύει ότι $P(\lambda)=0$.

Τέλος στην περίπτωση κοίλης συνάρτησης έχουμε ότι η $P(\lambda)$ είναι αύξουσα συνάρτηση ως προς λ και άρα ορίζουμε ως λ_{\max} τη τιμή που μεγιστοποιεί την συνάρτηση $P(\lambda)$ και άρα δίνεται από τον τύπο

$$\lambda_{\max} = \frac{U_i(\gamma_i(P_T))}{P_T} \quad (2.13)$$

Συγκεντρώνοντας τους παραπάνω τύπους έχω τον τελικό τύπο για το λ_{\max} ως εξής

$$\lambda_i^{\max} = \begin{cases} \frac{\partial U_i(\gamma_i(P))}{\partial P} \Big|_{P=0}, & \text{if } P_i^o = 0, \\ \frac{\partial U_i(\gamma_i(P))}{\partial P} \Big|_{P=P^*}, & \text{if } 0 < P_i^o < P_T \\ & \text{and } P^* \text{ exists,} \\ \frac{U_i(\gamma_i(P_T))}{P_T}, & \text{otherwise,} \end{cases} \quad (2.14)$$

Όπου P^* είναι η λύση της εξίσωσης

$$U_i(\gamma_i(P)) - P \frac{\partial U_i(\gamma_i(P))}{\partial P} = 0, \quad P_i^o \leq P \leq P_T \quad (2.15)$$

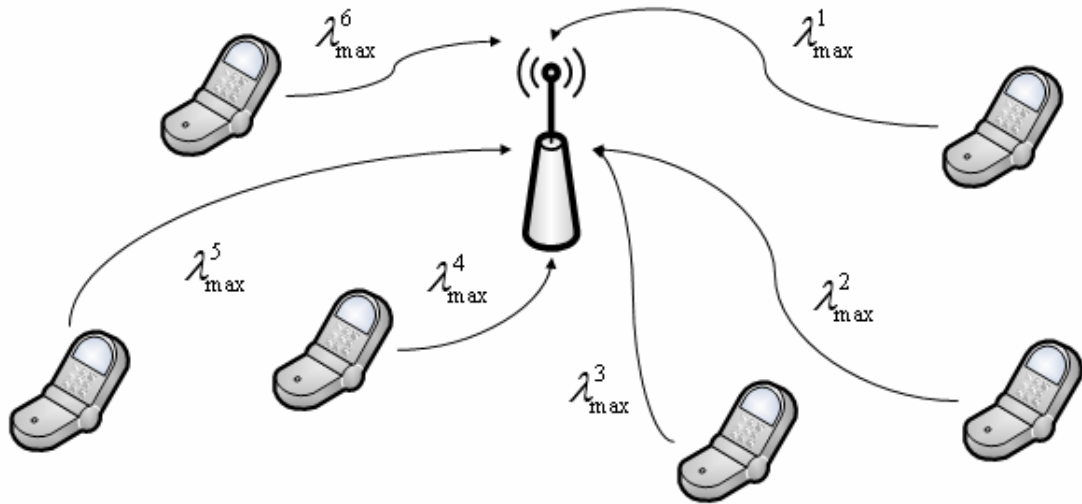
Με βάση τα παραπάνω μπορούμε να περιγράψουμε το λ_{\max} κάθε χρήστη σαν τη μέγιστη τιμή που αυτός μπορεί να «αντέξει», δηλαδή για τιμές μεγαλύτερες από αυτή τη τιμή δεν ανατίθεται τιμή ισχύος στο χρήστη άρα δεν συμμετέχει στον ανταγωνισμό ανάθεσης τιμών. Η παράμετρος λ_{\max} υπολογίζεται τοπικά από κάθε χρήστη και αναφέρεται στο σταθμό βάση κατά την εκτέλεση του

αλγορίθμου. Η παράμετρος αυτή είναι μια συμπυκνωμένη πληροφορία για κάθε κινητό σταθμό αφού περιέχει στοιχεία για την ποιότητα του καναλιού του χρήστη (η ποιότητα του καναλιού εκφράζεται από την παράμετρο A που περιέχεται στον υπολογισμό της συνάρτησης χρησιμοποίησης U κατά την εύρεση του $P(\lambda)$) και για το είδος της υπηρεσίας που απαιτεί (η πληροφορία αυτή περιέχεται στα διαφορετικά είδη συναρτήσεων χρησιμοποίησης που εκφράζουν την ευχαρίστηση κάθε χρήστη).

2.2 Αλγόριθμος επιλογής κινητών σταθμών

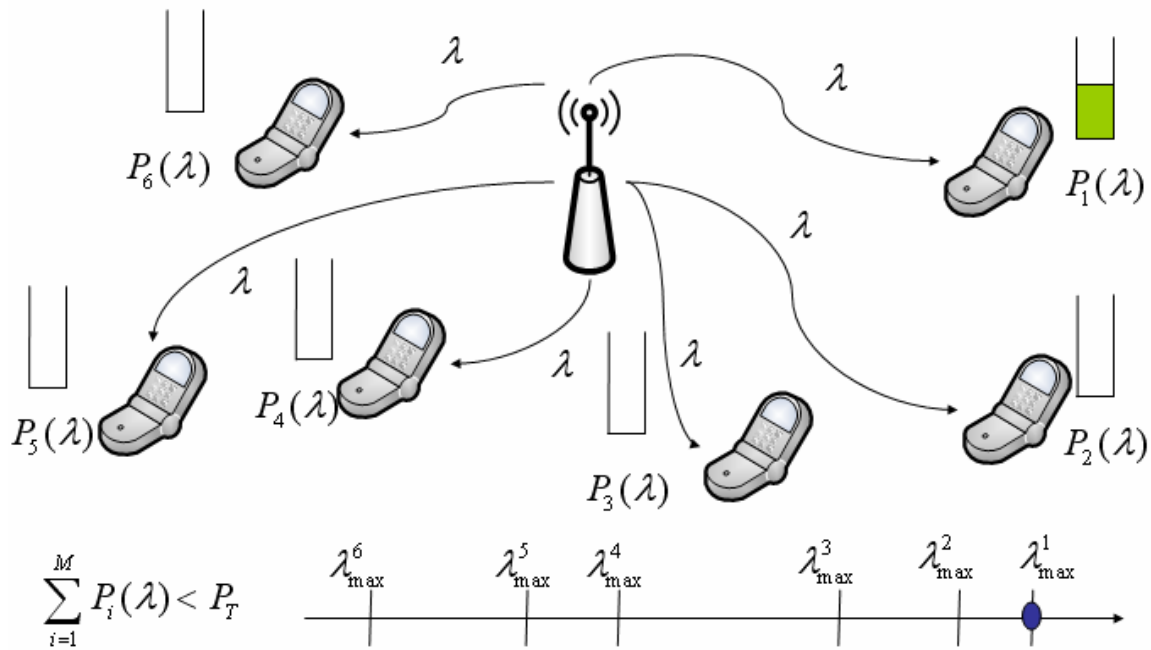
Με βάση τη φυσική ερμηνεία του μετασχηματισμένου προβλήματος βελτιστοποίησης και της παραμέτρου λ_{\max} κάθε χρήστη μπορούμε να περιγράψουμε την πρώτη φάση εκτέλεσης του αλγορίθμου. Σε αυτή τη φάση καθορίζεται η μέγιστη τιμή του λ που δίνει τη δυνατότητα στο σταθμό βάσης να αποδώσει όλη την ισχύ που μπορεί. Ο καθορισμός της τιμής λ επιτρέπει σε ένα υποσύνολο των M χρηστών να συμμετέχει στη δεύτερη φάση του αλγορίθμου δηλαδή στη τελική ανάθεση ισχύος. Ο λόγος είναι ότι η τιμή λ που επιλέχθηκε μπορεί να είναι μεγαλύτερη της παραμέτρου λ_{\max} κάποιων χρηστών στους οποίους αποδίδεται μηδενική ισχύς. Πρέπει να τονιστεί ότι ο αλγόριθμος UPR εκτελείται σε κάθε χρονοσχισμή άρα κάθε φορά επιλέγονται διαφορετικοί χρήστες και η επιλογή εξαρτάται από τη παράμετρο λ_{\max} άρα από το τρέχων κανάλι και την συνάρτηση χρησιμοποίησης τους. Για καλύτερη περιγραφή της εκτέλεσης του αλγορίθμου δίνονται στη συνέχεια τα βήματα εκτέλεσης με τη βοήθεια εικόνων

Στην αρχή κάθε χρονοσχισμός κάθε κινητός σταθμός ενημερώνει το σταθμό βάσης με το λ_{max}^i που έχει υπολογιστεί τοπικά όπως φαίνεται στο παρακάτω σχήμα



σχήμα 2.1

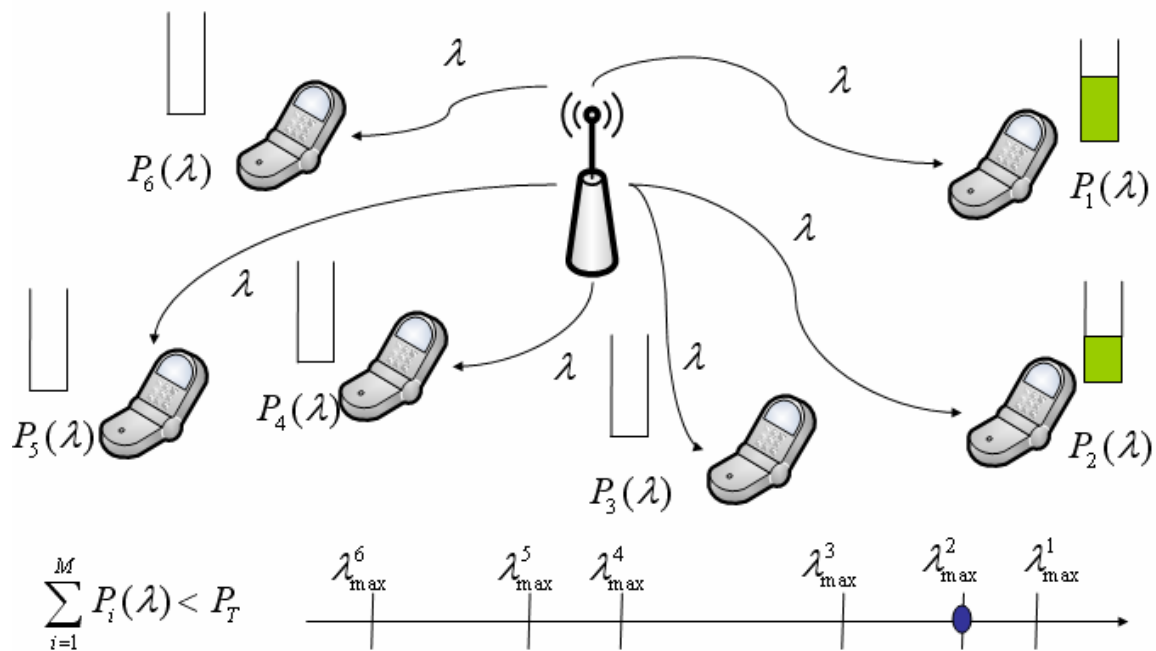
Στην συνέχεια ο σταθμός βάσης ταξινομεί τα λ_{max} από το μεγαλύτερο στο μικρότερο και αρχικά επιλέγει σαν τιμή κόστους το μέγιστο λ_{max} που έχει λάβει στο προηγούμενο βήμα. Αφού στείλει την τιμή αυτή στους κινητούς σταθμούς κάθε σταθμός υπολογίζει την τιμή $P^i(\lambda)$ και την στέλνει στη βάση. Η βάση αθροίζει τις επιμέρους τιμές $P(\lambda)$ και όσο αυτή είναι μικρότερη από τη συνολική ισχύ P_T που μπορεί να αποδοθεί επαναλαμβάνεται η παραπάνω λειτουργία με νέα τιμή επιλογής λ που αντιστοιχεί στην αμέσως μικρότερη τιμή λ_{max} από αυτή που μόλις εξετάστηκε. Η λειτουργία του αλγορίθμου φαίνεται παρακάτω σχηματικά



σχήμα 2.2

Στο παραπάνω σχήμα φαίνεται η ταξινόμηση των παραμέτρων λ_{\max} από το σταθμό βάσης και η επιλογή της τιμής λ^1_{\max} σαν τιμή κόστους που εκπέμπεται στους κινητούς σταθμούς (γαλάζιος κύκλος). Επίσης φαίνεται η τιμή ισχύος που ζητά κάθε χρήστης σαν τιμή της συνάρτησης $P(\lambda)$ (πράσινη στήλη). Όπως αναφέραμε παραπάνω αφού επιλέγεται η τιμή λ^1_{\max} μόνο ο χρήστης 1 θα έχει μη μηδενική τιμή ανάθεσης ισχύος καθώς η τιμή λ που επιλέχθηκε είναι μεγαλύτερη από τις τιμές λ_{\max} των υπολοίπων.

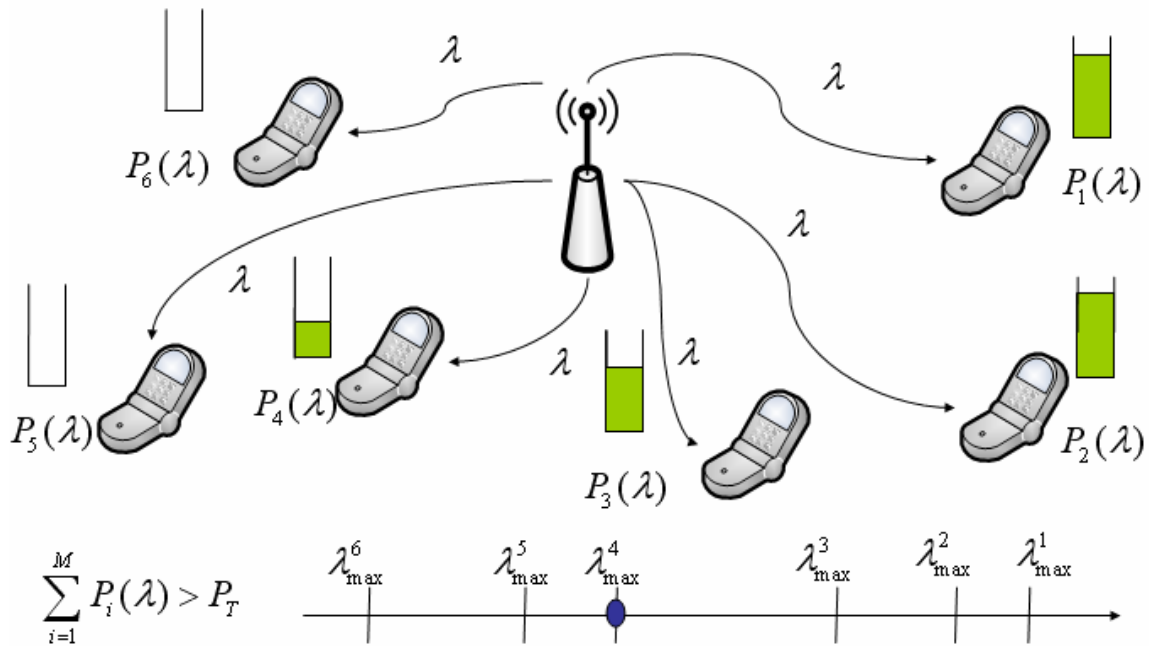
Κάτω αριστερά στην εικόνα φαίνεται η συνθήκη λήξης του αλγορίθμου. Όσο το άθροισμα των ισχύων που απαιτούν οι χρήστες είναι μικρότερο του P_T ο αλγόριθμος συνεχίζει στην επόμενη τιμή.



σχήμα 2.3

Στο παραπάνω σχήμα φαίνεται η επιλογή της αμέσως επόμενης τιμής λ που εισάγει ένα δεύτερο χρήστη στον ανταγωνισμό ισχύος και αυξάνει την ζήτησης ισχύος του χρήστη 1 (αυτό συμβαίνει γιατί η συνάρτηση $P(\lambda)$ είναι φθίνουσα άρα με μείωση της τιμής του λ η συνάρτηση αυξάνει).

Η διαδικασία συνεχίζεται εισάγοντας σε κάθε επιλογή του λ νέους χρήστες μέχρι να μην ικανοποιείται η συνθήκη τερματισμού $\sum_{i=1}^M P_i(\lambda) \geq P_T$. Στο σημείο ικανοποίησης που φαίνεται παρακάτω η πρώτη φάση του αλγορίθμου διακόπτεται και περνάμε στην φάση απόδοσης ισχύος.



σχήμα 2.4

Στην παραπάνω εικόνα φαίνεται ότι με την εισαγωγή του τέταρτου χρήστη στον ανταγωνισμό της ισχύος η συνθήκη τερματισμού ικανοποιείται άρα για τον τελικό ανταγωνισμό μένουν μόνο οι 3 πρώτοι χρήστες και οι υπόλοιποι απομακρύνονται.

Η βελτιστότητα της επιλογής χρηστών που δίνεται από την πρώτη φάση του αλγορίθμου αποδεικνύεται στην συνέχεια

Πρόταση 1: Η επιλογή K κινητών σταθμών που ικανοποιούν τα ακόλουθα κριτήρια αποτελεί τη βέλτιστη λύση του αρχικού προβλήματος βελτιστοποίησης

$$K = \max\{1 \leq j \leq M \mid \sum_{i=1}^j P_i(\lambda_j^{max}) \leq P_T\}. \quad (2.16)$$

Και

$$\sum_{i=1}^K P_i(\lambda_{K+1}^{max}) \geq P_T, \quad K < M \quad (2.17)$$

Η $K=M$

Απόδειξη: Από τα κριτήρια επιλογής έχουμε τις ακόλουθες περιπτώσεις

1. $K=M$. Από αυτό έχουμε ότι $\sum_{i=1}^M P_i(\lambda_{max}^M) \leq P_T$ αφού η συνάρτηση $P(\lambda)$ είναι φθίνουσα ως προς λ μπορούμε να υπολογίσουμε μια τιμή $\lambda^* \geq \lambda_{max}^M$ τέτοια ώστε $\sum_{i=1}^M P_i(\lambda^*) = P_T$ εξασφαλίζοντας την βέλτιστη λύση του αρχικού προβλήματος

2. $K < M$ και $\sum_{i=1}^K P_i(\lambda_{max}^{K+1}) \geq P_T$. Με την ίδια λογική όπως παραπάνω μπορούμε να βρούμε ένα λ^* τέτοιο ώστε $\lambda_{max}^{K+1} < \lambda^* < \lambda_{max}^K$ και ισχύει ότι $\sum_{i=1}^M P_i(\lambda^*) = P_T$

3. $K < M$ και $\sum_{i=1}^K P_i(\lambda_{max}^{K+1}) < P_T$. Τότε από τον ορισμό του K έχουμε ότι $\sum_{i=1}^{K+1} P_i(\lambda_{max}^{K+1}) \geq P_T$ και επιλέγουμε $\lambda^* = \lambda_{max}^{K+1}$. Τότε βλέπουμε ότι για ε οσοδήποτε

μικρό ισχύει ότι $\sum_{i=1}^M P_i(\lambda_{max}^{K+1}) = \sum_{i=1}^{K+1} P_i(\lambda_{max}^{K+1}) < P_T$ και

$\sum_{i=1}^M P_i(\lambda_{max}^{K+1} + \varepsilon) = \sum_{i=1}^K P_i(\lambda_{max}^{K+1} + \varepsilon) > P_T$ άρα δεν μπορεί να βρεθεί βέλτιστη τιμή λ

που κάνει την απόδοση ισχύος ακριβώς ίση με την συνολική ισχύς P_T . Σε αυτή τη περίπτωση η λύση μας είναι κοντά στο βέλτιστο αλλά δε μπορεί να αποδεχθεί η καθολική βελτιστότητα της. Για να βρούμε μία αρκετά καλή προσέγγιση στη βέλτιστη λύση επιλέγουμε τους K πρώτους κινητούς σταθμούς να ανταγωνιστούν την ανάθεση ισχύος και με βάση την ίδια

λογική που περιγράφηκε στις δύο παραπάνω περιπτώσεις μπορούμε να βρούμε ένα $\lambda^* < \lambda_{\max}^{K+1}$ τέτοιο ώστε $\sum_{i=1}^K P_i(\lambda^*) = P_T$

Τα ακριβή βήματα του αλγορίθμου επιλογής φαίνονται παρακάτω

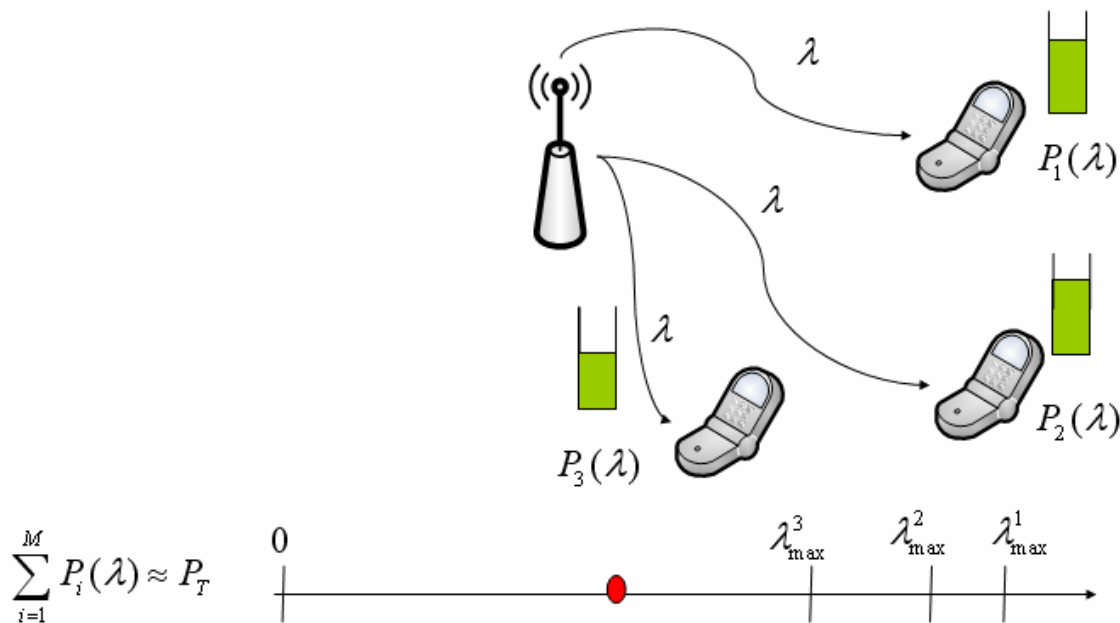
2.2.1 Αλγόριθμος Επιλογής κινητών σταθμών

- η κεραία εκπέμπει τη συνολική τιμή ισχύος προς ανάθεση P_T
- κάθε κινητός σταθμός εκπέμπει την παράμετρο λ_{\max}^i που του αντιστοιχεί
- $k=1$
- ο σταθμός βάσης εκπέμπει την τιμή λ_{\max}^k
- κάθε κινητός σταθμός εκπέμπει την τιμή $P_i(\lambda_{\max}^k)$ που έχει υπολογίσει
- αν $k=1$ και $P_1(\lambda_{\max}^1) = P_T$ τότε επέλεξε το κινητό 1 και σταμάτα, αλλιώς αν $k=1$ και $P_1(\lambda_{\max}^1) < P_T$ πήγαινε στο 9 αλλιώς πήγαινε στο 7
- αν $\sum_{i=1}^{k-1} P_i(\lambda_{\max}^k) > P_T$ επέλεξε τα κινητά από 1 ως $k-1$ και σταμάτα αλλιώς πήγαινε στο 8
- αν $\sum_{i=1}^{k-1} P_i(\lambda_{\max}^k) \leq P_T$ και $\sum_{i=1}^k P_i(\lambda_{\max}^k) > P_T$ τότε επέλεξε τα κινητά από 1 ως $k-1$ και σταμάτα αλλιώς πήγαινε στο 9
- $k=k+1$. αν $k \leq M$ πήγαινε στο 4 αλλιώς επέλεξε τα κινητά από 1 ως $k-1$ και σταμάτα

Μετά την επιλογή των κινητών σταθμών που θα ανταγωνιστούν για την απόδοση ισχύος γίνεται η τελική απόδοση με την εύρεση της τιμής λ στο

διάστημα $[0, \lambda_{\max}^k]$ όπου λ_{\max}^k είναι η μικρότερη τιμή λ_{\max} από όλα τα κινητά που παρέμειναν με σκοπό η συνολική ανάθεση ισχύος να γίνει ίση με P_T . Ο αλγόριθμος εκτελεί μία δυαδική αναζήτηση στο παραπάνω διάστημα με συνθήκη τερματισμού $\sum_{i=1}^k P_i(\lambda) \approx P_T$. Η τιμή $P_i(\lambda)$ κάθε κινητού σταθμού για την τελική τιμή του λ είναι η ανάθεση ισχύος που θα γίνει στο χρήστη i .

Μία σχηματική εκτέλεση του αλγορίθμου φαίνεται στη παρακάτω εικόνα



σχήμα 2.5

Όπου βλέπουμε ότι 3 κινητά ανταγωνίζονται την ανάθεση ισχύος και με αναζήτηση στο πεδίο $[0, \lambda_{\max}^3]$ βρίσκεται η τελική τιμή του λ (κόκκινος κύκλος) που κάνει την συνολική ανάθεση περίπου ίση ή ακριβώς ίση με P_T .

Ο ακριβής αλγόριθμος ανάθεσης φαίνεται παρακάτω

2.3 Αλγόριθμος ανάθεσης ισχύος

Έστω ϵ πολύ μικρή θετική σταθερά

- θέτω $\alpha=0$, $\beta=\lambda_{\max}^k$ και $n=1$
- ο σταθμός βάσης εκπέμπει την τιμή $\lambda = \frac{\alpha + \beta}{2}$ στους επιλεγμένους σταθμούς
- κάθε κινητός σταθμός εκπέμπει την τιμή $P_i(\lambda)$ που έχει υπολογίσει
- αν $|\beta - \alpha| \leq 2\epsilon$ ή $\sum_{i=1}^k P_i(\lambda) = P_T$ σταμάτα αλλιώς πήγαινε στο 5
- αν $\sum_{i=1}^k P_i(\lambda) > P_T$ κάνε $\alpha = \lambda$ αλλιώς κάνε $\beta = \lambda$
- κάνε $n=n+1$ και πήγαινε στο 2

Αν ο παραπάνω αλγόριθμος σταματήσει στην n^* επανάληψη έχουμε ότι $|\lambda^n - \lambda^*| < \epsilon$ όπου λ^* η βέλτιστη λύση του προβλήματος. Άρα μικρότερη τιμή του ϵ θα δίνει πιο ακριβή αποτελέσματα. Μπορεί να αποδειχθεί ότι η σχέση του ελάχιστου αριθμού επαναλήψεων για ένα συγκεκριμένο όριο ϵ δίνεται από τον παρακάτω τύπο

$$n^* = \min\left\{n \geq \frac{\log \lambda_K^{\max} - \log 2\epsilon}{\log 2}, n = 1, 2, \dots\right\}. \quad (2.18)$$

Που δίνει και την πολυπλοκότητα του αλγορίθμου.

2.4 Ασυμπτωτική βελτιστότητα και κατώτερο όριο επίδοσης του αλγορίθμου

Για τον ορισμό της ασυμπτωτικής βελτιστότητας αρχικά μελετάμε τον άνω φράγμα στην βέλτιστη λύση και για αυτό ορίζουμε την παρακάτω ποσότητα

$$U_i^u(P) = \begin{cases} \lambda_i^{max} P, & \text{if } 0 \leq P \leq P_i(\lambda_i^{max}) \\ U_i(\gamma_i(P)), & \text{if } P_i(\lambda_i^{max}) \leq P \leq P_T \end{cases} \quad (2.19)$$

Με βάση αυτό ορίζουμε το πρόβλημα βελτιστοποίησης

$$\begin{aligned} & \max \sum_{i=1}^M U_i^u(P_i) \\ & \text{subject to } \sum_{i=1}^M P_i \leq P_T, \\ & \quad 0 \leq P_i \leq P_T, \quad i = 1, 2, \dots, M. \end{aligned} \quad (2.20)$$

Αφού μπορούμε να δείξουμε ότι για κάθε p $U_i^u(P) \geq U_i(P)$ τότε η λύση του παραπάνω προβλήματος βελτιστοποίησης θα είναι το άνω φράγμα της λύσης του αρχικού μας προβλήματος.

Πρόταση 2: Αν $\bar{P}_p = (P_p^1, P_p^2, \dots, P_p^M)$ είναι η λύση που προτείνει ο παραπάνω αλγόριθμος και $\bar{P}_u = (P_u^1, P_u^2, \dots, P_u^M)$ είναι η βέλτιστη λύση του προβλήματος βελτιστοποίησης της νέας συνάρτησης χρησιμοποίησης $U_i^u(P)$ τότε

$$\sum_{i=1}^M U_i^u(P_i^u) - \sum_{i=1}^M U_i(\gamma_i(P_i^p)) \leq u_{max}, \quad (2.21)$$

$$\text{Όπου } u_{max} = \max_{1 \leq i \leq M} \{U_i(\gamma_i(P_T))\}$$

Απόδειξη: ορίζουμε τα $P^u(\lambda)$ και $\lambda_{i,u}^{max}$ με βάση τη νέα συνάρτηση χρησιμοποίησης ως εξής

$$P_{i,u}(\lambda) = \arg \max_{0 \leq P \leq P_T} \{U_i^u(P) - \lambda P\}$$

$$\lambda_{i,u}^{max} = \min\{\lambda \geq 0 \mid \max_{0 \leq P \leq P_T} \{U_i^u(P) - \lambda P\} = 0\}. \quad (2.22)$$

Από τα παραπάνω μπορούμε εύκολα να αποδείξουμε ότι

$$P_{i,u}(\lambda) = \begin{cases} P_i(\lambda) (> 0), & \text{if } \lambda < \lambda_i^{max} \\ \{P \mid 0 \leq P \leq P_i(\lambda_i^{max})\} (\ni P_i(\lambda)), & \text{if } \lambda = \lambda_i^{max} \\ P_i(\lambda) (= 0), & \text{if } \lambda > \lambda_i^{max} \end{cases} \quad (2.23)$$

Από τα παραπάνω έχουμε ότι $\max_{0 \leq P \leq P_T} \{U^u(P) - \lambda P\} > 0$ για $\lambda < \lambda_{max}$ και $\max_{0 \leq P \leq P_T} \{U^u(P) - \lambda P\} = 0$ για $\lambda \geq \lambda_{max}$ άρα έχουμε ότι $\lambda_{max} = \lambda_{i,u}^{max}$

Έστω ότι ισχύει μία από τις δύο πρώτες υποθέσεις που αναφέρθηκαν στην απόδειξη της πρότασης 1 τότε από την πρόταση 1 μπορούμε να βρούμε ένα λ^*

τέτοιο ώστε $\sum_{i=1}^M P_i(\lambda^*) = P_T$ το οποίο είναι ολικό βέλτιστο για το αρχικό πρόβλημα

που εξετάζουμε. Έτσι αν έχουμε $\bar{P}^u = \bar{P}^p$ και $P_i^p > 0$ τότε $P_i^p > P_i(\lambda_{i,u}^{max})$ άρα από τον ορισμό του U^u έχουμε ότι

$$\sum_{i=1}^M U_i^u(P_i^u) = \sum_{i=1}^M U_i(\gamma_i(P_i^p)). \quad (2.24)$$

Και άρα η πρόταση ισχύει.

Στην περίπτωση που δεν ισχύουν οι δύο πρώτες υποθέσεις και ισχύει η τρίτη σε αυτή τη περίπτωση επιλέγουμε τα K πρώτα κινητά για απόδοση ισχύος άρα $P_i^p = 0$ για $i=K+1, K+2, \dots, M$ και για τα υπόλοιπα επιλέγουμε λ^* τέτοιο ώστε

$$\sum_{i=1}^K P_i(\lambda^*) = P_T. \text{ Από αυτό έχουμε ότι } \lambda^* \leq \lambda_{\max}^{K+1} \text{ και έτσι } P_i^p \geq P_i(\lambda_{\max}^{K+1}). \text{ Άρα έχουμε}$$

$$\sum_{i=1}^M U_i(\gamma_i(P_i^p)) = \sum_{i=1}^K U_i(\gamma_i(P_i^p)) \geq \sum_{i=1}^K U_i(\gamma_i(P_i(\lambda_{\max}^{K+1}))). \quad (2.25)$$

Τώρα ορίζουμε

$$P_i^u = \begin{cases} P_i(\lambda_{K+1}^{max}), & \text{if } i \leq K \\ P_T - \sum_{i=1}^K P_i(\lambda_{K+1}^{max}), & \text{if } i = K + 1 \\ 0, & \text{if } i \geq K + 2 \end{cases} \quad (2.26)$$

Αφού $\sum_{i=1}^M P_i^u = P_T$ τότε αποτελεί βέλτιστη λύση του προβλήματος (U) . Επιπλέον

αφού ισχύει ότι $P_i(\lambda_{\max}^{K+1}) \geq P_i(\lambda_{\max}^i), i=1, 2, \dots, K$ τότε από τον ορισμό της U^u έχουμε ότι

$$U_i^u(P_i^u) = U_i(\gamma_i(P_i(\lambda_{K+1}^{max}))), \quad i = 1, 2, \dots, K.$$

Από τις παραπάνω σχέσεις έχουμε ότι

$$\begin{aligned}
& \sum_{i=1}^M U_i^u(P_i^u) - \sum_{i=1}^M U_i(\gamma_i(P_i^p)) \\
\leq & \sum_{i=1}^K U_i^u(P_i(\lambda_{K+1}^{max})) + U_{K+1}^u(P_T - \sum_{i=1}^K P_i(\lambda_{K+1}^{max})) \\
& - \sum_{i=1}^K U_i(\gamma_i(P_i(\lambda_{K+1}^{max}))) \\
= & U_{K+1}^u(P_T - \sum_{i=1}^K P_i(\lambda_{K+1}^{max})) \\
\leq & U_{K+1}^u(P_T) \\
= & U_{K+1}(\gamma_{K+1}(P_T)) \\
\leq & \max_{1 \leq i \leq M} \{U_i(\gamma_i(P_T))\}.
\end{aligned}
\tag{2.27}$$

Που αποδεικνύει την πρόταση.

Η παραπάνω πρόταση δείχνει ότι η μέγιστη διαφορά απόδοσης της λύσης του αλγορίθμου UPR από αυτήν που δίνεται αν θεωρήσουμε το άνω φράγμα της συνάρτησης χρησιμοποίησης είναι το πολύ η συνάρτηση χρησιμοποίησης ενός κινητού . Επιπλέον αν θεωρήσουμε ότι η συνάρτηση χρησιμοποίησης ενός κινητού είναι φραγμένη μπορούμε να αποδείξουμε την ασυμπτωτική ευστάθεια του αλγορίθμου ως εξής

Αν $\bar{P}_p = (P_p^1, P_p^2, \dots, P_p^M)$ είναι η λύση που προτείνεται από τον αλγόριθμο και $\bar{P}^* = (P_1^*, P_2^*, \dots, P_M^*)$ είναι η βέλτιστη λύση του αρχικού προβλήματος βελτιστοποίησης τότε έχουμε ότι αν ισχύει $M \rightarrow \infty \Rightarrow \sum_{i=1}^M U_i(\gamma_i(P_i^*)) \rightarrow \infty$

Ισχύει και

$$\frac{\sum_{i=1}^M U_i(\gamma_i(P_i^P))}{\sum_{i=1}^M U_i(\gamma_i(P_i^*))} \rightarrow 1 \quad (2.28)$$

Τέλος αναλύεται το κάτω όριο επίδοσης του αλγορίθμου ως εξής

Πρόταση 4: Αν $\bar{P}_p = (P_p^1, P_p^2, \dots, P_p^M)$ είναι η λύση που προτείνεται από τον αλγόριθμο και $\bar{P}^* = (P_1^*, P_2^*, \dots, P_M^*)$ είναι η βέλτιστη λύση του αρχικού προβλήματος βελτιστοποίησης τότε έχουμε

$$\frac{\sum_{i=1}^M U_i(\gamma_i(P_i^P))}{\sum_{i=1}^M U_i(\gamma_i(P_i^*))} \geq \frac{u_{min}}{u_{max} + u_{min}}, \quad (2.29)$$

Όπου $u_{min} = \min_{1 \leq i \leq M} \{U_i(\gamma_i(P_T))\}$

Και $u_{max} = \max_{1 \leq i \leq M} \{U_i(\gamma_i(P_T))\}$.

Απόδειξη: Από την πρόταση 3 έχουμε

$$\frac{\sum_{i=1}^M U_i(\gamma_i(P_i^P))}{\sum_{i=1}^M U_i(\gamma_i(P_i^*))} \geq \frac{\sum_{i=1}^M U_i(\gamma_i(P_i^P))}{u_{max} + \sum_{i=1}^M U_i(\gamma_i(P_i^P))}. \quad (2.30)$$

Και αφού η \bar{P}_p είναι βέλτιστη για τα επιλεγμένα κινητά ισχύει

$$\sum_{i=1}^M U_i(\gamma_i(P_i^P)) \geq U_j(\gamma_j(P_T)) \geq u_{min}, \quad (2.31)$$

Όπου j είναι ένα από τα κινητά που επιλέχθηκαν από τον αλγόριθμο επιλογής κινητών σταθμών. Από τα παραπάνω μπορούμε εύκολα να καταλήξουμε στην πρόταση 4.

2.5 Επέκταση του αλγορίθμου για συνδυασμένη ανάθεση ισχύος και ρυθμού μετάδοσης.

Για να επεκτείνουμε τον παραπάνω αλγόριθμο ούτος ώστε να μπορεί να αποδίδει ισχύς καθώς και ρυθμό μετάδοσης με βέλτιστο τρόπο σε κινητούς σταθμούς πρέπει να αλλάξουμε τη συνάρτηση χρησιμοποίησης. Αρχικά θεωρούμε τις ίδιες συνθήκες ορισμού του προβλήματος όπως παραπάνω δηλαδή μία κυψέλη ενός δικτύου που περιέχει ένα σταθμό βάσης (κεραία) και M κινητούς σταθμούς. Το σύστημα όπως πριν λειτουργεί σε προκαθορισμένες χρονοσχισμές τέτοιες που μπορούμε να υποθέσουμε ότι κατά τη διάρκεια μιας χρονοσχισμής ο θόρυβος, η ποιότητα καναλιού και η διακυβελωτή παρεμβολή κάθε χρήστη είναι σταθερή. Το σύστημα μας μπορεί να εξυπηρετήσει χρήστες που απαιτούν διαφορετικές υπηρεσίες και έχουν διαφορετικούς μέγιστους ρυθμούς μετάδοσης δεδομένων (ετερογενές περιβάλλον) και σκοπός είναι η βέλτιστη απόδοση ισχύος και ρυθμού μετάδοσης σε κάθε κινητό σταθμό από την κεραία κατά την επίλυση του προβλήματος downlink. Η συνάρτηση χρησιμοποίησης που μοντελοποιεί το σύστημα μας είναι η ακόλουθη

$$U_i(R_i, \bar{P}) = R_i f_i(\gamma_i(R_i, \bar{P})), \quad i = 1, 2, \dots, M. \quad (2.31)$$

Όπου R_i ο αποδιδόμενος ρυθμός μετάδοσης στον χρήστη i και P το διάνυσμα αποδιδόμενης ισχύος όπως παραπάνω. Η συνάρτηση f είναι τέτοια που ικανοποιεί τις απαιτήσεις μιας συνάρτησης χρησιμοποίησης όπως ορίστηκαν στη παράγραφο των συναρτήσεων χρησιμοποίησης και μπορεί να είναι κυρτή, κοίλη

ή σιγμοειδής. Όπως βλέπουμε από τον ορισμό της U είναι διαχωρίσιμη ως προς R αλλά όχι ως προς P . Βέβαια ακολουθώντας τη λογική που περιγράφηκε παραπάνω θα γίνει πλήρως διαχωρίσιμη .

Το πρόβλημα βελτιστοποίησης που ορίζεται από την παραπάνω συνάρτηση είναι το ακόλουθο

$$\begin{aligned}
 & \max_{\bar{P}, \bar{R}} \sum_{i=1}^M U_i(R_i, \bar{P}) \\
 & \text{subject to} \quad \sum_{i=1}^M P_i \leq P_T, \\
 & \quad 0 \leq P_i \leq P_T, \quad i = 1, 2, \dots, M, \\
 & \quad 0 \leq R_i \leq R_i^{max}, \quad i = 1, 2, \dots, M,
 \end{aligned} \tag{2.32}$$

Όπου αναζητούμε το βέλτιστο διάνυσμα απόδοσης ρυθμού και απόδοσης ισχύος που μεγιστοποιεί το άθροισμα των συναρτήσεων χρησιμοποίησης με βάση τους περιορισμούς συνολικής αποδιδόμενης ισχύος που θέτει η βάση και μέγιστου ρυθμού μετάδοσης που θέτει κάθε χρήστης. Η βέλτιστη επιλογή των δύο αυτών παραμέτρων οδηγεί τελικά σε μεγιστοποίηση της ρυθμαπόδοσης του δικτύου.

2.5.1 Αναγωγή στο πρόβλημα απόδοσης ισχύος

Η παρακάτω πρόταση μας βοηθάει να ανάγουμε το πρόβλημα απόδοσης ρυθμού μετάδοσης στο πρόβλημα ισχύος με αποτέλεσμα να μπορούμε να εφαρμόσουμε τον αλγόριθμο που περιγράψαμε ως τώρα.

Πρόταση 5: Αν $\sum_{i=1}^M P_i = P_T$ τότε για μία συγκεκριμένη ανάθεση ισχύος η βέλτιστη ανάθεση ρυθμού μετάδοσης δίνεται από τον τύπο

$$R_i^*(P_i) = \begin{cases} \frac{WP_i}{\gamma_i^*(\theta P_T - \theta P_i + A_i)}, & \text{if } P_i \leq \frac{R_i^{max} \gamma_i^*(\theta P_T + A_i)}{W + \theta R_i^{max} \gamma_i^*} \\ R_i^{max}, & \text{otherwise} \end{cases} \quad (2.33)$$

$$\text{Όπου } \gamma_i^* = \arg \max_{\gamma \geq 1} \left\{ \frac{1}{\gamma} f_i(\gamma) \right\}$$

Απόδειξη: Ας υποθέσουμε αρχικά ότι $R_{max} = \infty$ και $N_i^{min} \triangleq W/R_i^{max} = 0$.

Τότε από την πρόταση 1 του [11] έχουμε ότι

$$R_i^*(P_i) = \frac{WP_i}{\gamma_i^*(\theta P_T - \theta P_i + A_i)}, \quad (2.34)$$

$$\text{Όπου } \gamma_i^* = \arg \max_{\gamma \geq 1} \left\{ \frac{1}{\gamma} f_i(\gamma) \right\}$$

Αφού στην περίπτωση μας έχουμε μέγιστο ρυθμό μετάδοσης για κάθε χρήστη η παραπάνω σχέση μετατρέπεται ως εξής

$$R_i^*(P_i) = \begin{cases} \frac{WP_i}{\gamma_i^*(\theta P_T - \theta P_i + A_i)}, & \text{if } \frac{WP_i}{\gamma_i^*(\theta P_T - \theta P_i + A_i)} \leq R_i^{max} \\ R_i^{max}, & \text{otherwise} \end{cases} \quad (2.34)$$

Που είναι ισοδύναμη μετά από πράξεις με

$$R_i^*(P_i) = \begin{cases} \frac{WP_i}{\gamma_i^*(\theta P_T - \theta P_i + A_i)}, & \text{if } P_i \leq \frac{R_i^{max} \gamma_i^*(\theta P_T + A_i)}{W + \theta R_i^{max} \gamma_i^*} \\ R_i^{max}, & \text{otherwise} \end{cases} \quad (2.35)$$

Η παραπάνω πρόταση μας επιτρέπει να αλλάξουμε τον ορισμό της συνάρτησης χρησιμοποίησης ως εξής

$$\begin{aligned}
 U_i^{R_i^*}(P_i) &\triangleq U_i(R_i^*(P_i), P_i) \\
 &= \begin{cases} \frac{W}{\gamma_i^*} \frac{P_i}{\theta P_T - \theta P_i + A_i} f_i(\gamma_i^*), \\ \text{if } P_i \leq \frac{R_i^{max} \gamma_i^* (\theta P_T + A_i)}{W + \theta R_i^{max} \gamma_i^*} \\ \\ R_i^{max} f_i(\gamma_i(R_i^{max}, P_i)), \\ \text{otherwise} \end{cases}, \tag{2.36}
 \end{aligned}$$

Που όπως βλέπουμε εξαρτάται μόνο από την απόδοση ισχύος. Η παραπάνω συνάρτηση χρησιμοποίησης είναι δίκλαδη και πιο συγκεκριμένα είναι κοίλη στο πρώτο μέρος μέχρι το σημείο διαχωρισμού και στην συνέχεια ακολουθεί τη μορφή της f . Με βάση αυτή τη παρατήρηση βλέπουμε ότι η $U_i^{R_i^*}(P_i)$ μπορεί να είναι είτε κοίλη είτε σιγμοειδής (αφού το πρώτο μέρος είναι κοίλο δε μπορεί να είναι κυρτή)

Αντίστοιχα μπορούμε να αλλάξουμε το βασικό πρόβλημα βελτιστοποίησης ως εξής

$$\begin{aligned}
 \max_{\bar{P}} \sum_{i=1}^M U_i^{R_i^*}(P_i) \\
 \text{subject to } \sum_{i=1}^M P_i \leq P_T, \\
 0 \leq P_i \leq P_T, \quad i = 1, 2, \dots, M. \tag{2.37}
 \end{aligned}$$

Που όπως βλέπουμε είναι ανάλογο του προβλήματος που μελετήσαμε στην προηγούμενη ενότητα και πλέον εξαρτάται μόνο από την απόδοση ισχύος στους κινητούς σταθμούς. Η αναγωγή αυτή μας επιτρέπει να χρησιμοποιήσουμε

αυτούσιο τον αλγόριθμο UPR που περιγράψαμε παραπάνω για την επίλυση του συνδυασμένου προβλήματος βελτιστοποίησης της ρυθμαπόδοσης (throughput).

2.5.2 Μελέτη της φυσικής σημασίας του αλγορίθμου για το πρόβλημα ρυθμαπόδοσης

Στην συνέχεια θα αναλυθούν κάποιες χρήσιμες παρατηρήσεις για την φυσική σημασία του αλγορίθμου επίλυσης του παραπάνω προβλήματος.

Αρχικά ορίζουμε την έννοια της αποδοτικότητας. Θεωρούμε ένα κινητό σταθμό i πιο αποδοτικό από έναν σταθμό j αν

$$U_i^{R_i^*}(P) \geq U_j^{R_j^*}(P), \text{ για κάθε } P \in [0, P_T]$$

Σημείωση 1: Αν ένα κινητό i είναι αποδοτικότερο από ένα κινητό j τότε ισχύει ότι

$$\lambda_i^{\max} \geq \lambda_j^{\max}$$

Απόδειξη: Αρχικά θεωρούμε τις ακόλουθες ποσότητες

$$w_i(\lambda) = \max_{0 \leq P \leq P_T} \{U_i^{R_i^*}(P) - \lambda P\}, \quad (2.38)$$

Και

$$w_j(\lambda) = \max_{0 \leq P \leq P_T} \{U_j^{R_j^*}(P) - \lambda P\}. \quad (2.39)$$

Αφού ο κινητός σταθμός i είναι αποδοτικότερος από τον j έχουμε ότι $w_i(\lambda) \geq w_j(\lambda), \forall \lambda$ και από τον ορισμό του λ_i^{\max} έχουμε ότι

$$w_i(\lambda_j^{max}) \geq w_j(\lambda_j^{max}) = 0 \quad (2.40)$$

Και αφού πρέπει να έχουμε $w_i(\lambda_i^{max}) = 0$ και η συνάρτηση w όπως έχει αναφερθεί παραπάνω είναι φθίνουσα ως προς λ πρέπει να ισχύει ότι

$$\lambda_i^{max} \geq \lambda_j^{max} \quad (2.41)$$

Από την παραπάνω σημείωση φαίνεται ότι η αποδοτικότητα ενός κινητού άρα η προτεραιότητα του στον αλγόριθμο επιλογής εκφράζεται από την παράμετρο λ^{max} .

Στην συνέχεια δίνονται τρεις παράμετροι που επηρεάζουν την αποδοτικότητα ενός κινητού

- αν $f_i(\gamma) = f_j(\gamma)$ και $A_i = A_j$ τότε αν $R_i^{max} \geq R_j^{max}$ ισχύει ότι $\lambda_i^{max} \geq \lambda_j^{max}$. Άρα για κινητά με την ίδια συνάρτηση χρησιμοποίησης και το ίδιο κανάλι προτεραιότητα έχει αυτό με το μεγαλύτερο μέγιστο ρυθμό μετάδοσης.
- αν $R_i^{max} = R_j^{max}$ και $A_i = A_j$ τότε αν $f_i(\gamma) \geq f_j(\gamma)$ ισχύει ότι $\lambda_i^{max} \geq \lambda_j^{max}$. Άρα για κινητά με τον ίδιο μέγιστο ρυθμό μετάδοσης και το ίδιο κανάλι, προτεραιότητα έχει αυτό με τη μεγαλύτερη συνάρτηση χρησιμοποίησης
- αν $R_i^{max} = R_j^{max}$ και $f_i(\gamma) = f_j(\gamma)$ τότε αν $A_i \leq A_j$ ισχύει ότι $\lambda_i^{max} \geq \lambda_j^{max}$. Άρα για κινητά με τα ίδια χαρακτηριστικά δίνεται προτεραιότητα σε αυτό με το καλύτερο κανάλι (μικρότερη τιμή παραμέτρου A)

Οι παραπάνω παρατηρήσεις δίνουν τη φυσική ερμηνεία της βελτιστότητας στην επιλογή των κινητών σταθμών καθώς σε κάθε περίπτωση επιλέγεται το κινητό που υπερτερεί είτε λόγω μεγαλύτερου ρυθμού είτε λόγω καλύτερο καναλιού είτε λόγω καλύτερης συνάρτησης χρησιμοποίησης.

2.6 Μελέτη πολιτικής

Παρακάτω θα αναλυθεί ποια είναι η βέλτιστη πολιτική δρομολόγησης (CDMA ή TDMA) για την βελτιστοποίηση της συνολικής ρυθμαπόδοσης του συστήματος

Πρόταση 4:

- Αν ισχύει ότι $P_i(\lambda_2^{\max}) \geq P_T$ όπου $\lambda_1^{\max} > \lambda_2^{\max} > \dots > \lambda_M^{\max}$ τότε η βέλτιστη στρατηγική είναι TDMA. Αυτό ισχύει γιατί ικανοποιείται η περίπτωση 3 της πρότασης 1 που εκφράζει την περίπτωση αδυναμίας εγγύσης της βελτιστοτητας του αλγορίθμου και μπορούμε να αποδείξουμε ότι η βέλτιστη απόδοση ικανοποιείται αν επιλεγεί μόνο ο πρώτος χρήστης με το καλύτερο κανάλι και του ανατεθεί η συνολική ισχύς και ρυθμός μετάδοσης.
- Αν ισχύει ότι $\sum_{i=1}^M P_i(\lambda_M^{\max}) \leq P_T$ όπου $\lambda_1^{\max} > \lambda_2^{\max} > \dots > \lambda_M^{\max}$ τότε η βέλτιστη πολιτική είναι CDMA. Όπως έχει αποδειχθεί παραπάνω αν ισχύει η παραπάνω συνθήκη μπορεί να επιλεγεί μία τιμή παραμέτρου λ τέτοια ώστε να συμμετέχουν όλοι οι χρήστες στον ανταγωνισμό ισχύος και να ισχύει $\sum_{i=1}^M P_i(\lambda^*) = P_T$ άρα να αποτελεί λύση του μετασχηματισμένου προβλήματος βελτιστοποίησης που αναφέρθηκε παραπάνω.

Τέλος θα δοθεί μία συνθήκη που καθορίζει πότε συμφέρει η καθεμία από τις διαφορετικές πολιτικές CDMA ή TDMA.

Πρόταση 6:

Αν ισχύει $R_i^{\max} A_i \geq \frac{P_T W}{\gamma_i^*}$ για κάθε χρήστη τότε βέλτιστη πολιτική είναι η TDMA

Απόδειξη:

Αν ισχύει η παρακάτω πρόταση τότε με απλές πράξεις μπορούμε να δείξουμε ότι ισχύει

$$P_T \leq \frac{r_i^{max} \gamma_i^* (\theta P_T + A_i)}{W + \theta R_i^{max} \gamma_i^*} \quad (2.41)$$

Άρα κάθε συνάρτηση χρησιμοποίησης είναι κύλη συνάρτηση άρα κάθε λ_i^{max} δίνεται από τον τύπο $\lambda_i^{max} = \frac{U_i^{R_i^*}(P_T)}{P_T}$ και η ισχύς που του ανατίθεται είναι $P_i(\lambda_i^{max}) = P_T$. Άρα επιλέγεται ένα μόνο κινητό και η βέλτιστη πολιτική είναι η TDMA.

Στην συνέχεια δίνεται μία πολύ χρήσιμη παρατήρηση για την απόδοση ρυθμών μετάδοσης στους κινητούς σταθμούς.

Σημείωση 2: αν ένα κινητό επιλεγεί από τον αλγόριθμο επιλογής τότε ο ρυθμός μετάδοσης που θα του αποδοθεί δίνεται από τον τύπο

$$R_i^* = \begin{cases} \frac{P_T W}{\gamma_i^* A_i} < R_i^{max}, & \text{if } A_i > \frac{P_T W}{R_i^{max} \gamma_i^*} \\ R_i^{max}, & \text{otherwise} \end{cases} \quad (2.42)$$

Απόδειξη: Έστω ότι για το κινητό i που επιλέχθηκε δόθηκε ισχύς P_i επίσης έστω

ότι $A_i > \frac{P_T W}{R_i^{max} \gamma_i^*}$ άρα ισχύει ότι

$$P_T < \frac{R_i^{\max} \gamma_i^* (\theta P_T + A_i)}{W + \theta R_i^{\max} \gamma_i^*} \quad (2.43)$$

Και από τον ορισμό της συνάρτησης χρησιμοποίησης έχουμε ότι πρέπει να είναι

κοίλη άρα θα ισχύει ότι $\lambda_i^{\max} = \frac{U_i^{R_i^*}(P_T)}{P_T}$ και $P_i(\lambda) = P_T$ για $\lambda \leq \lambda_i^{\max}$. Σε αυτή τη

περίπτωση έχουμε από την πρόταση 5 ότι $R_i^* = \frac{P_T W}{\gamma_i^* A_i}$ και $R_i^* < R^{\max}$.

Στην περίπτωση που $A_i \leq \frac{P_T W}{R_i^{\max} \gamma_i^*}$ έχουμε ότι

$$P_T \geq \frac{R_i^{\max} \gamma_i^* (\theta P_T + A_i)}{W + \theta R_i^{\max} \gamma_i^*}. \quad (2.44)$$

Τότε η U μπορεί να είναι είτε κοίλη είτε σιγμοειδής. Στην περίπτωση που είναι κοίλη η λογική της απόδειξης είναι όπως παραπάνω. Στην περίπτωση που είναι σιγμοειδής το κινητό i θα δέχεται ισχύ P_i ώστε η $U_i^{R_i^*}(P_i)$ να βρίσκεται στην κυρτή περιοχή για να έχουμε αρνητική δεύτερη παράγωγο που απαιτεί η μεγιστοποίηση του $P(\lambda)$. Λόγω αυτού έχουμε ότι για να ανήκει το P στη κυρτή περιοχή πρέπει να είναι μεγαλύτερο από το σημείο διαχωρισμού άρα

$$P_i(\lambda^*) > \frac{R_i^{\max} \gamma_i^* (\theta P_T + A_i)}{W + \theta R_i^{\max} \gamma_i^*} \quad (2.45)$$

Και άρα από την πρόταση 5 $R_i^* = R^{\max}$

Από την πρόταση 5 έχουμε ότι η απόδοση ρυθμού μετάδοσης τελικά δεν εξαρτάται από την απόδοση ισχύος και κάθε κινητό ή θα λαμβάνει το μέγιστο ρυθμό μετάδοσης ή θα του δίνεται η συνολική ισχύς P_T του σταθμού βάσης άρα θα μεταδίδει μόνο αυτό.

Σημείωση 3: από τα κινητά που επιλέχθηκαν στη πρώτη φάση του αλγορίθμου το πολύ σε ένα να αποδοθεί ρυθμός μετάδοσης μικρότερος από το μέγιστο ρυθμό μετάδοσης του

Απόδειξη: Για την απόδειξη αρχικά αποδεικνύουμε το παρακάτω λήμμα

Βοηθητικό λήμμα 1: Για όλα τα κινητά που έχουν επιλεγεί προς μετάδοση ισχύει

$$\frac{\partial U_i^{R_i^*}(P_i)}{\partial P_i} \Big|_{P_i=P_i^*} = \frac{\partial U_j^{R_j^*}(P_j)}{\partial P_j} \Big|_{P_j=P_j^*} \quad (2.46)$$

Όπου P_i^* η βέλτιστη απόδοση ισχύος για το κινητό i .

Απόδειξη: Η απόδειξη θα γίνει με άτοπο. Έστω ότι ισχύει

$$\frac{\partial U_i^{R_i^*}(P_i)}{\partial P_i} \Big|_{P_i=P_i^*} > \frac{\partial U_j^{R_j^*}(P_j)}{\partial P_j} \Big|_{P_j=P_j^*} \quad (2.47)$$

Τότε μπορούμε να βρούμε δP τέτοιο ώστε για κάθε $P_i \in [P_i^*, P_i^* + \delta P], P_j \in [P_j^* - \delta P, P_j^*]$ να ισχύει η ίδια ανισότητα όπως παραπάνω.

Με βάση αυτό έχουμε ότι

$$\int_{P_i^*}^{P_i^* + \delta P} \frac{\partial U_i^{R_i^*}(P_i)}{\partial P_i} dp > \int_{P_j^* - \delta P}^{P_j^*} \frac{\partial U_j^{R_j^*}(P_j)}{\partial P_j} dp. \quad (2.48)$$

Άρα ισχύει ότι

$$\begin{aligned}
U_i^{R_i^*}(P_i^* + \delta P) - U_i^{R_i^*}(P_i^*) &> U_j^{R_j^*}(P_j^*) - U_j^{R_j^*}(P_j^* - \delta P), \\
U_i^{R_i^*}(P_i^* + \delta P) + U_j^{R_j^*}(P_j^* - \delta P) &> U_i^{R_i^*}(P_i^*) + U_j^{R_j^*}(P_j^*),
\end{aligned}
\tag{2.49}$$

Άρα αποδεικνύουμε ότι η P_i^* δεν είναι βέλτιστη αφού μπορούν να αποδοθούν διαφορετικές τιμές ισχύος στα κινητά ij που να αυξάνουν την συνολική συνάρτηση χρησιμοποίησης. Άτοπο.

Στην συνέχεια θα αποδείξουμε ότι κατά την βέλτιστη απόδοση ισχύος P_i^* δεν μπορούμε να έχουμε δύο κινητά των οποίων οι ισχύς να βρίσκονται στο κοίλο κομμάτι της συνάρτησης χρησιμοποίησης. Αυτό αποδεικνύεται γιατί από το βοηθητικό λήμμα έχουμε ότι

$$\frac{\partial U_i^{R_i^*}(P_i)}{\partial P_i} \Big|_{P_i=P_i^*} = \frac{\partial U_j^{R_j^*}(P_j)}{\partial P_j} \Big|_{P_j=P_j^*}
\tag{2.50}$$

Και αφού και τα δύο κινητά βρίσκονται στη κοίλη περιοχή μπορούμε να βρούμε ένα δP τέτοιο ώστε $U_i^{R_i^*}(P_i) \Big|_{P_i=P_i^*+\delta P}$ και $U_j^{R_j^*}(P_j) \Big|_{P_j=P_j^*-\delta P}$ να συνεχίζουν να βρίσκονται στη κοίλη περιοχή. Από αυτό έχουμε ότι

$$\int_{P_i^*}^{P_i^*+\delta P} \frac{\partial U_i^{R_i^*}(P_i)}{\partial P_i} dp > \int_{P_j^*-\delta P}^{P_j^*} \frac{\partial U_j^{R_j^*}(P_j)}{\partial P_j} dp
\tag{2.51}$$

Αφού οι παράγωγοι της U στην κοίλη περιοχή είναι αύξουσες συναρτήσεις. Αυτό είναι άτοπο με βάση τη λογική απόδειξης του βοηθητικού λήμματος 1.

Άρα το πολύ ένα κινητό από τα επιλεχθέντα μπορεί να δεχθεί ρυθμό μετάδοσης μικρότερο από το μέγιστο μόνο αν αυτό δεχθεί ισχύ που να βρίσκεται στο κοίλο μέρος της συνάρτησης χρησιμοποίησης του.

3 Βασικοί στόχοι

Μια ουσιαστική επέκταση στο μοντέλο UPR [11] που παρουσιάστηκε στο προηγούμενο κεφάλαιο θα ήταν η αναβάθμισή του έτσι ώστε να καλύπτει υπηρεσίες χρηστών πραγματικού χρόνου (Real Time). Ο μόνος τρόπος για να επιτευχθεί αυτός ο στόχος είναι να ενσωματωθεί στην συνάρτηση ευχαρίστησης κάθε χρήση η πληροφορία που αντιπροσωπεύει την ικανοποίησή του σε σχέση με την υπηρεσία που χρησιμοποιεί [12].

Αν και το υπάρχον μοντέλο λύνει με τον καλύτερο δυνατό τρόπο το πρόβλημα της μεγιστοποίησης της ευχαρίστησης των χρηστών ενός συστήματος (εκφρασμένη μέσα από την συνολική ρυθμαπόδοση), δεν λαμβάνει σε καμία περίπτωση υπόψη την υπηρεσία που χρησιμοποιεί ο εκάστοτε χρήστης. Συγκεκριμένα, στην περίπτωση που ο χρήστης θέλει να χρησιμοποιήσει μια υπηρεσία πραγματικού χρόνου (Real Time) όπως μια τηλεφωνική κλήση ή μια βιντεοδιάσκεψη το σύστημα δεν μπορεί να του εξασφαλίσει την ποιότητα της υπηρεσίας του. Αυτό κυρίως οφείλεται στο γεγονός ότι υπάρχουσες συναρτήσεις ευχαρίστησης δεν εμπεριέχουν σε καμία περίπτωση πληροφορία σχετικά με την ικανοποίηση του χρήστη σε σχέση με την υπηρεσία που χρησιμοποιεί, γιατί έχουν σχεδιαστεί με το σκεπτικό να χρησιμοποιηθούν από έναν χρονοδρομολογιτή που είναι στα κατώτερα επίπεδα της στοίβας TCP/IP. Η περιορισμένη χωρητικότητα ενός ασύρματου δικτύου, κάνει επιτακτική την ανάγκη να εξελιχθούν οι χρονοδρομολογητές του επιπέδου μετάδοσης έτσι ώστε να λαμβάνουν υπόψη τους πληροφορίες του επιπέδου υπηρεσίας.

Στην περίπτωση των χρηστών πραγματικού χρόνου, η ποιότητα της υπηρεσίας που παρέχεται από το δίκτυο μπορεί να εξασφαλιστεί μέσα από τους

βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης. Όπως θα αναλυθεί και στην συνέχεια, μέσα από τέτοιου είδους περιορισμούς, κάθε υπηρεσία θέτει ένα κατώφλι στην ρυθμαπόδοσης κάτω από το οποίο η υπηρεσία δεν είναι αποδεκτή. Μια τροποποίηση των συναρτήσεων χρησιμοποίησης που θα οδηγούσε στην έκφραση των περιορισμών αυτών, θα έδινε στον χρονοδρομολογική την δυνατότητα να εξασφαλίσει την επιτυχή συνύπαρξη ετερογενών υπηρεσιών μέσα στο ίδιο σύστημα.

Επιλέξαμε να επεκτείνουμε το σύστημα που βασίζεται σε συναρτήσεις χρησιμοποίησης του Lee [11] λόγω των πλεονεκτημάτων αυτής της λύσης όπως αυτά έχουν αναλυθεί στο προηγούμενο κεφάλαιο. Μια «καθαρή» επέκταση του συστήματος αυτού θα πρέπει να χαρακτηρίζεται από δύο βασικά χαρακτηριστικά. Πρώτον, πρέπει να εξασφαλίζεται ότι οι επεκτάσεις επηρεάζουν όσο το δυνατόν λιγότερο την βέλτιστη λύση στις ανάγκες του αρχικού συστήματος. Δεν θα είχε νόημα να χρησιμοποιήσουμε ένα σύστημα που αποδεδειγμένα πετυχαίνει τους στόχους του, και το τροποποιήσουμε με τέτοιο τρόπο ώστε να μην ικανοποιεί ένα υπερσύνολο αυτών των στόχων. Κατά συνέπεια, πρέπει οι επεκτάσεις αυτές να είναι τόσο δυναμικές και ευέλικτες έτσι ώστε να μην επεμβαίνουν καταστροφικά στην διαδικασία της λύσης. Δεύτερον, πρέπει οι τροποποιήσεις που θα λάβουν χώρα στην διατύπωση του προβλήματος και στην διαδικασία επίλυσής του να μπορούν εύκολα να αρθούν έτσι ώστε να υπάρχει συμβατότητα με χρήστες μη πραγματικού χρόνου (Non Real Time).

Κατά συνέπεια, ο στόχος αυτής της εργασίας είναι διττός. Πρέπει αφενός να επεκταθεί το πρόβλημα και η λύση του, ώστε να καλύπτει τις ανάγκες των νέων υπηρεσιών. Αφαιρετού η νέα λύση πρέπει να σέβεται τους υπόλοιπους χρήστες και την βελτιστοποίηση της απόδοσης του δικτύου όπως αυτή εξασφαλιζόταν παλιότερα.

3.1 Βραχυπρόθεσμοι περιορισμοί ρυθμαπόδοσης (Sort-term throughput constraints)

Τα προηγούμενα χρόνια η έρευνα έχει αποδώσει χρονοδρομολογητές που κατανέμουν τους πόρους του συστήματος στους χρήστες με δικαιοσύνη και με σεβασμό στις απαιτήσεις τους. Το πόσο επιτυχημένος είναι ένας χρονοδρομολογητής προφανώς εξαρτάται από το πόσο επιτυχημένα έχουν εκφραστεί οι απαιτήσεις των χρηστών. Μέχρι στιγμής, όπως έχει αναφερθεί και στην εισαγωγή, μια υπηρεσία πραγματικού χρόνου έχει εκφραστεί είτε μέσα από περιορισμούς στην καθυστέρηση ανάμεσα σε διαδοχικές προσβάσεις στο δίκτυο (delay constraints) είτε μέσα από πιθανοτικούς ή αυστηρούς περιορισμούς πρόσβασης στο δίκτυο σε μη επικαλυπτόμενα παράθυρα παρακολούθησης.

Είναι κοινώς αποδεκτό ότι οι υπηρεσίες πραγματικού χρόνου είναι υπηρεσίες ευαίσθητες στην καθυστέρηση. Μέσα από τους περιορισμούς καθυστέρησης εξασφαλίστηκε η πρόσβαση ενός χρήστη στο δίκτυο ανά τακτά χρονικά διαστήματα. Δεν εξασφαλίστηκε όμως αν η πρόσβαση αυτή στο δίκτυο θα γίνει υπό συνθήκες που θα επιτρέψουν την μετάδοση δεδομένων. Ο χρονοδρομολογητής δεν χρησιμοποιεί με ικανοποιητικό τρόπο την πληροφορία του καναλιού του χρήστη και κατά συνέπεια χάνει τον ευκαιριακό χαρακτήρα του.

Παρόμοια μειονεκτήματα έχουν οι περιορισμοί πρόσβασης στο δίκτυο. Είτε φράσσοντας από πάνω την πιθανότητα ένας χρήστης να μην αποκτήσει πρόσβαση στο δίκτυο ανάλογη με το βάρος του μέσα σε μη επικαλυπτόμενα παράθυρα παρατήρησης είτε έχοντας αυστηρούς περιορισμούς που επιβάλλουν την ελάχιστη πρόσβαση στο δίκτυο, δεν μπορούμε να εξασφαλίσουμε την

μετάδοση δεδομένων. Κατά συνέπεια, είναι πιθανόν, παρόλο που εξασφαλίζονται οι περιορισμοί των υπηρεσιών πραγματικού χρόνου, η ποιότητα της υπηρεσίας να μην είναι αποδεκτή.

Σε αυτή την ενότητα μελετάμε την και προσδιορίζουμε τις επιθυμητές ιδιότητες και τις απαιτήσεις QoS που χαρακτηρίζουν την υπηρεσία ενός χρήστη πραγματικού χρόνου έτσι ώστε να εξασφαλίζεται η ικανοποίησή του. Οι απαιτήσεις ενός χρήστη υπηρεσίας πραγματικού χρόνου είναι άρρηκτα συνδεδεμένες με σταθερούς ρυθμούς μετάδοσης και με βραχυπρόθεσμη εγγύηση ρυθμαπόδοσης. Για να πετύχουμε τους παραπάνω στόχους:

- i. χωρίζουμε τον χρόνο σε διαδοχικά επικαλυπτόμενα παράθυρα,
- ii. προκαθορίζουμε ένα κατώφλι εκφρασμένο σε ποσότητα δεδομένων
- iii. μελετούμε την πιθανότητα ο χρήστης να λάβει υπηρεσία χειρότερη από το παραπάνω κατώφλι.

Οι παραπάνω απαιτήσεις εκφράζονται μέσα από τον ακόλουθο τύπο

$$\Pr[\hat{\beta}_{RT,i}(t) \leq B_{RT,i}]_{W_i} \quad \forall t \text{ (slot)} \quad \forall i \in S_{RT} \quad (3.1)$$

όπου, S_{RT} αποτελεί το σύνολο των χρηστών υπηρεσίας πραγματικού χρόνου,

το W_i αποτελεί το παράθυρο παρατήρησης του χρήστη i ,

το $B_{RT,i}$ αποτελεί το προκαθορισμένο κατώφλι δεδομένων που πρέπει να λάβει ο χρήστης μέσα στο παράθυρο W_i ,

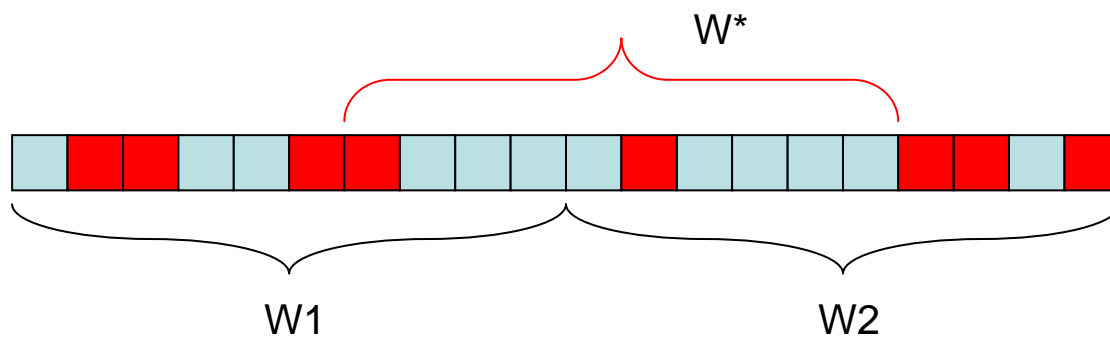
το $\hat{\beta}_{RT,i}(t)$ αποτελεί την ποσότητα των δεδομένων που έχει λάβει ο χρήστης i στο παρελθόν κατά το διάστημα $[t-W_i+1, t]$.

Όσο μικρότερη είναι η πιθανότητα, που εξασφαλίζεται από το σύστημα, ο χρήστης να λάβει λιγότερα δεδομένα από το προκαθορισμένο του κατώφλι, τόσο καλύτερη είναι η ποιότητα της υπηρεσίας που το παρέχεται, και το αντίστροφο.

Γνωρίζοντας τον ρυθμό με τον οποίο ο χρήστης μετέδωσε σε κάθε χρονοσχισμή του παρελθόντος $R_{RT,i}$, αλλά και την διάρκεια μιας χρονοσχισμής t_s μπορούμε να εκφράσουμε την ποσότητα δεδομένων που έχει λάβει ο χρήστης i την μέσα στην προηγούμενη χρονοσχισμή ως

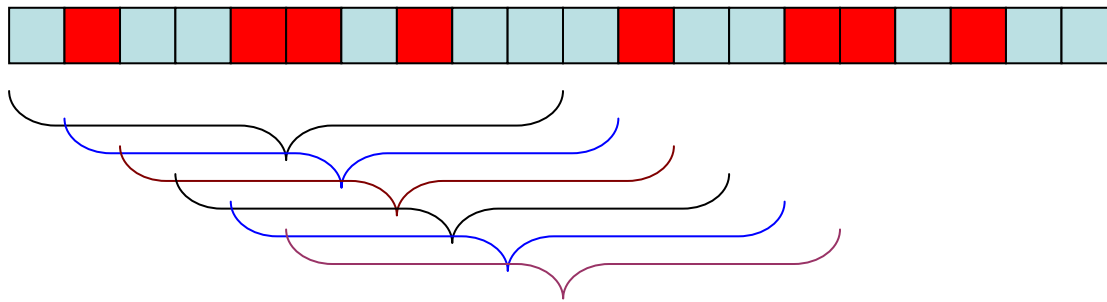
$$B_{RT,i} = R_{RT,i} \cdot W_i \cdot t_s \quad (3.2)$$

Ο λόγος που χρησιμοποιούμε επικαλυπτόμενα παράθυρα χρησιμοποίησης, είναι επειδή μόνο μέσα από αυτά μπορούμε να παρατηρήσουμε τον στιγμιαίο ρυθμό μετάδοσης των χρηστών. Με μη επικαλυπτόμενα παράθυρα παρατήρησης μπορούμε μεν αν πετύχουμε τον ίδιο μέσο ρυθμό μετάδοσης, αλλά υπάρχουν περιπτώσεις στιγμιαία να μην εξασφαλίζεται η υπηρεσία.



Σχήμα 3.1

Ας υποθέσουμε ότι κάθε φορά που εξυπηρετείται ένας χρήστης τότε λαμβάνει την μέγιστη δυνατή ρυθμαπόδοση x (εκφρασμένη σε δεδομένα) και ότι το προκαθορισμένο κατώφλι που θέτει η υπηρεσία του είναι $4x$. Όπως φαίνεται στο σχήμα 3.1 τα μη επικαλυπτόμενα παράθυρα $W1$ και $W2$ καλύπτουν το κριτήριο της υπηρεσίας, σε αντίθεση με το W^* όπου η υπηρεσία αποτυγχάνει.



Σχήμα 3.2

Αντίθετα, με επικαλυπτόμενα παράθυρα, δεν απαιτούμε περισσότερα δεδομένα από το σύστημα, απλά εξασφαλίζουμε ότι τα δεδομένα που θα λάβουμε, κατανέμονται στον χρόνο με τέτοιο τρόπο ώστε να εξασφαλίζεται πάντα η υπηρεσία του χρήστη, πετυχαίνοντας εν τέλει την ίδια ρυθμαπόδοση.

Χρησιμοποιώντας την πληροφορία από δίδεται από τον τύπο (3.2) ένας χρονοδρομολογητής μπορεί να εντοπίσει:

- 1) τους χρήστες που παίρνουν σε κάθε παράθυρο περισσότερα δεδομένα από το κατώφλι που έχουν θέσει και κατά συνέπεια μεγαλύτερη από την επιθυμητή ρυθμαπόδοση.
- 2) Τους χρήστες που έχουν παράθυρα που δεν ικανοποιούν τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης.

Ένας «έξυπνος» χρονοδρομολογητής θα μπορούσε, λοιπόν να μειώσει την ρυθμαπόδοση των χρηστών της πρώτης περίπτωσης προς όφελος των χρηστών της δεύτερης περίπτωσης αλλά και των χρηστών υπηρεσιών μη πραγματικού χρόνου, αυξάνοντας ταυτόχρονα και την χωρητικότητα του συστήματος.

3.2 Οι προδιαγραφές των νέων συναρτήσεων χρησιμοποίησης

Ο μόνος τρόπος να επιτευχθούν οι παραπάνω στόχοι είναι η τροποποίηση των συναρτήσεων χρησιμοποίησης. Αυτό οφείλεται στο γεγονός ότι η ευχαρίστηση των χρηστών (σε επίπεδο χρονοσχισμής) πλέον δεν εξαρτάται μόνο από τον ρυθμό και την ισχύ με τις οποίες θα επιτευχθεί η μετάδοση του αλλά και με την επίτευξη των πραγματικού χρόνου (Real time) αναγκών του. Οι επεκτάσεις που θα προτείνονται κινούνται σε 3 άξονες. Πρώτον, πρέπει να συμπληρωθεί η εκφραστικότητα των συναρτήσεων χρησιμοποίησης έτσι ώστε αν εμπεριέχουν πληροφορία σχετικά με τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης. Δεύτερον, πρέπει να διατηρήσουν την εκφραστικότητά τους ως προς το περιβάλλον επικοινωνίας τους χρήστη και κατά συνέπεια το κανάλι του, έτσι ώστε ο χρονοδρομολογητής που θα προκύψει να διατηρεί τα ευκαιριακά χαρακτηριστικά του (opportunistic scheduling). Τρίτον η νέα μορφή των συναρτήσεων χρησιμοποίησης πρέπει να είναι άρρηκτα συνδεδεμένη με την μαθηματική επίλυση του προβλήματος και τον αλγόριθμο UPR.

Με βάση τα παραπάνω, οι νέες συναρτήσεις χρησιμοποίησης πρέπει να εμπεριέχουν πληροφορία σχετικά με το κριτήριο,

$$\Pr[\hat{\beta}_{RT,i}(t) \leq B_{RT,i}]_{W_i} \quad \forall t \text{ (slot)} \quad \forall i \in S_{RT}$$

όπως αυτό παρουσιάζεται στην προηγούμενη παράγραφο. Η πληροφορία αυτή πρέπει να χρησιμοποιηθεί με τέτοιο τρόπο ώστε αφενός να εκφράζεται η δυσαρέσκεια του χρήστη όταν η υπηρεσία που του παρέχεται τείνει να υποβαθμιστεί ή είναι ήδη υποβαθμισμένη, αφετέρου να εκφράζεται η ευχαρίστηση του χρήστη στην αντίθετη περίπτωση. Για να συνδυαστεί η

παραπάνω λογική με την ισχύ και την συχνότητα υπό τις οποίες θα πραγματοποιηθεί η μετάδοση, που είναι βασικά χαρακτηριστικά των παλαιών συναρτήσεων χρησιμοποίησης, μια επιθυμητή συμπεριφορά των νέων συναρτήσεων χρησιμοποίησης θα μπορούσε να εκφραστεί από την ακόλουθη πρόταση:

«όταν η υπηρεσία του χρήστη υποβαθμίζεται πρέπει η ικανοποίηση αυτού να αυξάνεται έστω και με μικρότερα επίπεδα ισχύος και ρυθμού μετάδοσης – όταν η υπηρεσία του χρήστη εκπληρώνεται τότε η επιπλέον ευχαρίστησή του εμπεριέχει μόνο μεγάλα επίπεδα ισχύος και ρυθμού μετάδοσης»

Κατά συνέπεια ένας χρονοδρομολογητής που βασίζεται στις νέες συναρτήσεις χρησιμοποίησης θα μπορεί να μειώσει τα μέγιστα επίπεδα ρυθμαπόδοσης του συστήματος για να εξυπηρετήσει ένα χρήστη στον οποίο παρέχεται κακής ποιότητας υπηρεσία πραγματικού χρόνου (που παραβιάζει το κάτω φράγμα των απαιτήσεών του). Αντίθετα θα μπορεί να χρησιμοποιήσει προς κοινό όφελος του συστήματος τις παραχωρήσεις που κάνει ένας χρήστης με καλής ποιότητας υπηρεσία πραγματικού χρόνου (όταν υπερβαίνεται το άνω φράγμα των απαιτήσεών του).

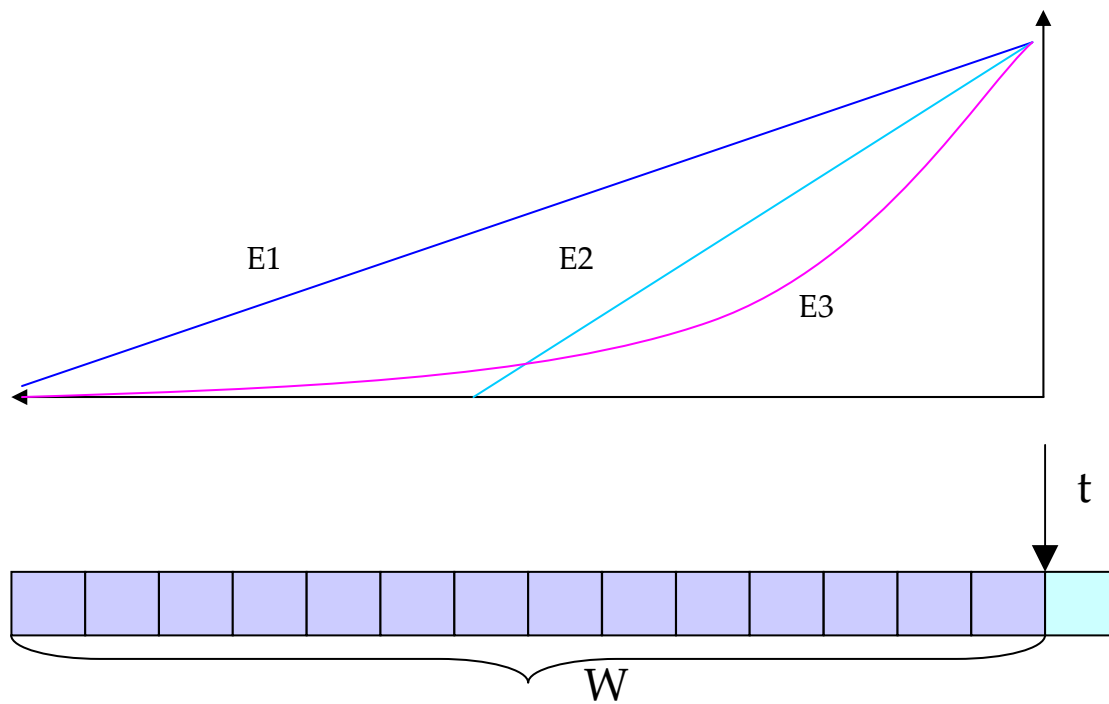
Οι όποιες επεκτάσεις χρησιμοποιηθούν, θα πρέπει να γίνουν με σεβασμό ως προς την εκφραστικότητα των συναρτήσεων χρησιμοποίησης σε σχέση με το κανάλι του χρήστη. Μια τέτοια προδιαγραφή θα έδινε στον χρονοδρομολογητή τα απαιτούμενα ευκαιριακά χαρακτηριστικά του. Κατά συνέπεια ανά χρονοσχισμή θα πρέπει να συνυπάρχουν δύο παράγοντες που να ανταγωνίζονται ή να συμμαχούν έτσι ώστε να επηρεάζουν την απόφαση χρονοδρομολόγησης, το επίπεδο παροχής της βραχυπρόθεσμης υπηρεσίας και το κανάλι του χρήστη.

Τέλος όσες τροποποιήσεις λάβουν χώρα, δεν πρέπει σε καμία περίπτωση να επηρεάζουν τα χαρακτηριστικά των συναρτήσεων χρησιμοποίησης μέσω των

οποίων ο αλγόριθμος UPR βρίσκει βέλτιστη λύση. Σε αυτά τα χαρακτηριστικά εμπεριέχονται το πεδίο τιμών, η συνέχειά τους και η μονοτονία τους.

3.3 Η πληροφορία που θα εκμεταλλευτούμε

Όπως έχει γίνει προφανές μέχρι στιγμής, η ποιότητα των υπηρεσιών πραγματικού χρόνου είναι άρρηκτα συνδεδεμένη με την εξασφάλιση των βραχυπρόθεσμων απαιτήσεων ανά μικρά χρονικά διαστήματα. Η πληροφορία που πρέπει ένας χρονοδρομολογιτή να χρησιμοποιήσει για να δώσει μια λύση η οποία σέβεται αυτές τις απαιτήσεις, είναι το σύνολο των δεδομένων που έχει πάρει ο εκάστοτε χρήστης υπηρεσίας πραγματικού χρόνου στο άμεσο παρελθόν. Για να γίνει αυτό εφικτό πρέπει να αποθηκεύεται η ποσότητα των δεδομένων που λαμβάνει κάθε χρήστης την στιγμή που εξυπηρετείται. Με αυτόν τον τρόπο, κάθε χρήστης υπηρεσιών πραγματικού χρόνου συνοδεύεται από ένα παράθυρο (ίδια διάστασης με το παράθυρο που προκύπτει από τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης την υπηρεσίας που χρησιμοποιεί.) όπου κάθε χρονοσχισμή του παρελθόντος εμπεριέχει την ποσότητα των δεδομένων που μεταδόθηκε.



Σχήμα 3.3

Όπως φαίνεται στο σχήμα 3.3, πριν ο χρονοδρομολογητής πάρει μια απόφαση σχετικά με την κατανομή της ισχύος στην επόμενη χρονοσχισμή (γαλάζιο κουτί), μπορεί να γνωρίζει τον όγκο της πληροφορίας που έχει μεταδοθεί στις W προηγούμενες χρονοσχισμές (μοβ κουτιά).

Η πληροφορία αυτή μπορεί να χρησιμοποιηθεί με δύο τρόπους:

1. Αθροίζοντας τον όγκο των δεδομένων που μεταδόθηκαν στις προηγούμενες χρονοσχισμές μπορούμε να συμπεράνουμε αν ο χρήστης ικανοποιεί τους περιορισμούς της υπηρεσίας του στο τρέχον παράθυρο και με βάση αυτό το δεδομένο να ρυθμίσουμε την προτεραιότητά του.
2. Επιπλέον μπορούμε να αποκτήσουμε αίσθηση την κατάστασης της υπηρεσίας του χρήστη χρησιμοποιώντας την ποσότητα των δεδομένων που έχει λάβει σε κάθε χρονοσχισμή. Είναι προφανές ότι η πληροφορία που έχει πάρει σε κάθε χρονοσχισμή έχει διαφορετική ισχύ και κατά συνέπεια διαφορετικό αντίκτυπο στην ποιότητα της υπηρεσίας στο άμεσο μέλλον. Αυτό συμβαίνει γιατί η i -οστή χρονοσχισμή πριν την τρέχουσα,

πρόκειται να συμπεριληφθεί στον καθορισμό της επιτυχίας ή της αποτυχίας της υπηρεσίας στα επόμενα W-i παράθυρα. Κατά συνέπεια ανάμεσα σε δύο χρήστες με επιτυχημένα παράθυρα, σε ευνοϊκότερη θέση βρίσκεται αυτός που έχει πάρει μεγαλύτερο όγκο δεδομένων στις πλησιέστερες προς την τρέχουσα χρονοσχισμές.

Όπως φαίνεται από τις εξισώσεις του παραπάνω σχήματος, υπάρχουν πολλοί τρόποι για να αποδοθεί ισχύς ή η «σημαντικότητα» της κάθε χρονοσχισμής. Ανάλογα με το αν θα χρησιμοποιήσουμε γραμμική (E1) ή εκθετική (E3) συνάρτηση καθορισμού της ισχύος κάθε χρονοσχισμής και με το αν θα χρησιμοποιήσουμε την πληροφορία όλου του παραθύρου ή μέρος αυτού μπορούμε να ρυθμίσουμε το πόσο δυναμικά θα μεταβάλλεται η αίσθηση της εύνοιας ή της αποτυχίας κάθε χρήστη σε σχέση με τον χρόνο.

3.4 Προτεινόμενες τροποποιήσεις στις συναρτήσεις χρησιμοποίησης

Έχοντας προδιαγράψει τις απαιτήσεις που πρέπει να πληρούνται και έχοντας παρουσιάσει την πληροφορία που πρέπει να αποτυπωθεί μέσα από αυτές, είμαστε σε θέση να παρουσιάσουμε τις αλλαγές που προτείνουμε πάνω στις συναρτήσεις χρησιμοποίησης.

Βασισμένοι στους τύπους (3.1),(3.2),(3.3), μπορούμε να δώσουμε την πλήρη ανάπτυξη της συνάρτησης χρησιμοποίησης ενός χρήστη με σιγμοειδή (sigmoid) απεικόνιση ρυθμού σφάλματος ως προς την ισχύ μετάδοσης (bit error rate).

$$U_i^*(R_i, P) = R_i c_i \left\{ \frac{1}{1 + e^{-a_i \left(\left(\frac{W}{R_i} \frac{P_i}{\theta \sum_{j=1}^{N(t)} P_j - \theta P_i + (A_i)} \right)^{-(b_i)} \right)}} - d_i \right\} \quad (3.3)$$

Όπως φαίνεται στον παραπάνω τύπο, η συνάρτηση χρησιμοποίησης ενός χρήστη είναι μια συνάρτηση της ισχύος και του ρυθμού μετάδοσης και εξαρτάται από τις παραμέτρους της συνάρτησης ρυθμού σφάλματος a_i , b_i , c_i , d_i , και του περιβάλλοντος επικοινωνίας A_i , όπου $A_i \triangleq I_i/G_i$. Πριν από κάθε χρονοσχιμή ο παράγοντας που θα παίξει τον σημαντικότερο ρόλο στην απόφαση του χρονοδρομολογητή είναι το περιβάλλον επικοινωνίας του κάθε χρήστη και κατά συνέπεια η ποιότητα του καναλιού του A_i .

Μελετώντας αυτό το πολυπαραγοντικό πρόβλημα, καταλήξαμε στο να προσδώσουμε στην παράμετρο b_i τα χαρακτηριστικά που θα προσδιορίζουν τις υπηρεσίες πραγματικού χρόνου. Ο λόγος που καταλήγαμε σε αυτή την απόφαση περιγράφεται με διαισθητικό τρόπο αρχικά και αποδεικνύεται με μαθηματική ανάλυση στην συνέχεια της εργασίας.

Όπως αναλύθηκε στην στο κεφάλαιο 2 αν αυξήσουμε την παράμετρο b_i τότε η συνάρτηση ρυθμού σφάλματος και κατά συνέπεια και η συνάρτηση χρησιμοποίησης μεταφέρονται προς τα δεξιά. Στην ουσία η συνάρτηση χρησιμοποίησης μικραίνει για οποιοδήποτε P από την αρχική της μορφή διατηρώντας όμως την ίδια μορφή. Όπως θα αποδειχθεί στην συνέχεια αυτή η μετακίνησης της συνάρτησης χρησιμοποίησης οδηγεί στην μείωση του λ_{MAX} που προκύπτει από την ανάλυση του αλγόριθμου UPR και κατά συνέπεια στην μείωση της προτεραιότητας του χρήστη κατά την εκτέλεση του αλγόριθμου MSA (Mobile Selection Algorithm). Αντίθετα η μείωση του b_i οδηγεί στην μετακίνηση

της συνάρτησης χρησιμοποίησης προς τα αριστερά και κατά συνέπεια στην αύξηση του λ_{\max} αλλά και τις προτεραιότητας του χρήστη στον MSA.

Θεωρώντας, λοιπόν, το λ_{\max} κάθε χρήστη την «μέγιστη τιμή» που είναι διατεθειμένος να «πληρώσει» για την υπηρεσία που θα λάβει, αυξάνοντας το b_i μειώνουμε την προθυμία του χρήστη να πληρώσει για την υπηρεσία που θα λάβει, κατά συνέπεια τον κάνουμε λιγότερο ανταγωνιστικό ανάμεσα στους χρήστες. Αντίθετα μειώνοντας το b_i αυξάνεται η προθυμία του χρήστη να πληρώσει για την υπηρεσία που θα λάβει κατά συνέπεια αυξάνουμε την πιθανότητα να εξυπηρετηθεί.

Κατά συνέπεια κάθε κινητός χρήστης που χρησιμοποιεί υπηρεσία πραγματικού χρόνου θα έχει μια συνάρτηση χρησιμοποίησης που θα εκφράζει τον βαθμό ικανοποίησής του σε σχέση με την αναμενόμενη ισχύ μετάδοσης που μπορεί να πετύχει, η οποία μπορεί πλέον να εκφραστεί όπως φαίνεται στον ακόλουθο τύπο:

$$U_i^{R_i^*}(P_i, a_i, \hat{b}_{RT,i}(t)) \triangleq U_i^{R_i^*}(P_i, \hat{b}_{RT,i}(t)) \quad \forall i \in S_{RT} \quad (3.4)$$

όπου f_i είναι η συνάρτηση της πιθανότητας μιας επιτυχούς μετάδοσης ενός πακέτου ενός χρήστη i και είναι μια αύξουσα συνάρτηση της ενέργειας ανά bit προς τον ρυθμό παρεμβολής γ_i σε κάθε χρονοσχισμή, S_{RT} είναι το σύνολο των χρηστών που απαιτούν υπηρεσίες πραγματικού χρόνου και $\hat{b}_{RT,i}(t)$ είναι η παράμετρος b ενός χρήστη υπηρεσίας πραγματικού χρόνου (RT) την χρονοσχισμή t .

Πρόταση 1: αν $A_i = A_j$, $R_i^{\max} = R_j^{\max}$ και $b_i < b_j$ τότε $\lambda_i^{\max} > \lambda_j^{\max}$.

Απόδειξη: [12]

Η παραπάνω πρόταση αποδεικνύει ότι υπό τις ίδιες συνθήκες, ένας χρήστης i με μικρότερη παράμετρο b_i έχει μεγαλύτερη προτεραιότητα στο να επιλεγεί από έναν χρήστη j με μεγαλύτερη παράμετρο b_j . Επιπλέον, αν παρατηρούνται, σε βραχυπρόθεσμο επίπεδο, μικρότερες από τις επιθυμητές τιμές ρυθμαπόδοσης ενός χρήστη, τότε η μείωση της παραμέτρου b θα οδηγήσει στην αύξηση της πιθανότητας αυτός να εξυπηρετηθεί και το αντίστροφο.

3.5 Ανάλυση των τροποποιήσεων της παραμέτρου b

Έχοντας ορίσει τις τροποποιήσεις που θα γίνουν στην συνάρτηση χρησιμοποίησης των χρηστών υπηρεσιών πραγματικού χρόνου που σχετίζονται με την παράμετρο b , είμαστε σε θέση να δώσουμε περισσότερες λεπτομέρειες σχετικά με τον τρόπο θα επηρεάσουμε την παράμετρο αυτή. Στις επόμενες παραγράφους παρουσιάζουμε το πεδίο τιμών της παραμέτρου και τις συναρτήσεις που τις αποδίδουν τιμές και προφανώς σχετίζονται μην την ποιότητα της υπηρεσίας που ο εκάστοτε χρήσης πραγματικού χρόνου λαμβάνει.

3.5.1 Πεδίο τιμών b

Από την στιγμή που αποδείχθηκε ότι η μετακίνηση της παραμέτρου b μπορεί να οδηγήσει σε μια λύση του προβλήματός μας, το επόμενο θέμα που πρέπει να λυθεί είναι το εύρος μέσα στο οποίο είναι δυνατόν να κινείται αυτή η παράμετρος.

Το πεδίο τιμών της παραμέτρου b για κάθε χρήστη πραγματικού χρόνου δίνεται από την παρακάτω σχέση:

$$\hat{b}_{RT,i}(t) = b_{RT,i} + b_{RT,i}(t) \quad \forall i \in S_{RT}$$

$$\hat{b}_{RT,i}(t) \in [b_{\min,i}(t), b_{\max,i}(t)]$$

$$\text{where } 0 \leq b_{\min,i}(t) \leq b_{RT,i} \leq b_{\max,i}(t) \quad (3.5)$$

όπου $b_{RT,i}$ είναι η πραγματική παράμετρος b ενός χρήστη υπηρεσίας πραγματικού χρόνου που σχετίζεται με το σχήμα μετάδοσής σου (transmission scheme) και $b_{RT,i}(t)$ είναι ο παράγοντας που δυναμικά επηρεάζει την παράμετρο αυτή έτσι ώστε να ρυθμίζεται η προτεραιότητα του εκάστοτε χρήστη σε σχέση με την επίδοσή του ως προς τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης (short term throughput constraints) που έχει θέσει. Οι παράμετροι $b_{\min,i}(t)$ και $b_{\max,i}(t)$ αποτελούν το κάτω και άνω φράγμα της παραμέτρου $b_{RT,i}$ ενός χρήστη υπηρεσίας πραγματικού χρόνου αντίστοιχα.

Ο τρόπος με τον οποίο φράσσεται η παράμετρος b σχετίζεται με τον τρόπο που αυτή επηρεάζει την συνάρτηση χρησιμοποίησης του χρήστη. Ορίζοντας τα παραπάνω φράγματα λάβαμε υπόψη τα όρια εκτός των οποίων μια περαιτέρω αλλαγή στο b δεν θα είχε ιδιαίτερη επίδραση στην προτεραιότητα του χρήστη. Σε [ακόλουθο κεφάλαιο](#) μπορείτε να παρακολουθήσετε το σκεπτικό και τους αλγόριθμους μέσα από τους οποίους υπολογίζονται τα όρια αυτά.

3.5.2 Συναρτήσεις προσδιορισμού της τιμής της παραμέτρου $b(t)$

Στόχος αυτής της ενότητας είναι η αποτύπωση της μεθόδου μέσα από την οποία μπορεί να επιτευχθεί αποδοτική ενημέρωση των παραμέτρων b των χρηστών υπηρεσιών πραγματικού χρόνου. Βασισμένοι στο πεδίο τιμών που ορίστηκε στην προηγούμενη ενότητα και στις βραχυπρόθεσμες απαιτήσεις ρυθμαπόδοσης, εισάγουμε τον αλγόριθμο όπου πριν από κάθε χρονοσχισμή συντονίζει την παράμετρο b κάθε χρήστη υπηρεσίας πραγματικού χρόνου.

Στην συνέχεια ορίζουμε την συνολική ποσότητα δεδομένων που έχει λάβει ένας χρήσης i μέσα στο παράθυρο παρατήρησης W_i από την χρονοσχισμή $t-W_i+1$, μέχρι την χρονοσχισμή $t-1$ ως:

$$B_{RT,i}^{W_i-1}(t) = \sum_{k=1}^{W_i-1} \beta_{RT,i}(t-k) \quad \forall i \in S_{RT} \quad (3.6)$$

όπου $\beta_{RT,i}(t) = U_i^{R_i^*}(P_i(t), a, b_{RT,i}) \cdot t_s$ προσδιορίζει την πραγματική ποσότητα δεδομένων που έλαβε ένας χρήστης i στην χρονοσχισμή t δεδομένης της ισχύος $P_i(t)$ υπό την οποία πραγματοποιήθηκε η μετάδοση.

Προσδιορίζοντας την παράμετρο $B_{RT,i}^{W_i-1}(t)$ κάθε χρήση υπηρεσίας πραγματικού χρόνου και συγκρίνοντάς την με το προκαθορισμένο κατώφλι δεδομένων που αυτός έχει ορίσει μπορούμε να μεταβάλλουμε την παράμετρο $\hat{b}_{RT,i}(t)$ της συνάρτησης χρησιμοποίησης που τον αντιπροσωπεύει όπως φαίνεται στην παρακάτω σχέση:

$$\hat{b}_{RT,i}(t) = \begin{cases} b_{RT,i} - I_{RT,i}(t)[b_{RT,i} - b_{\min,i}(t)] & \text{if } B_{RT,i}^{W_i-1}(t) \leq (Tr+1)B_{RT,i} \\ b_{RT,i} + G_{RT,i}(t)[b_{\max,i}(t) - b_{RT,i}] & \text{if } B_{RT,i}^{W_i-1}(t) > (Tr+1)B_{RT,i} \end{cases} \quad (3.7)$$

όπου

$$I_{RT,i}(t) \in [0,1] \quad \text{και} \quad G_{RT,i}(t) \in [0,1]$$

αποτελούν δύο κανονικοποιημένους δείκτες που αντικατοπτρίζουν την επιθυμία του χρήστη i να αποκτήσει πρόσβαση στο δίκτυο, ανάλογα με το αν έχει έλλειψη ή πλεονασμό δεδομένων στο παράθυρο παρατήρησης.

Επιπλέον ο παράγοντας Tr ($Tr \geq 1$), στον οποίον θα αναφερόμαστε ως παράγοντας πρόβλεψης του συστήματος, εκφράζει την προνοητικότητα του χρονοδρομολογητή και ο ακριβής ρόλος του θα αποδοθεί στην συνέχεια.

Σύμφωνα με την (7), όταν ένας χρήστης υπηρεσίας πραγματικού χρόνου i έχει λάβει λιγότερα δεδομένα από το προκαθορισμένο κατώφλι (αυξημένο κατά τον παράγοντα Tr), μέχρι την χρονοσχισμή $t-1$, τότε για να εμποδίσουμε την υποβάθμιση της υπηρεσίας του, η παράμετρος $\hat{b}_{RT,i}(t)$ της συνάρτησης χρησιμοποίησής του μειώνεται και κατά συνέπεια, η πιθανότητα αυτός ο χρήστης να εξυπηρετηθεί την επόμενη χρονοσχισμή αυξάνεται. Ο βαθμός της μείωσης της παραμέτρου $\hat{b}_{RT,i}(t)$ σε σχέση με την πραγματική παράμετρο $b_{RT,i}$ της συνάρτησης χρησιμοποίησης καθορίζεται από την τιμή του κανονικοποιημένου δείκτη $I_{RT,i}(t)$, την συγκεκριμένη χρονοσχισμή.

Ο δείκτης $I_{RT,i}(t)$ ενός χρήστη υπηρεσίας πραγματικού χρόνου εκφράζει την ανάγκη του χρήστη για πρόσβαση στους πόρους του συστήματος ανάλογα με την κατανομή των είδη ληφθέντων δεδομένων μέσα στο παράθυρο παρατήρησης δίνεται από τον τύπο

$$I_{RT,i}(t) = 1 - \sum_{k=1}^{W_i-1} \frac{\hat{w}_{RT,i}(k, W_i) \beta_{RT,i}(t-k)}{(Tr+1)B_{RT,i}} \quad \text{if } B_{RT,i}^{W_i-1}(t) \leq (Tr+1)B_{RT,i} \quad (3.8)$$

$$\text{where } \hat{w}_{RT,i}(k, W_i) = \frac{W_i - k}{W_i - 1} \quad \text{for } k = 1, \dots, W_i \quad \forall i \in S_{RT}$$

όπως φαίνεται από τις παραπάνω εξισώσεις, η συνάρτηση ισχύος που προσδιορίζει τα δεδομένα που έχει λάβει ο χρήστης i ανάλογα με το πόσο κοντά είναι στην τρέχουσα χρονοσχισμή είναι γραμμική.

Αντίθετα, όταν ένας χρήστης υπηρεσίας πραγματικού χρόνου i έχει λάβει περισσότερα δεδομένα από το προκαθορισμένο κατώφλι (αυξημένο κατά τον παράγοντα Tr), μέχρι την χρονοσχισμή $t-1$, τότε για να εμποδίσουμε την περαιτέρω εξυπηρέτησή του που δεν θα ωφελούσε ούτε τον ίδιο αλλά ούτε και την συνολική απόδοση του συστήματος, η παράμετρος $\hat{b}_{RT,i}(t)$ της συνάρτησης χρησιμοποίησής του αυξάνεται και κατά συνέπεια, η πιθανότητα αυτός ο χρήστης να εξυπηρετηθεί την επόμενη χρονοσχισμή μειώνεται. Όσο μεγαλύτερη είναι η ποσότητα δεδομένων που έχει λάβει ένας χρήστης σε σχέση με τις απαιτήσεις του, στο παράθυρο παρατήρησης, τόσο θα πρέπει να μειωθεί η προτεραιότητά του. Κατά συνέπεια ο κανονικοποιημένος δείκτης προσαρμογής της παραμέτρου b της συνάρτησης χρησιμοποίησης ενός «καλού» χρήστη πρέπει να έχει την ακόλουθη μορφή:

$$G_{RT,i}(t) = 1 - \frac{\sum_{k=1}^{W_i-1} \hat{w}_{RT,i}(k, W_i) \beta_{RT,i}(t-k)}{\sum_{k=1}^{W_i-1} \hat{w}_{RT,i}(k, W_i) R_i^{\max} t_s} \quad \text{if } B_{RT,i}^{W_i-1}(t) > (Tr+1)B_{RT,i} \quad (3.9)$$

Όπως φαίνεται από τον παραπάνω τύπο, όσο μεγαλύτερη είναι η ποσότητα δεδομένων που έχει λάβει ο χρήστης i στο παράθυρο παρατήρησης W και όσο πιο κοντά αυτή είναι στην μέγιστη ποσότητα δεδομένων που μπορεί να λάβει ο χρήστης, τόσο μειώνεται η προτεραιότητά του και κατά συνέπεια η πιθανότητα να εξυπηρετηθεί στην επόμενη χρονοσχισμή. Ο χρήστης δηλαδή, στην επόμενη χρονοσχισμή, πιθανόν να παραχωρήσει την θέση του σε άλλους χρήστες υπηρεσιών πραγματικού χρόνου ή μη πραγματικού χρόνου επειδή στην ουσία δεν χρειάζεται επιπλέον πρόσβαση στους πόρους του συστήματος. Κατά συνέπεια θα έπρεπε αυτός να εξυπηρετηθεί μόνο αν το κανάλι του ήταν τόσο

καλό όπου να υπερισχύουν τα ευκαιριακά χαρακτηριστικά του χρονοδρομολογητή.

Οι παράμετροι $I_{RT,i}(t) \in [0,1]$ και $G_{RT,i}(t) \in [0,1]$ όπως φαίνεται στους τύπους

(8) και (9) είναι κανονικοποιημένοι ο μεν $I_{RT,i}(t) \in [0,1]$ ως προς το κατώφλι του χρήστη i αυξημένο κατά τον παράγοντα Tr , ο δε $G_{RT,i}(t) \in [0,1]$ ως προς την μέγιστη ποσότητα δεδομένων που μπορεί ο χρήστης να λάβει μέσα στο παράθυρο παρατήρησης.

Η διαδικασία προσαρμογής των παραμέτρων b των συναρτήσεων χρησιμοποίησης επηρεάζεται επιπλέον από τον παράγοντα πρόβλεψης του συστήματος, Tr , που αποδίδει το πόσο προεκτοπιστικό (pre-emption) είναι το σύστημα. μεγαλύτερες τιμές της παραμέτρου Tr θα οδηγούσε σε ένα σύστημα που «υπερ-προστατεύει» τους χρήστες βραχυπρόθεσμων υπηρεσιών και τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης που θέτουν, μειώνοντας με αυτόν τον τρόπο την πιθανότητα της μη εκπλήρωσής τους. Αντίθετα πολύ μικρές τιμές του δείκτη Tr ίσως να οδηγούσε σε μια ταλάντωση των παραμέτρων b των χρηστών πάνω στους δύο κλάδους της εξίσωσης (7) η οποία θα αύξανε τις πιθανότητες υποβάθμισης της υπηρεσίας των χρηστών.

3.6 Προτεινόμενος αλγόριθμος χρονοδρομολόγησης – M_UPR

Βήμα 1^ο: Υπολογισμός των άνω και κάτω φραγμάτων τις παραμέτρου b των συναρτήσεων χρησιμοποίησης των χρηστών υπηρεσίας πραγματικού χρόνου. Η μεθοδολογία και οι αλγόριθμοι υπολογισμού του πεδίου τιμών της b δίνεται σε ακόλουθο κεφάλαιο.

Βήμα 2^ο: Υπολογισμός των νέων παραμέτρων \mathbf{b} με βάση την σχέση (7) και ενημέρωση των συναρτήσεων χρησιμοποίησης

Βήμα 3^ο: Το μη κυρτό (non convex) πρόβλημα βελτιστοποίησης που παρουσιάστηκε στο κεφάλαιο ξανα-διατυπώνεται με βάση τις νέες συναρτήσεις χρησιμοποίησης (οι οποίες είναι $\hat{U}_{RT,i}^{R_i^*}(P, a, \hat{\mathbf{b}}_{RT,i}(t)) \forall i \in S_{RT}$ και $\hat{U}_{NRT,j}^{R_i^*}(P, a, \mathbf{b}_{NRT,j}) \forall j \in S_{NRT}$, όπου $\mathbf{b}_{NRT,j}$ είναι η παράμετρος \mathbf{b} των συναρτήσεων χρησιμοποίησης των χρηστών υπηρεσιών μη πραγματικού χρόνου.

Βήμα 4^ο: Πραγματοποιείται η επιλογή των χρηστών που θα συμμετάσχουν στην κατανομή των πόρων του συστήματος με βάση τον αλγόριθμο επιλογής κινητών (Mobile Selection Algorithm – MSA) του UPR.

Βήμα 5^ο: Υπολογίζεται η ισχύς και ο ρυθμός μετάδοσης του κάθε χρήστη που συμμετέχει στην συγκεκριμένη χρονοσχισμή από τους τύπους $\hat{R}_{NRT,j}^{R_i^*}(t) \triangleq \hat{U}_{NRT,j}^{R_i^*}(P, a, \mathbf{b}_{NRT,j}) \forall j \in S_{NRT}$ για υπηρεσίες μη πραγματικού χρόνου και $\hat{R}_{RT,i}^{R_i^*}(t) \triangleq \hat{U}_{RT,i}^{R_i^*}(P, a, \mathbf{b}_{RT,i}) \forall i \in S_{RT}$ για υπηρεσίες πραγματικού χρόνου με βάση τον αλγόριθμο ισχύος και ρυθμού (Power and rate allocation algorithm –PAA).

Πρέπει να τονίσουμε ότι η ισχύς και ρυθμός των χρηστών υπηρεσιών πραγματικού χρόνου υπολογίζεται με τις πραγματικές συναρτήσεις χρησιμοποίησης που προβλήματος που δεν περιέχουν

ανανεωμένες παραμέτρους b έτσι ώστε να λαμβάνουμε πραγματικά αποτελέσματα ισχύος και ρυθμού μετάδοσης.

4 Αριθμητικά αποτελέσματα

Σε αυτό το μέρος της διπλωματικής αξιολογούμε τον προτεινόμενο αλγόριθμο χρονοδρομολόγησης μέσα από την μοντελοποίηση και προσομοίωση του. Με στόχο την καλύτερη απεικόνιση της επίδοσης και τις αποδοτικότητας των προτεινόμενων αλγορίθμων, σε σχέση με τη μέση συνολική πραγματική ρυθμαπόδοση του συστήματος και την ικανοποίηση των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης των χρηστών υπηρεσιών πραγματικού χρόνου, την συγκρίνουμε με την επίδοση του αλγόριθμου UPR, ο οποίος μεγιστοποιεί ασυμπτωτικά, ως προς το πλήθος των χρηστών, την επίδοση του συστήματος. Η υλοποίηση της προσομοίωσης έγινε σε Matlab 7.0 και βασίστηκε στις αρχές του δομημένου κατηγορηματικού προγραμματισμού. Ο κώδικας της προσομοίωσης υπάρχει στο CD που συνοδεύει την διπλωματικό εργασία. Η επεξεργασία των αποτελεσμάτων στο Microsoft Office Excel.

4.1 Εισαγωγικές πληροφορίες.

Κατά στην διάρκεια της μελέτης μας θεωρούμε ένα σύστημα μιας κυψέλης. Το σύστημα εξυπηρετεί τους χρήστες με πολύπλεξη κώδικα πάνω σε πολύπλεξη χρόνου (CDMA over TDMA). Η διάρκεια μιας χρονοσχισμής θεωρείται 1.67 msec και κάθε προσομοίωση διαρκεί 1000 χρονοσχισμές. Θεωρούμε ότι η σταθμός βάσης (Base station) είναι τοποθετημένος στο κέντρο της κυψέλης και έχει μέγιστη ισχύ εκπομπής $P_{max}=10$. Μοντελοποιήσαμε το κέρδος διαδρομής (gain path) από τον σταθμό βάσης μέχρι έναν χρήστη i , G_i , ως:

$$G_i = K_i / d_i^{\alpha}$$

όπου d_i είναι η απόσταση μεταξύ του χρήστη i και του σταθμού βάσης,

α' είναι ο εκθέτης απώλειας διαδρομής (distance loss exponent) ($\alpha'=4$)

και K_i είναι η τυχαία μεταβλητή που ακολουθεί λογαριθμική κανονική κατανομή διασποράς με μέση τιμή 0 και διακύμανση $\sigma^2 = 8(\text{dB})$, η οποία οφείλεται σε φαινόμενα σκίασης.

Υποθέσαμε ότι το εύρος ζώνης του συστήματος (spreading bandwidth) είναι $W=10^8$ και ότι ο μέγιστος ρυθμός μετάδοσης για όλους τους χρήστες είναι

$R_i^{\max} = 2 \cdot 10^3$ Kbps. Συνολικός αριθμός των χρηστών οι οποίοι είναι συνεχόμενα

δηλωμένοι στο σύστημα (continuously backlogged users) είναι $N=30$. Θεωρούμε

δύο κατηγορίες χρηστών στο σύστημα ανάλογα με την υπηρεσία που

χρησιμοποιούν, τους χρήστες υπηρεσιών πραγματικού χρόνου (real time users)

(NRT) και τους χρήστες υπηρεσιών μη πραγματικού χρόνου (non real time users)

(NNRT). Οι χρήστες υπηρεσιών πραγματικού χρόνου χρησιμοποιούν παράθυρο

παρατήρησης $W_i=20$ slots $\forall i \in S_{RT}$, και απαιτούν σταθερούς ρυθμούς μετάδοσης

δεδομένων $R_{RT,i} = 512 \text{ kbps}/728 \text{ kbps}$ $\forall i \in S_{RT}$. Με βάση τα παραπάνω το κατώφλι

δεδομένων που ικανοποιεί τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης

των χρηστών υπηρεσιών πραγματικού χρόνου είναι $B_{RT,i}=17002 \text{ bits}/24175 \text{ bits}$. Η

παράμετρος πρόβλεψης του συστήματος είναι $T_r=0,3$. Υποθέτουμε ότι και οι δύο

κατηγορίες χρηστών έχουν ίδιο σχήμα μετάδοσης (transmission scheme).

Επιπλέον, υποθέσαμε ότι οι παράμετροι της συνάρτησης $f_i(\gamma)$ κάθε χρήστη είναι

are $a=2$ and $b_{NRT,j} = b_{RT,i} = 3$ $\forall i \in S_{RT}, \forall j \in S_{NRT}$. Για να υπολογίσουμε την μέγιστη

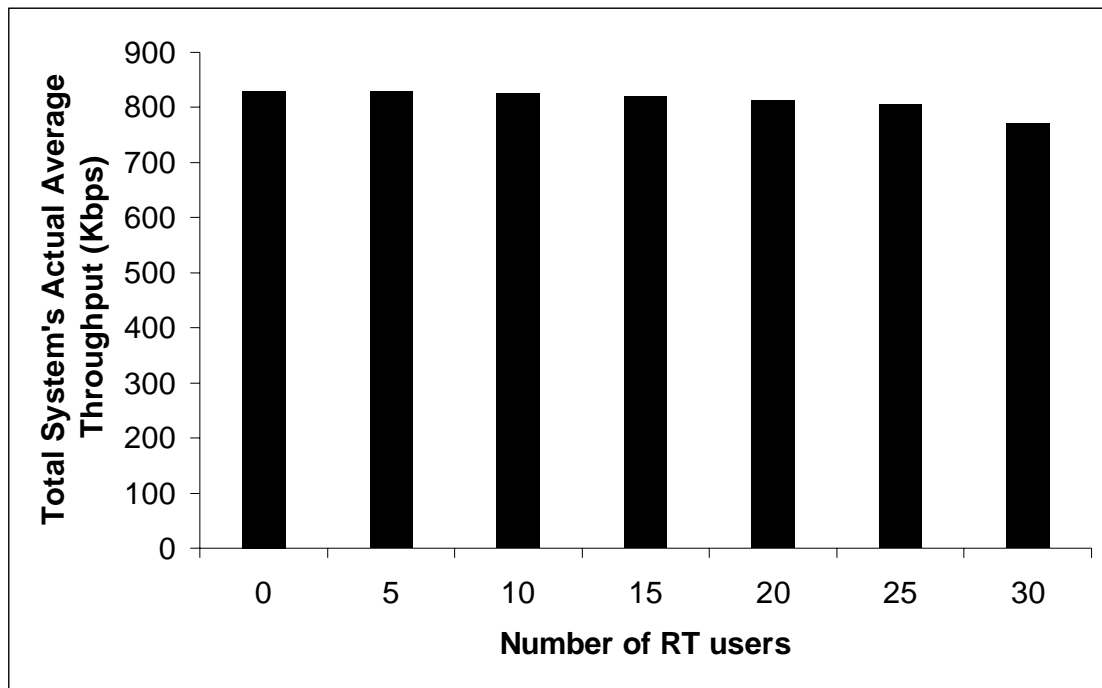
και την ελάχιστη τιμή του b για τους χρήστες υπηρεσιών πραγματικού χρόνου

θεωρήσαμε ότι $\epsilon = 10^{-5}$ και $L_{\max} = 10^5$.

4.2 Πρώτο σενάριο (Χρήστες ιδίου μέσου καναλιού – CBR 512 Kbps)

Στο πρώτο σενάριο χρονοδρομολόγησης θεωρούμε ότι το σύστημα έχει 30 χρήστες (continuously backlogged active users). Όλοι οι χρήστες έχουν τις ίδιες συνθήκες καναλιού κατά μέση τιμή (100 μέτρα απόσταση από τον σταθμό βάσης), και τον ίδιο μέγιστο ρυθμό μετάδοσης $R_{\max} = 2\text{Mbps}$. Θεωρούμε δύο κατηγορίες χρηστών, τους χρήστες υπηρεσιών πραγματικού χρόνου και αυστηρούς βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης και παράθυρο παρακολούθησης 20 χρονοσχισμών (Constant Bit Rate CBR = 512 Kbps και short-term window 20 time-slots) και τους χρήστες υπηρεσιών μη πραγματικού χρόνου που είναι ανεκτικοί στις καθυστερήσεις αλλά απαιτούν υψηλούς μέσους ρυθμούς μετάδοσης δεδομένων.

Μελετάμε τις επιδόσεις των χρηστών αφενός ως προς τις πιθανότητες αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης που περιγράφηκαν παραπάνω (αν πρόκειται για χρήστες υπηρεσιών πραγματικού χρόνου) και ως προς την μέση ρυθμαπόδοσή τους ($E[R*BER]$). Οι δύο αλγόριθμοι που συγκρίνονται είναι ο UPR (utility based power and rate allocation algorithm) και ο νέος αλγόριθμος που προτείναμε στα παραπάνω κεφάλαια M_UPR .



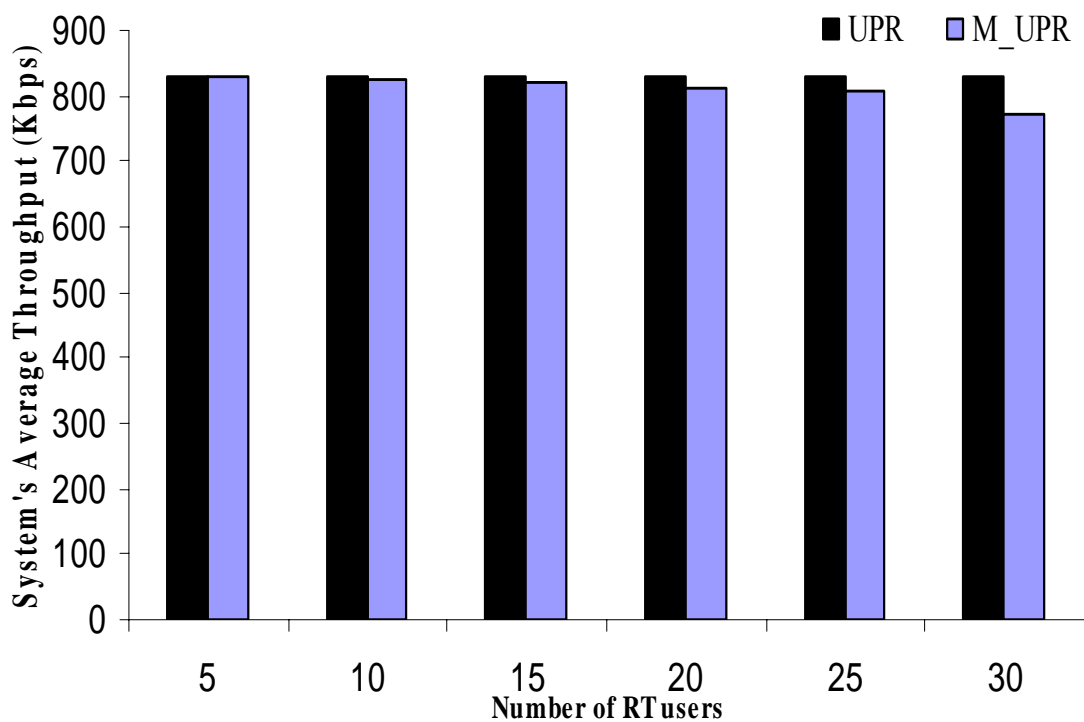
Σχήμα 4.1

Στο σχήμα 4.1 μπορούμε να δούμε την μέση συνολική ρυθμαπόδοση του συστήματος κάτω από τον αλγόριθμο M_UPR. Στον κατακόρυφο άξονα παρακολουθούμε την μέση ρυθμαπόδοση του συστήματος και στον οριζόντιο άξονα τον αριθμό των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα, καθώς αυτός αυξάνει από 0 έως 30 ανά πέντε χρήστες.

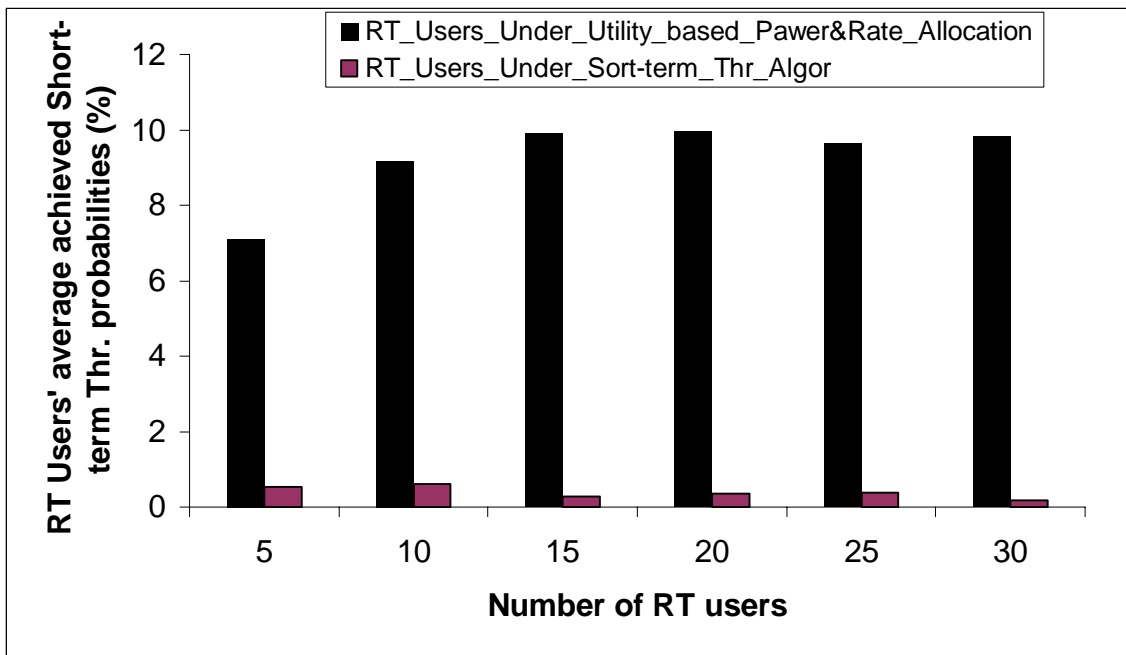
Η πρώτη στήλη στην ουσία περιγράφει την επίδοση του συστήματος κάτω από τον αλγόριθμο UPR από την στιγμή που δεν υπάρχουν χρήστες υπηρεσιών πραγματικού χρόνου στο σύστημα. Παρατηρούμε ότι η συνολική μέση ρυθμαπόδοση του συστήματος είναι μια φθίνουσα συνάρτηση του αριθμού των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα. Αυτό είναι κάτι απόλυτα αναμενόμενο από την στιγμή που όσο αυξάνονται οι χρήστες αυτοί στο σύστημα, τότε αυξάνονται και οι βραχυπρόθεσμοι περιορισμοί ρυθμαπόδοσης που το σύστημα πρέπει να εξασφαλίσει. Η μικρή μείωση όμως αυτή δεν οφείλεται μόνο στο γεγονός ότι περισσότεροι περιορισμοί εισάγονται στο σύστημα, αλλά και στο ότι χρήστες υπηρεσιών πραγματικού χρόνου με καλό

κανάλι και εξασφαλισμένους τους περιορισμούς πραγματικού χρόνου, παραχωρούν την προτεραιότητά τους σε χρήστες με χειρότερο κανάλι έτσι ώστε να καλύψουν και αυτοί τους περιορισμούς τους. Οι χρήστες με το χειρότερο κανάλι, όπως είναι αναμενόμενο θα πετύχουν χειρότερες συνθήκες μετάδοσης και κατά συνέπεια θα μειώσουν την μέση συνολική ρυθμαπόδοση του συστήματος. Η παραπάνω παρατήρηση θα γίνει πιο προφανής στην συνέχεια της ανάλυσης.

Τονίζεται ιδιαίτερα ότι η μείωση που παρατηρείται στην συνολική μέση ρυθμαπόδοση του συστήματος που επιτυγχάνει ο από εμάς προτεινόμενος αλγόριθμος είναι εξαιρετικά μικρή σε σχέση με αυτή που υπήρχε στο σύστημα όταν υπήρχαν μόνο χρήστες υπηρεσιών μη πραγματικού χρόνου (UPR). Η παρατήρηση αυτή φαίνεται πιο ξεκάθαρα στο σχήμα 4.2 όπου παρουσιάζει την μέση συνολική ρυθμαπόδοση του συστήματος (ανάλογα με τον αριθμό των χρηστών υπηρεσιών πραγματικού χρόνου) κάτω από τους αλγόριθμους UPR και M_UPR.



Σχήμα 4.2



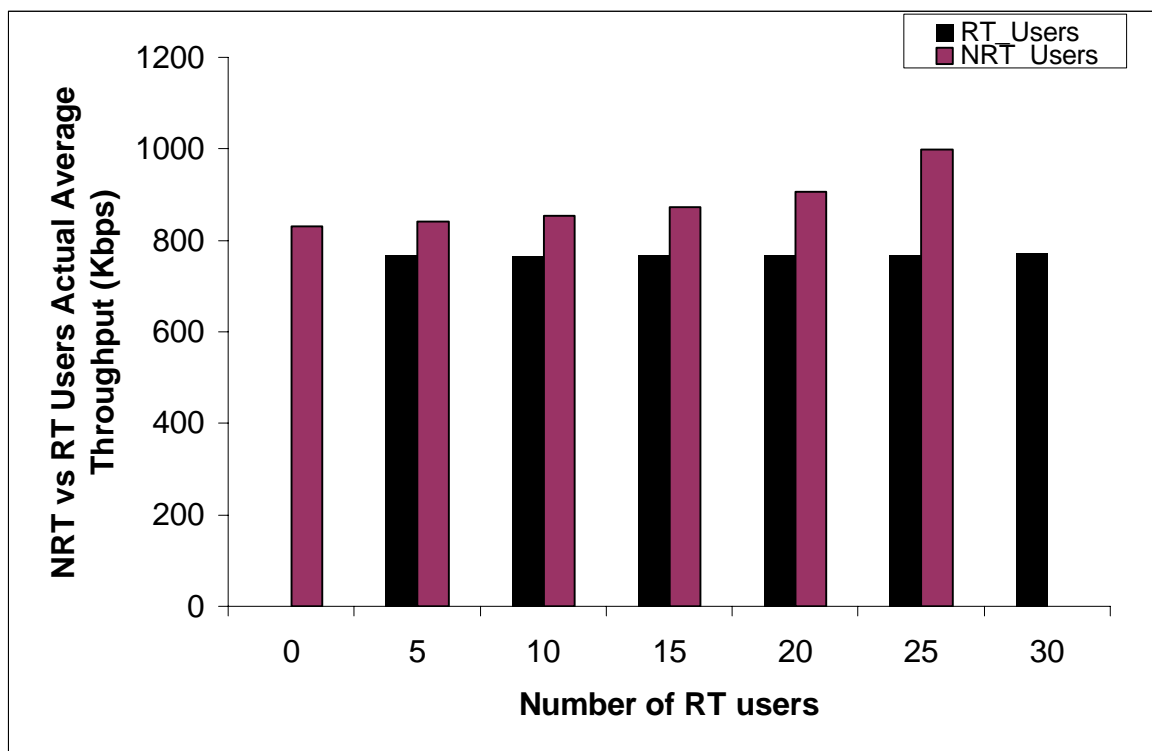
Σχήμα 4.3

Επιπροσθέτως, στο σχήμα 4.3 βλέπουμε τις πιθανότητες αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης των χρηστών υπηρεσιών πραγματικού χρόνου όσο αυτοί αυξάνονται από 5 ως 30 μέσα στο σύστημα.

Παρατηρούμε ότι οι πιθανότητες σφάλματος είναι εξαιρετικά μεγάλες στον UPR καθώς αυτός ο αλγόριθμος προσπαθεί μόνο να μεγιστοποιήσει την συνολική ρυθμαπόδοση των χρηστών. Οι πιθανότητες σφάλματος του UPR αυξάνουν ανάλογα με τον αριθμό των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα. Η παραπάνω εξέλιξη είναι απόλυτα αναμενόμενη από την στιγμή που αυξάνονται και οι περιορισμοί βραχυπρόθεσμης ρυθμαπόδοσης.

Αντίθετα, ο προτεινόμενος αλγόριθμος M_UPR πετυχαίνει εξαιρετικά μικρά ποσοστά σφάλματος όπως βλέπουμε στο σχήμα 4.3. Ακόμα και όταν ο αριθμός των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα αυξάνει, τότε οι πιθανότητες σφάλματος όχι μόνον δεν αυξάνονται αλλά μειώνονται. Η παραπάνω παρατήρηση οφείλεται στο ότι όσο αυξάνεται ο αριθμός των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα τόσο πιο πολλές συναρτήσεις χρησιμοποίησης μπορεί αυτό να ρυθμίσει, και κατά συνέπεια αυξάνεται η ευχέρεια και η δυναμικότητα του συστήματος. Επιπλέον, όσο

μικρότερος είναι ο αριθμός των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα τόσο περισσότεροι είναι οι χρήστες που ενδιαφέρονται για μεγάλους μέσους ρυθμούς ρυθμαπόδοσης (χρήστες μη πραγματικού χρόνου), οι οποίοι επιτυγχάνονται λόγω των ευκαιριακά χαρακτηριστικών χρονοδρομολόγησης του M_UPR. Αυξάνεται λοιπόν οι πιθανότητα σφάλματος, αφού αυξάνεται η πιθανότητα κάποιος χρήστης (υπηρεσίας μη πραγματικού χρόνου) να έχει τόσο καλό κανάλι που να του δίνει προτεραιότητα μεγαλύτερη και από την προτεραιότητα που το σύστημα δίνει σε έναν αδικημένο χρήση υπηρεσίας πραγματικού χρόνου με χειρότερο κανάλι.



Σχήμα 4.4

Στο σχήμα αυτό (4.4) βλέπουμε την μέση ρυθμαπόδοση που επιτυγχάνεται για χρήστες υπηρεσιών πραγματικού χρόνου και για χρήστες υπηρεσιών μη πραγματικού χρόνου κάτω από τον αλγόριθμο M_UPR και καθώς ο αριθμός της πρώτης κατηγορίας αυξάνει από 0 ως 30 μέσα στο σύστημα.

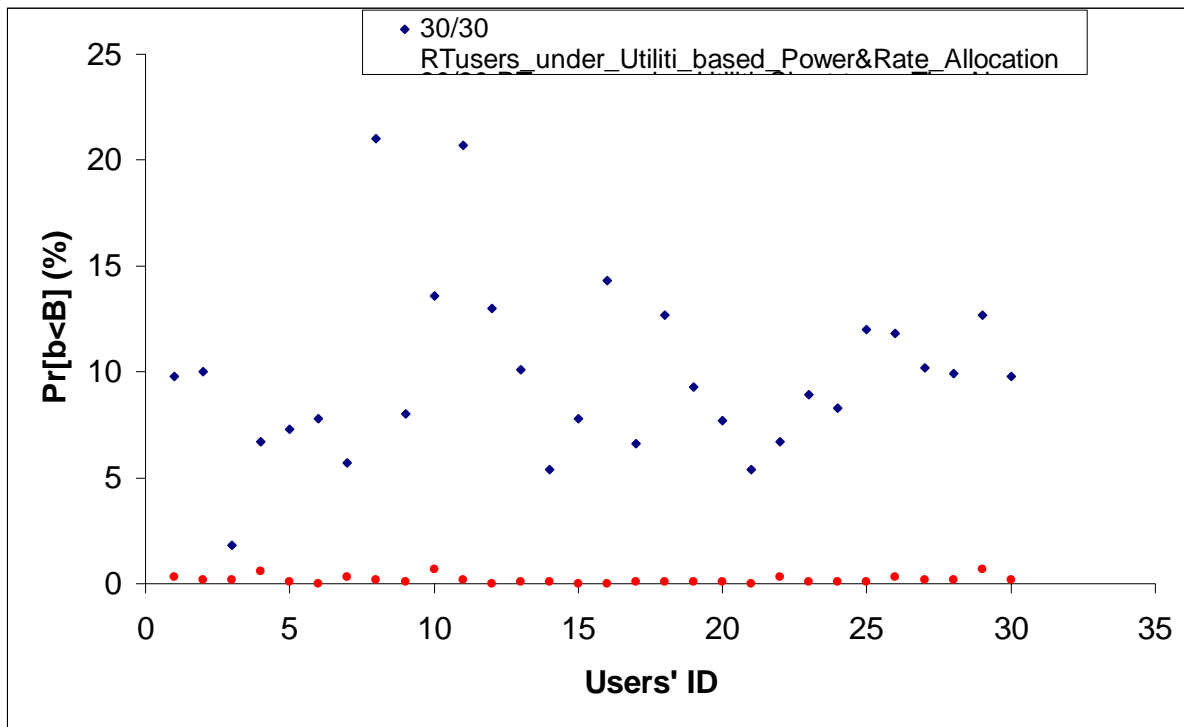
Παρατηρούμε ότι η μέση ρυθμαπόδοση των χρηστών υπηρεσιών πραγματικού χρόνου παραμένει σταθερή υπό M_UPR ανεξάρτητα με τον αριθμό τους στο

σύστημα. Αυτό αποδεικνύει την σταθερότητα που επιδεικνύει ο M_UPR ως προς την εξυπηρέτηση των υπηρεσιών αυτών.

Αντίθετα παρατηρούμε ότι η μέση ρυθμαπόδοση των χρηστών υπηρεσιών μη πραγματικού χρόνου αυξάνεται όσο ο αριθμός τους μειώνεται στο σύστημα. Αυτό οφείλεται στο γεγονός ότι όσο αυξάνονται οι χρήστες υπηρεσιών πραγματικού χρόνου τόσο αυξάνεται η πιθανότητα κάποια να ικανοποιεί τις προσδοκίες του και να παραχωρήσει την προτεραιότητά του στο δίκτυο. Αν λάβουμε υπόψη ότι ταυτόχρονα μειώνεται η ποσότητα των χρηστών υπηρεσιών μη πραγματικού χρόνου και κατά συνέπεια μοιράζεται σε λιγότερους χρήστες οι περισευούμενοι πόροι του συστήματος, μπορούμε να δικαιολογήσουμε την αύξηση αυτή της ρυθμαπόδοσης τους. Η παραπάνω παρατήρηση είναι χαρακτηριστική της συμπεριφοράς του M_UPR ο οποίος αποδεικνύει ότι ευνοεί και τους χρήστες υπηρεσιών μη πραγματικού χρόνου, παρόλο που έχει σχεδιαστεί για να εξυπηρετήσει κυρίες πραγματικού χρόνου υπηρεσίες.

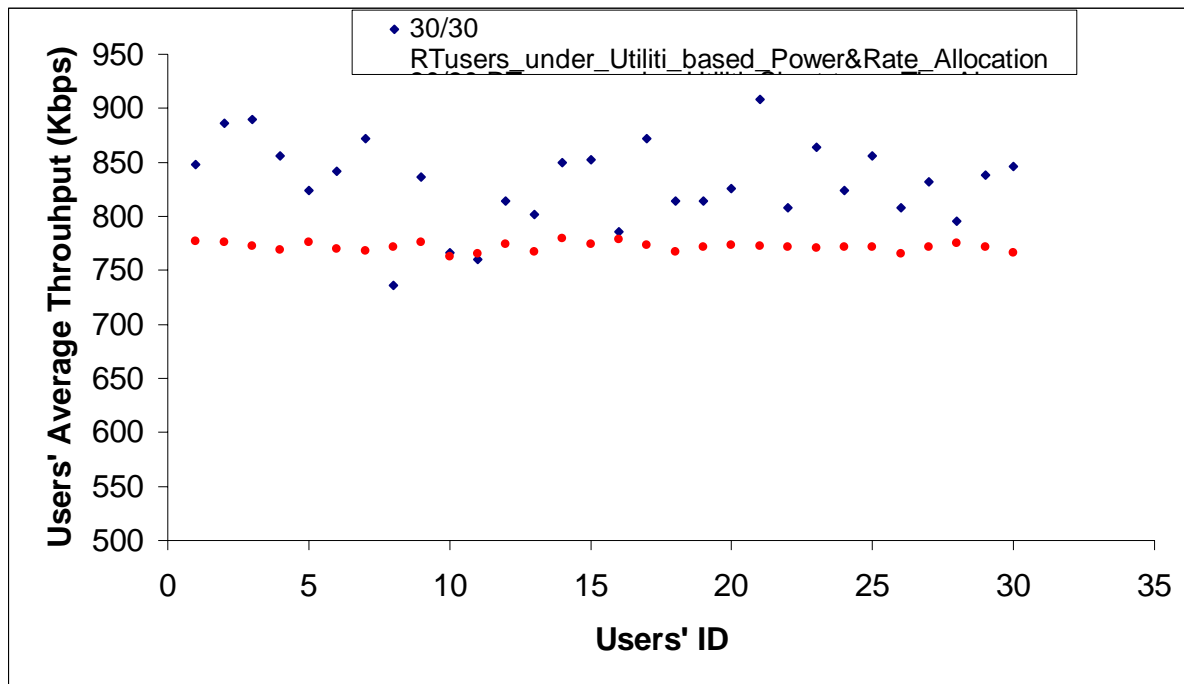
Στα δύο ακόλουθα σχήματα παρουσιάζεται η ρυθμαπόδοση και τα επίπεδα σφάλματος που πετυχαίνουν οι αλγόριθμοι UPR και M_UPR όταν στο σύστημα είναι συνδεδεμένοι 30 χρήστες που απαιτούν υπηρεσίες πραγματικού χρόνου.

Στα σχήματα 4.5 και 4.6 ο κάθετος άξονας περιγράφει τα επίπεδα αποτυχίας ανά χρήστη και την μέση ρυθμαπόδοση ανά χρήστη αντίστοιχα. Και στα δύο σχήματα ο οριζόντιος άξονας έχει το ID του κάθε χρήστη. Επιπλέον με μπλε τελείες αναπαρίστανται οι επιδόσεις του UPR ενώ με κόκκινες οι επιδόσεις του M_UPR.



Σχήμα 4.5

Από το σχήμα 4.5 γίνεται ολοφάνερο το πόσο επιτυχημένο τρόπο έχουν εκφραστεί στον χρονοδρομολογητή οι απαιτήσεις ρυθμαπόδοσης και πόσο έχει βελτιωθεί η συμπεριφορά του σε σχέση με έναν καθαρά ευκαιριακό αλγόριθμο. Παρατηρούμε ότι οι πιθανότητες αποτυχίας του UPR με 30 χρήστες υπηρεσιών πραγματικού χρόνου κυμαίνονται από 2% έως 20,7% με μέση τιμή 9.8%. Σε αντίθεση, βλέπουμε τις κόκκινες τελείες πολύ κοντά στον οριζόντιο άξονα, γεγονός που οφείλεται στις πολύ μικρές πιθανότητες αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης που πετυχαίνει ο M_UPR. Οι παραπάνω επιδώσεις αποδεικνύουν ταυτόχρονα το πόσο καλά οι βραχυπρόθεσμοι περιορισμοί εκφράζονται στον χρονοδρομολογητή, αλλά και πόσο καλά αυτός τους διαχειρίζεται.



Σχήμα 4.6

Ας ερμηνεύσουμε το σχήμα 4.6 που περιγράφει την μέση ρυθμαπόδοση κάθε χρήστη i . Παρατηρούμε ότι ο UPR οδηγεί σχεδόν πάντα σε μεγαλύτερη ρυθμαπόδοση από τον M_UPR. Το γεγονός αυτό στην προκειμένη περίπτωση δεν είναι μειονέκτημα του M_UPR όπως μπορεί να φαίνεται αλλά μειονέκτημα του UPR.

Αναλυτικότερα, μπορεί μεν οι χρήστες να έχουν μεγαλύτερη ρυθμαπόδοση κάτω από τον UPR αλλά δεν την χρειάζονται από την στιγμή που οι βραχυπρόθεσμες απαιτήσεις ρυθμαπόδοσής τους είναι άνω φραγμένες. Για παράδειγμα, ένας χρήστης που βλέπει τηλεόραση μέσω του δικτύου, έχει εκ των προτέρων καθορίσει την ρυθμαπόδοση που χρειάζεται κατά συνέπεια επιπλέον ρυθμαπόδοση θα του ήταν άχρηστη.

Ο δεύτερος λόγος που δεν ωφελεί αυτή η επιπλέον μέση ρυθμαπόδοση, είναι επειδή έχει προκύψει από μια ταλαντούμενη ρυθμαπόδοση σε σχέση με τον χρόνο που έτρεχε το σύστημα. Συγκεκριμένα, αν παρατηρήσουμε ταυτόχρονα τα σχήματα 4.5 και 4.6 θα παρατηρήσουμε την εξής αντίφαση: τα μεγαλύτερα

επίπεδα ρυθμαπόδοσης του UPR συνοδεύονται και από μεγαλύτερα σφάλματα σε σχέση με τον M_UPR.

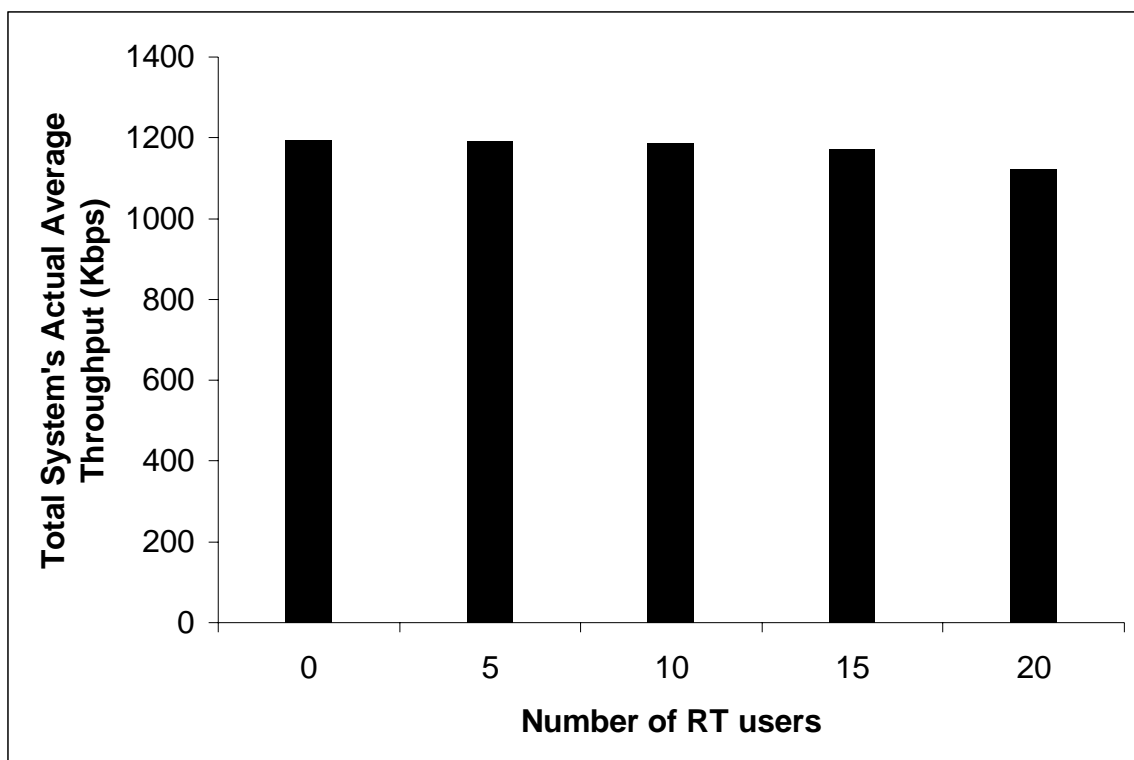
Η αντίφαση αυτή μπορεί να εξηγηθεί αν λάβουμε υπόψη ότι ο M_UPR πετυχαίνει να σταθεροποιήσει την στιγμιαία ρυθμαπόδοση κάθε χρήστη κοντά στις απαιτήσεις του, αντίθετα ο UPR προσφέρει στιγμιαία ρυθμαπόδοση που τείνει να ακολουθεί το κανάλι του χρήστη και κατά συνέπεια μεταβάλλεται με τυχαίο τρόπο.

4.3 Δεύτερο σενάριο (Χρήστες ιδίου μέσου καναλιού – CBR 728 Kbps)

Στο δεύτερο σενάριο χρονοδρομολόγησης θεωρούμε ότι το σύστημα έχει 20 χρήστες (continuously backlogged active users). Όλοι οι χρήστες έχουν τις ίδιες συνθήκες καναλιού κατά μέση τιμή (100 μέτρα απόσταση από τον σταθμό βάσης), και τον ίδιο μέγιστο ρυθμό μετάδοσης $R_{\max} = 2\text{Mbps}$. Θεωρούμε δύο κατηγορίες χρηστών, όπως και στο πρώτο σενάριο, τους χρήστες υπηρεσιών πραγματικού χρόνου και αυστηρούς βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης και παράθυρο παρακολούθησης 20 χρονοσχισμών (Constant Bit Rate CBR = 728 Kbps – αντίθετα με το πρώτο σενάριο – και short-term window 20 time-slots) και τους χρήστες υπηρεσιών μη πραγματικού χρόνου που είναι ανεκτικοί στις καθυστερήσεις αλλά απαιτούν υψηλούς μέσους ρυθμούς μετάδοσης δεδομένων.

Μελετάμε τις επιδόσεις των χρηστών αφενός ως προς τις πιθανότητες αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης που περιγράφηκαν παραπάνω (αν πρόκειται για χρήστες υπηρεσιών πραγματικού χρόνου) και ως προς την μέση ρυθμαπόδοσή τους ($E[R*BER]$). Οι δύο αλγόριθμοι που συγκρίνονται είναι ο UPR (utility based power and rate allocation algorithm) και ο νέος αλγόριθμος που προτείναμε στα παραπάνω κεφάλαια M_UPR.

Με αυτόν τον τρόπο θα μελετήσουμε πως αντιδρούν οι δύο αλγόριθμοι ως προς την μέση ρυθμαπόδοση και πιθανότητα σφάλματος ως προς την εκπλήρωση των βραχυπρόθεσμων απαιτήσεων (στο σύστημα, αλλά και ανά χρήση) υπό μεγαλύτερο φορτίο.



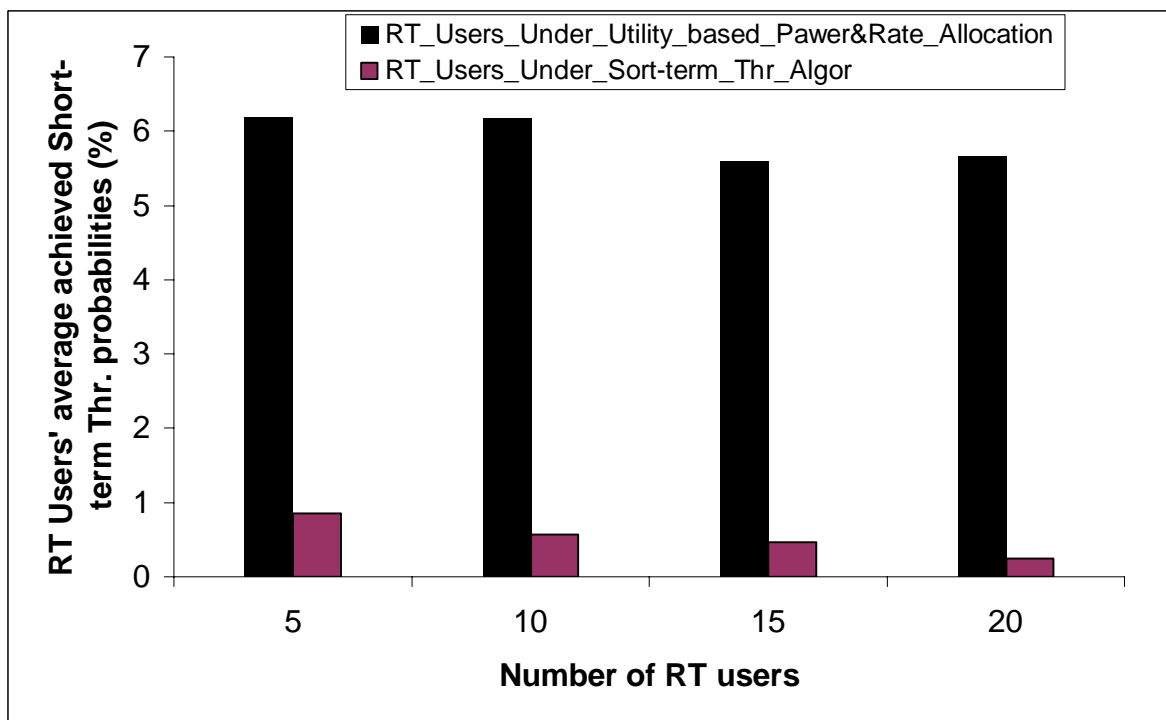
Σχήμα 4.7

Στο σχήμα 4.7 μπορούμε να δούμε την μέση συνολική ρυθμαπόδοση του συστήματος κάτω από τον αλγόριθμο M_UPR. Στον κατακόρυφο άξονα παρακολουθούμε την μέση ρυθμαπόδοση του συστήματος και στον οριζόντιο

άξονα τον αριθμό των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα, καθώς αυτός αυξάνει από 0 έως 20 ανά πέντε χρήστες.

Όπως και στο σχήμα 4.1, παρατηρούμε μια μικρή μείωση της μέσης ρυθμαπόδοσης όσο οι χρήστες υπηρεσιών πραγματικού χρόνου αυξάνονται στο σύστημα. όπως έχουμε εξηγήσει αυτό οφείλεται σε τρεις λόγους

- 1) Αφενός, αυξάνονται οι περιορισμοί που πρέπει να ικανοποιήσει το σύστημα
- 2) Αφετέρου, οι υπηρεσίες των χρηστών πραγματικού χρόνου είναι άνω φραγμένες.
- 3) Τέλος, οι ευνοημένοι χρήστες παρατηρούν την προτεραιότητά τους στους «στους κακούς χρήστες» οι οποίοι πετυχαίνουν χειρότερες συνθήκες μετάδοσης και μειώνουν την μέση ρυθμαπόδοση σε αντίθεση με έναν καθαρά ευκαιριακό αλγόριθμο.

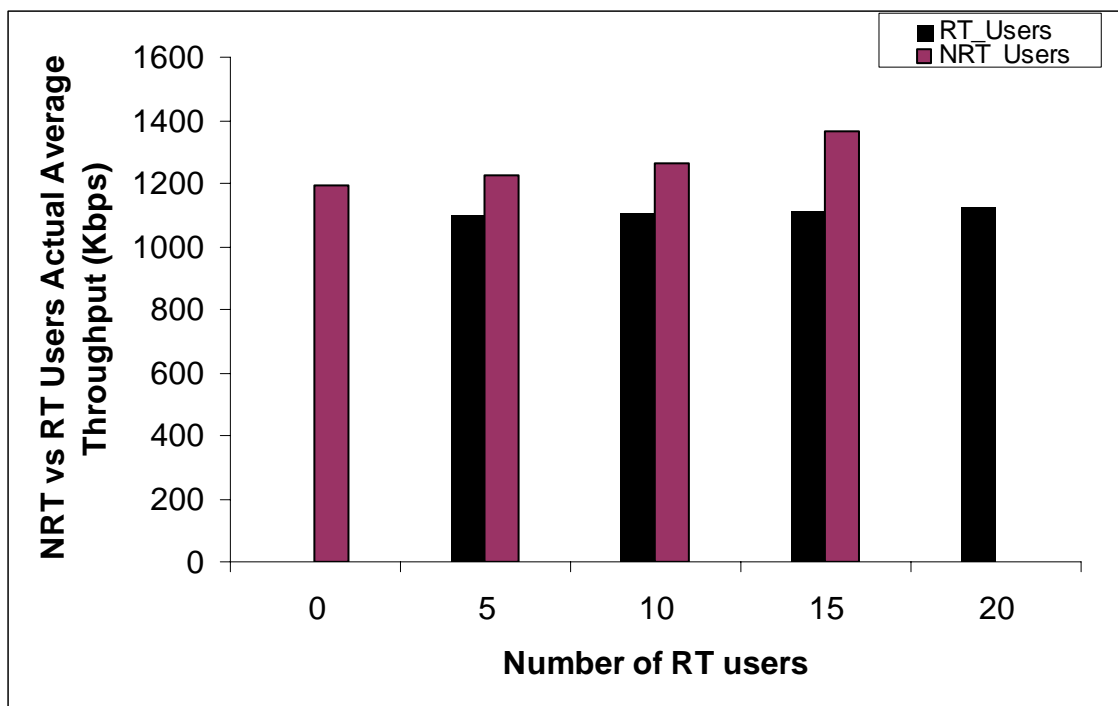


Σχήμα 4.8

Στο σχήμα 4.8 βλέπουμε τις πιθανότητες αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης των χρηστών υπηρεσιών πραγματικού χρόνου όσο αυτοί αυξάνονται από 5 ως 20 μέσα στο σύστημα. Παρατηρούμε, όπως και στο σχήμα 4.3 μεγάλη διαφορά ανάμεσα στις δύο αλγορίθμους.

Παρόλο που οι απαιτήσεις των χρηστών υπηρεσιών πραγματικού χρόνου έχουν αυξηθεί πολύ παρατηρούμε ότι ο M_UPR συνεχίζει να πετυχαίνει ποσοστά αποτυχίας μικρότερα από 1%.

Οι λόγοι που τα ποσοστά αυτά μεταβάλλονται αντιστρόφως ανάλογα με τον αριθμό των χρηστών υπηρεσιών πραγματικού χρόνου στο σύστημα έχουν εξηγηθεί στο σχήμα 4.3.



Σχήμα 4.9

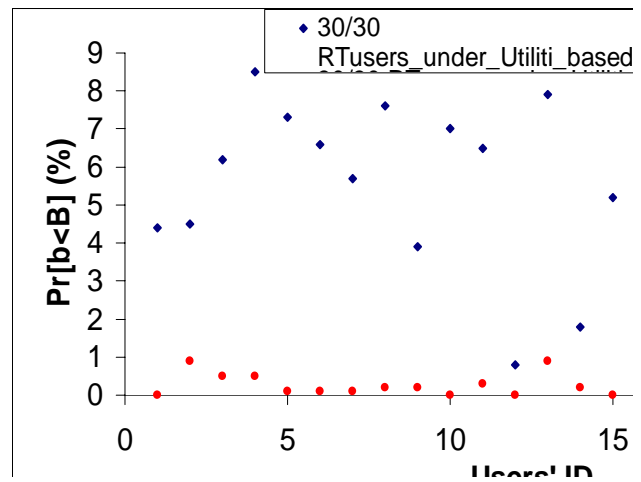
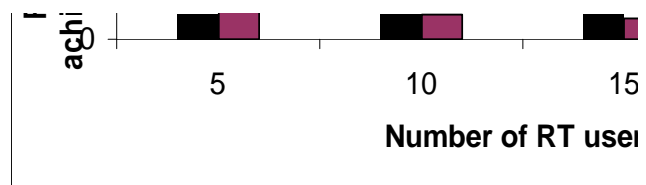
Στο σχήμα αυτό (4.9) βλέπουμε την μέση ρυθμαπόδοση που επιτυγχάνεται για χρήστες υπηρεσιών πραγματικού χρόνου και για χρήστες υπηρεσιών μη πραγματικού χρόνου κάτω από τον αλγόριθμο M_UPR και καθώς ο αριθμός της πρώτης κατηγορίας αυξάνει από 0 ως 20 μέσα στο σύστημα.

Παρατηρούμε την σταθεροποίηση της ρυθμαπόδοσης των χρηστών υπηρεσιών πραγματικού χρόνου ανεξάρτητα με τον αριθμό τους στο σύστημα

Αντίθετα οι υπηρεσίες μη πραγματικού χρόνου εξυπηρετούνται με όλο και μεγαλύτερη ρυθμαπόδοση, για του λόγους που έχουμε εξηγήσει και στο σχήμα 4.4, όσο ο αριθμός των χρηστών που τις ζητούν μικραίνει στο σύστημα. παρατηρούμε ότι όταν έχουμε μόνο 5 χρήστες υπηρεσιών μη πραγματικού χρόνου στο σύστημα, τότε, επιτυγχάνεται ρυθμαπόδοση της τάξης των 1400 Kbps.

Στα δύο ακόλουθα σχήματα παρουσιάζεται η ρυθμαπόδοση και τα επίπεδα σφάλματος που πετυχαίνουν οι αλγόριθμοι UPR και M_UPR όταν στο σύστημα είναι συνδεδεμένοι 20 χρήστες που απαιτούν υπηρεσίες πραγματικού χρόνου.

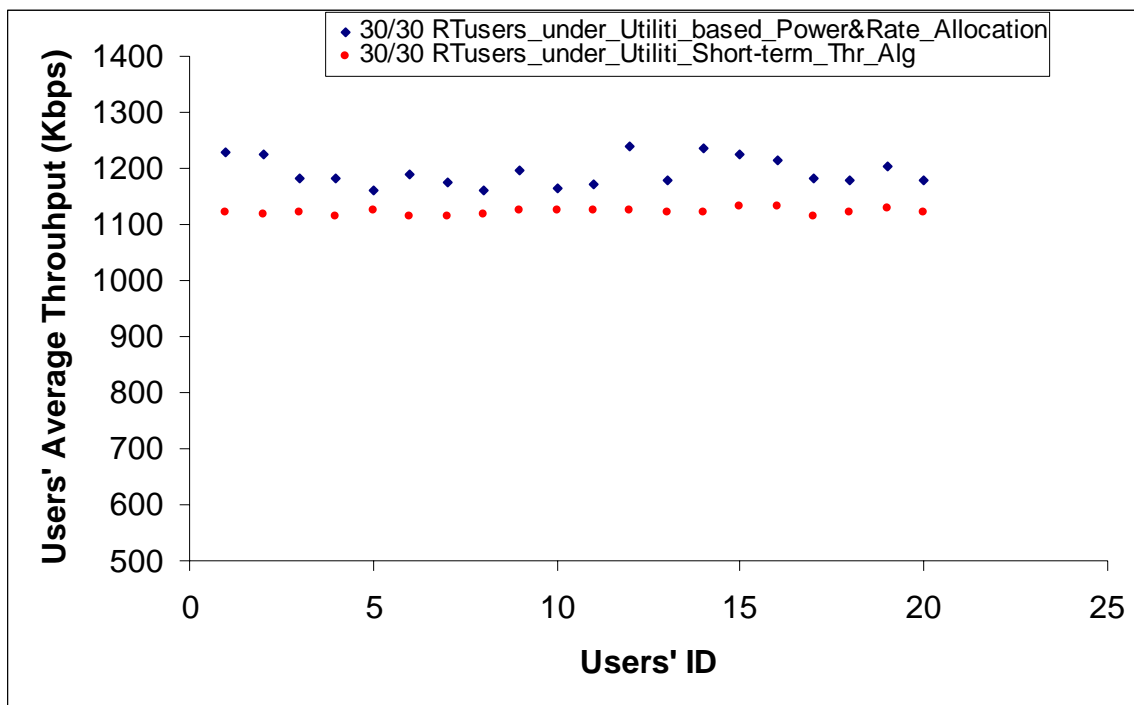
Στα σχήματα 4.10 και 4.11 ο κάθετος άξονας περιγράφει τα επίπεδα αποτυχίας ανά χρήστη και την μέση ρυθμαπόδοση ανά χρήστη αντίστοιχα. Και στα δύο σχήματα ο οριζόντιος άξονας έχει το ID του κάθε χρήστη. Επιπλέον με μπλε τελείες αναπαρίστανται οι επιδόσεις του UPR ενώ με κόκκινες οι επιδόσεις του M_UPR.



Σχήμα 4.10

Στα δύο τελευταία σχήματα παρακολουθούμε είναι εμφανές ότι ο M_UPR πετυχαίνει να σταθεροποιήσει την ρυθμαπόδοση των χρηστών κοντά στις βραχυπρόθεσμες απαιτήσεις που θέτουν οι πραγματικού χρόνου υπηρεσίες τους. Σε αντίθεση ο UPR έχοντας μεγάλες ταλαντώσεις την στιγμιαία ρυθμαπόδοση (η οποία τείνει να ακολουθήσει την ποιότητα καναλιού των χρηστών) επιτυγχάνει μεν μεγαλύτερες τιμές ρυθμαπόδοσης, χωρίς όμως να προσφέρει ουσιαστικά στο δίκτυο από την στιγμή που οι χρήστες δεν χρειάζονται τίποτα παραπάνω από τις υπηρεσίες που εκφράζονται μέσα από τις απαιτήσεις τους και επιπλέον οι πόροι που προκύπτουν από αυτόν τον πλεονασμό δεν μπορούν να μοιραστούν σε χειρότερους χρήστες.

Τα ποσοστά αποτυχίας του M_UPR συνεχίζουν να είναι μικρότερα της μιας ποσοστιαίας μονάδας σε αντίθεση με τον UPR που όπως έχουμε επαναλάβει ενδιαφέρεται μόνο για την μεγιστοποίηση της ρυθμαπόδοσης.



σχήμα 4.11

4.4 Τρίτο σενάριο (Χρήστες διαφορετικών μέσων καναλιών)

Στο δεύτερο σενάριο χρονοδρομολόγησης θεωρούμε ότι το σύστημα έχει 30 χρήστες (continuously backlogged active users). Όλοι οι χρήστες έχουν τον ίδιο μέγιστο ρυθμό μετάδοσης $R_{max} = 2\text{Mbps}$.

Για να μελετήσουμε τις επιδόσεις των δύο αλγορίθμων κάτω από ετερογενές, ως προς την ποιότητα καναλιού, δείγμα χρηστών, χωρίζουμε τους χρήστες σε δύο κατηγορίες:

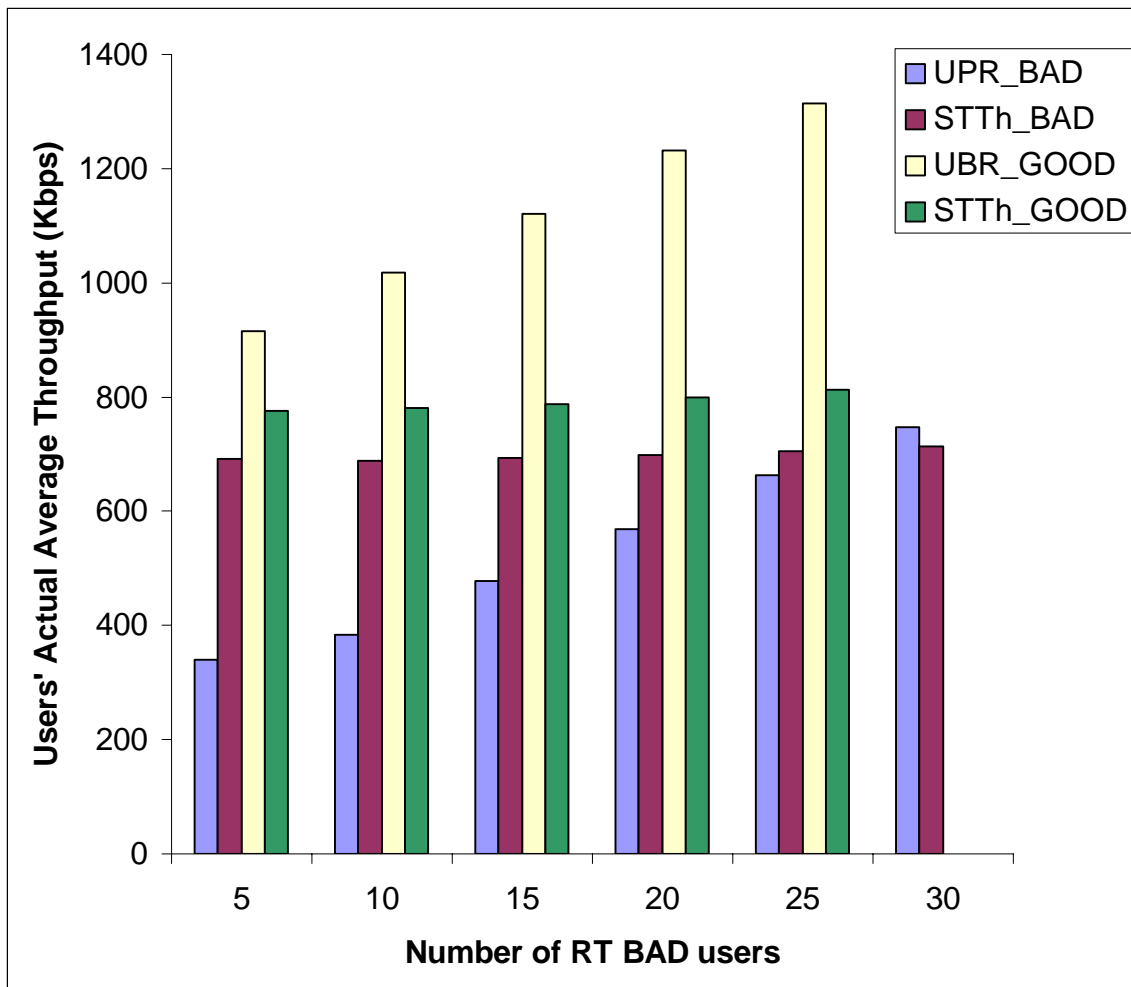
- I. χρήστες με καλό κανάλι,
απόσταση από το κέντρο της κυψέλης $d_g = 100\text{m}$

- II. χρήστες με κακό κανάλι,
απόσταση από το κέντρο της κυψέλης $d_g = 150\text{m}$

όπου με βάση τον εκθέτη απωλειών λόγω της απόστασης η διαφορά στο κανάλι των χρηστών είναι της τάξης των 7dB.

Θεωρούμε ότι το σύστημα έχει χρήστες υπηρεσιών πραγματικού χρόνου και αυστηρούς βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης και παράθυρο παρακολούθησης 20 χρονοσχημάτων (Constant Bit Rate CBR = 512 Kbps – αντίθετα με το πρώτο σενάριο – και short-term window 20 time-slots)

Μελετάμε τις επιδόσεις των χρηστών αφενός ως προς τις πιθανότητες αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης που περιγράφηκαν παραπάνω (αν πρόκειται για χρήστες υπηρεσιών πραγματικού χρόνου) και ως προς την μέση ρυθμαπόδοσή τους ($E[R*BER]$) όσο ο αριθμός των κακών χρηστών αυξάνεται στο δίκτυο. Οι δύο αλγόριθμοι που συγκρίνονται είναι ο UPR (utility based power and rate allocation algorithm) και ο νέος αλγόριθμος που προτείναμε στα παραπάνω κεφάλαια M_UPR.



Σχήμα 4.12

Στο σχήμα 4.12 μπορούμε να δούμε την μέση συνολική ρυθμαπόδοση του συστήματος κάτω από τους αλγόριθμους UPR και M_UPR. Στον κατακόρυφο άξονα παρακολουθούμε την μέση ρυθμαπόδοση του συστήματος και στον οριζόντιο άξονα τον αριθμό των χρηστών με κακό κανάλι στο σύστημα, καθώς αυτός αυξάνει από 0 έως 30 ανά πέντε χρήστες.

Κοιτώντας τις στήλες από τα αριστερά έχουμε:

- i. Η πρώτη στήλη περιγράφει την ρυθμαπόδοση των κακών χρηστών κάτω από τον UPR.

- ii. Η δεύτερη στήλη περιγράφει την ρυθμαπόδοση των κακών χρηστών κάτω από τον M_UPR.
- iii. Η τρίτη στήλη περιγράφει την ρυθμαπόδοση των καλών χρηστών κάτω από τον UPR.
- iv. Η τέταρτη στήλη περιγράφει την ρυθμαπόδοση των καλών χρηστών κάτω από τον M_UPR.

Το σχήμα 4.12 μπορούμε να το μελετήσουμε από πολλές πλευρές.

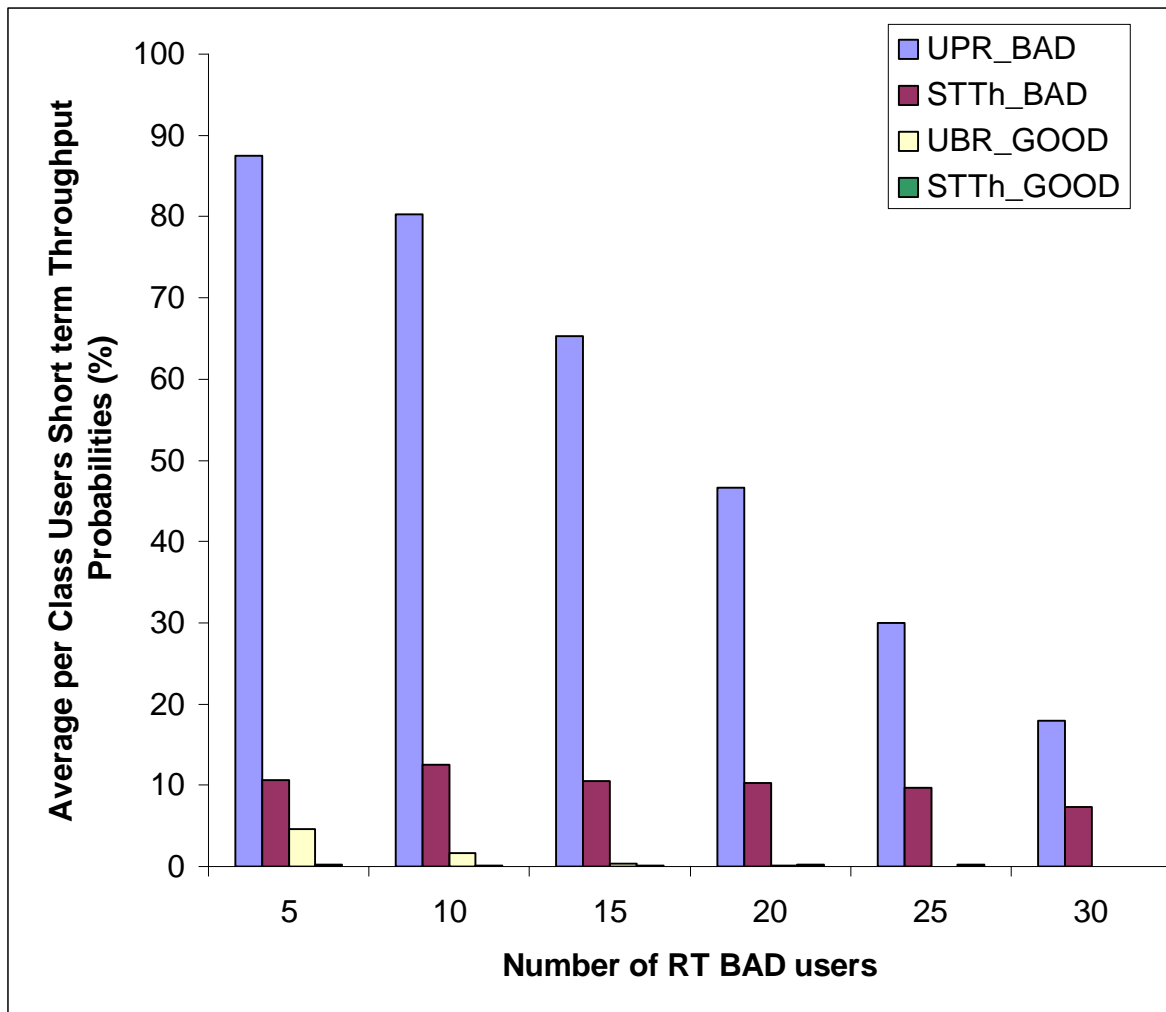
Αρχικά ας μελετήσουμε τις επιδώσεις του UPR. Παρατηρούμε ότι υπάρχει μεγάλη διαφορά ανάμεσα στην μέση ρυθμαπόδοση των χρηστών με καλό και κακό κανάλι. Αυτό είναι απολύτως λογικό γιατί ο UPR θέλει να μεγαλώσει την μέση ρυθμαπόδοση του συστήματος (και είναι προσεγγιστικά βέλτιστος σε αυτό) χωρίς όμως να έχει αίσθηση δικαιοσύνης ανάμεσα στους χρήστες αλλά και σεβασμού ως προς βραχυπρόθεσμες απαιτήσεις. Οι χρήστες με καλό κανάλι θα επιλέγονται περισσότερες φορές από τον MSA και θα πετυχαίνουν καλύτερο κανάλι αυξάνοντας την ρυθμαπόδοσή τους, αντίθετα με τους κακούς χρήστες που θα βρίσκονται συνήθως στις τελευταίες θέσεις του MSA και κατά συνέπεια θα τυγχάνουν χαμηλής εξυπηρέτησης.

Αντίθετα ο M_UPR κρατάει και τις δύο κατηγορίες χρηστών πολύ κοντά μεταξύ τους αλλά και πάνω από τις βραχυπρόθεσμες απαιτήσεις ρυθμαπόδοσης που έχουν θέσει στο σύστημα. Ας εξηγήσουμε για ποιο λόγο όμως υπάρχει αυτή η απόκλιση ανάμεσα στους καλούς και τους κακούς χρήστες. Όπως έχουμε πει προηγουμένως, ένας χρήστης υπηρεσίας πραγματικού χρόνου ταλαντεύεται (στον M_UPR) γύρω από το κατώφλι που θέτουν οι απαιτήσεις τους (αυξημένο κατά τον παράγοντα πρόβλεψης μέλλοντος). Όταν ένας χρήστης είναι κάτω από αυτό το κατώφλι γίνεται πιο απαιτητικός ενώ αντίθετα όταν είναι πάνω από αυτό, τότε τείνει να παραχωρήσει την προτεραιότητά του. Παρακολουθώντας την μέση ρυθμαπόδοση που επιτυγχάνεται πρέπει να έχουμε υπόψη το μέσο b

των χρηστών και κατά συνέπεια που ισορροπούν κατά μέσο όρο σε σχέση με αυτό το όριο. Με βάση τα παραπάνω συμπεραίνουμε ότι οι χρήστες με κακό κανάλι ισορροπούν λίγο παρακάτω από το αυξημένο κατά Tr κατώφλι αλλά πάντα παραπάνω από το κατώφλι των απαιτήσεών του. Αυτό σημαίνει ότι τείνουν να απαιτούν μεγαλύτερη προτεραιότητα από αυτή που αξίζουν από το σύστημα. Αντίθετα οι χρήστες με καλό κανάλι ισορροπούν λίγο παραπάνω από το αυξημένο κατά Tr κατώφλι άρα τείνουν να παραχωρούν την προτεραιότητά τους (και κατά συνέπεια να εξυπηρετούνται μόνο όταν μπορούν να πετύχουν εξαιρετικές συνθήκες μετάδοσης).

Όσο οι κακοί χρήστες αυξάνονται στο σύστημα τότε αυξάνεται η επίδοση των καλών και των κακών χρηστών στο σύστημα. Αυτό οφείλεται στην μείωση των μεγάλων αποκλίσεων που δημιουργούνται από τον UPR όταν έχουμε πολλούς καλούς χρήστες.

Η λόγος που για 30 καλούς χρήστες οι επιδώσεις ρυθμαπόδοσης του UPR είναι καλύτερες από τον M_UPR έχουν εξηγηθεί σε προηγούμενα σενάρια και έχει αποδειχθεί ότι το γεγονός αυτό οφείλεται στις πιθανότητες αποτυχίας που επιτυγχάνουν τα δύο μοντέλα και απεικονίζονται στο σχήμα 4.13.



Σχήμα 4.13

Στο σχήμα 4.13 βλέπουμε τις πιθανότητες αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης των χρηστών υπηρεσιών πραγματικού χρόνου όσο οι χρήστες με κακό κανάλι αυξάνονται από 5 ως 30 μέσα στο σύστημα.

Παρατηρούμε ότι οι κακοί χρήστες κάτω από τον UPR αδικούνται όπως φαίνεται από το ποσοστά αποτυχίας των υπηρεσιών τους που αγγίζουν το 90%.

Στον M_UPR αντίθετα και οι δύο κατηγορίες επιτυγχάνουν πολύ μικρά ποσοστά αποτυχίας. Ειδικά τους χρήστες με καλό κανάλι η πράσινη στείλει δεν υπάρχει καν, το οποίο σημαίνει σχεδόν μηδενική πιθανότητα αποτυχίας των περιορισμών ρυθμαπόδοσης.

5 Επίλογος

Στον επίλογο της διπλωματικής εργασίας θα συνοψίσουμε την ανάλυση και τα αποτελέσματα της εργασίας και θα αναφέρουμε περαιτέρω προβλήματα που πρέπει να λυθούν και περαιτέρω έρευνα που μπορεί να γίνει στο θέμα που εξετάσαμε.

5.1 Σύνοψη και συμπεράσματα

Στην παρούσα διπλωματική εργασία αναλύσαμε το πρόβλημα απόδοσης πόρων κατά την ζεύξη καθόδου σε ασύρματα δίκτυα που υποστηρίζουν ετερογενή περιβάλλοντα και υπηρεσίες πολυμέσων. Οι αλγόριθμοι απόδοσης πόρων που μελετήθηκαν επιτυγχάνουν την ικανοποίηση βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης κρατώντας σε επίπεδα κοντά στο βέλτιστο την συνολική ρυθμαπόδοση (Throughput) του συστήματος. Με την ικανοποίηση των βραχυπρόθεσμων περιορισμών ποιότητας υπηρεσίας εξασφαλίζουμε στους χρήστες πραγματικού χρόνου την εξασφάλιση όχι μόνο της εξυπηρέτησης του σε τακτά χρονικά διαστήματα από το σύστημα αλλά και την απόδοση της ελάχιστης αναγκαίας ρυθμαπόδοσης που απαιτείται για την επιτυχή εξυπηρέτηση των υπηρεσιών πραγματικού χρόνου. Αυτό είναι ιδιαίτερα σημαντικό για τους χρήστες πραγματικών εφαρμογών που εκτελούν εφαρμογές τηλεδιάσκεψης, προβολής βίντεο ή λοιπές εφαρμογές πολυμέσων γιατί η ποιότητα της υπηρεσίας εξαρτάται σε πολύ μεγάλο βαθμό από την εξασφάλιση κατ' ελάχιστον, της απόδοσης ενός ποσοστού των διαθέσιμων πόρων του συστήματος για διαδοχικά χρονικά παράθυρα της εφαρμογής.

Οι αλγόριθμοι που προτάθηκαν χρησιμοποιούν την έννοια της συνάρτησης χρησιμοποίησης που εκφράζει την ευχαρίστηση κάθε χρήστη από την συνολική ικανοποίηση των περιορισμών ποιότητας υπηρεσιών που έχει δεχθεί μέχρι το σημείο εξέτασης. Οι αλγόριθμοι που προτάθηκαν έχουν στηριχθεί σε ιδέες που έχουν αναπτυχθεί στην βιβλιογραφία και βελτιστοποιούν την συνολική

ρυθμαπόδοση του συστήματος αγνοώντας όμως τους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης που αναφέραμε παραπάνω. Η ικανοποίηση των περιορισμών αυτών γίνεται με κατάλληλη μετατροπή των συναρτήσεων χρησιμοποίησης μέσω «ρύθμισης» μίας παραμέτρου ελέγχου για κάθε χρήστη ούτως ώστε να μπορέσουν να εκφράζουν την πληροφορία ικανοποίησης των βραχυπρόθεσμων περιορισμών που έχει θέσει κάθε χρήστης στο παράθυρο ανάλυσης. Μελετώνται επίσης τα όρια ρύθμισης της παραμέτρου ελέγχου και μετατροπής της συνάρτησης μεταφοράς και αποδεικνύονται εκτεταμένα θεωρήματα που εξασφαλίζουν την αποδοτικότητα και ορθότητα των αλγορίθμων που προτάθηκαν ως προς την ικανοποίηση των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης. Τέλος αναλύονται παραδείγματα προσομοίωσης για διάφορες περιπτώσεις και σενάρια που αποδεικνύουν ότι οι αλγόριθμοι που προτείνονται ικανοποιούν με μεγάλη επιτυχία τους διάφορους βραχυπρόθεσμους περιορισμούς ρυθμαπόδοσης που έχει θέσει κάθε χρήστης πραγματικού χρόνου χωρίς να υποβαθμίζουν την συνολική ρυθμαπόδοση του συστήματος. Η σημασία του αλγορίθμου φαίνεται από το γεγονός ότι διατηρεί μια καλή προσέγγιση στην βέλτιστη απόδοση ρυθμαπόδοσης στους χρήστες του δικτύου και παράλληλα μπορεί να αποδώσει δικαιοσύνη στους χρήστες πραγματικών εφαρμογών εξασφαλίζοντας πρακτικά μηδενική πιθανότητα αποτυχίας στην αποδοχή των αναγκαίων πόρων για την επιτυχία της υπηρεσίας τους μέσα σε αλληλοεπικαλυπτόμενα παράθυρα ελέγχου. Επίσης προβάλλεται η απόδοση των αλγορίθμων και σε σενάρια με υψηλό ποσοστό «κακών» χρηστών, δηλαδή χρηστών με χαμηλή μέση τιμή ποιότητας καναλιού, όπου παρατηρούμε επίσης πολύ μικρά ποσοστά αποτυχίας των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης με πολύ μικρές απώλειες στην βέλτιστη απόδοση συνολικής ρυθμαπόδοσης του συστήματος όπως προκύπτει από την σύγκριση του αλγορίθμου μας με έναν αλγόριθμο από την βιβλιογραφία που εξασφαλίζει βελτιστότητα στην συνολικά αποδιδόμενη ρυθμαπόδοση αλλά αδυναμία στην

ικανοποίηση των βραχυπρόθεσμων περιορισμών που εμείς προσπαθούμε να επιλύσουμε.

5.2 Μελλοντικές επεκτάσεις

Το πρόβλημα απόδοσης πόρων που εξετάσαμε στην παρούσα διπλωματική είναι ένα πρόβλημα που έχει γνωρίσει πολλές προεκτάσεις στην διεθνή βιβλιογραφία. Στην προσέγγιση που προτείνουμε υπάρχουν ακόμα διάφορες προεκτάσεις που μπορούν να αναλυθούν και να ερευνηθούν περαιτέρω. Μια σημαντική προσέγγιση είναι η απόδειξη των συνθηκών συνολικής ικανοποιησιμότητας του συστήματος (feasibility constraints) για την υποστήριξη των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης κάθε χρήστη που έχει εισαχθεί στο σύστημα. Μέχρι τώρα έχουμε υποθέσει ότι υπάρχει μία υπηρεσία ελέγχου αποδοχής χρηστών (call admission control) που ελέγχει αν επιτρέπεται ένας χρήστης να εισέλθει στο σύστημα με το κριτήριο αν μπορεί η βάση να ικανοποιήσει τους περιορισμούς ρυθμαπόδοσης που έχει θέσει. Η αδυναμία ικανοποίησης των περιορισμών έγκειται στο πλήθος των ήδη εισαχθέντων χρηστών και στις φυσικές αδυναμίες του σταθμού βάσης (για παράδειγμα η μέγιστη συνολική αποδιδόμενη ισχύς από την κεραία). Μια εκτενής και μαθηματικά αποδεδειγμένη θεωρία που παρουσιάζει τους περιορισμούς ικανοποιησιμότητας είναι μέσα στα θέματα επέκτασης της παρούσας διπλωματικής. Επίσης ένα δεύτερο σημαντικό θέμα είναι η επίλυση του ίδιου προβλήματος εκχώρησης πόρων για την ικανοποίηση των βραχυπρόθεσμων περιορισμών ρυθμαπόδοσης κατά την ζεύξη ανόδου. Τα προβλήματα που σχετίζονται με την ζεύξη ανόδου και καθόδου αν και είναι δυαδικά μεταξύ τους έχουν εντελώς διαφορετική προσέγγιση λύσης γιατί στο μεν πρώτο οι περιορισμοί του συστήματος έρχονται κατανεμημένοι από τους περιορισμούς απόδοσης πόρων από κάθε χρήστη ενώ στο δεύτερο οι περιορισμοί έρχονται από τον σταθμό βάσης. Άλλοι τομείς επέκτασης της εργασίας μας θα μπορούσαν να είναι η δυναμική μετακίνηση χρηστών στο δίκτυο, (έχουμε υποθέσει ότι οι χρήστες στο δίκτυο έχουν εισαχθεί και δεν απομακρύνονται από

το δίκτυο ούτε εισάγονται νέοι χρήστες) όπως επίσης και το η άρση της υπόθεσης ότι κάθε χρήστης στο δίκτυο διαθέτει σε κάθε στιγμή δεδομένα προς αποστολή. Αυτά είναι πεδία έρευνας που πολύ θα θέλαμε να ερευνηθούν στο άμεσο μέλλον.

6 Βιβλιογραφία

- [1] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and S. Viterbi, "CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users," *Communications Magazine, IEEE* , vol.38, no.7pp.70-77, Jul 2000., pp. 1-10.
- [2] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMAHDR a high efficiency-high data rate personal communication wireless system," in *Proc. VTC*, pp. 1854--1858, 2000
- [3] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Transactions on Wireless Communications* vol.3, no.4pp. 1250- 1259, July 2004.
- [4] X. Liu, E.K.P. Chong and N.B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol.19, no.10pp.2053-2064, Oct 2001.
- [5] X. Liu, E. Chong, and N. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, pp. 451–474, March 2003.
- [6] F. Berggren and R. Jantti, "Asymptotically fair transmission scheduling over fading channels," *IEEE Transactions on Wireless Communications*, vol.3, no.1pp. 326- 336, Jan. 2004.
- [7] Li Chuxiang, and Wang Xiaodong, "Adaptive opportunistic fair scheduling over multiuser spatial channels," *Communications, IEEE Transactions on* , vol.53, no.10pp. 1708- 1717, Oct. 2005
- [8] Chengzhou Li and Symeon Papavassiliou, "Joint Throughput Maximization and Fair Scheduling in Uplink DS-CDMA Systems", in *Proc. 2004 IEEE Sarnoff Symposium on Advances in Wired and Wireless Communications*, pp. 193-196, April 2004.

- [9] T. Kastrinogiannis and S. Papavassiliou, "Probabilistic Short-term Delay and Throughput Requirements of Multimedia Services in High Throughput Wireless Networks," IEEE Sarnoff Symposium 2007.
- [10] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Joint resource allocation and base-station assignment for the downlink in CDMA networks," IEEE/ACM Trans. Netw., vol. 14, no. 1, Feb. 2006.
- [11] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Downlink power allocation for multi-class wireless systems," IEEE/ACM Trans. Netw., vol. 13, no. 4, pp. 854-867, Aug. 2005.
- [12] T. Kastrinogiannis, S. Papavassiliou, K. Kastrinogiannis and D. Soulios, "A Utility-based Resource Allocation Approach for the Downlink in CDMA Wireless Networks with Multimedia Services" to appear in IEEE International Symposium on Personal, Indoor and Mobile Radio Communications September 2007.