ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Ενισχυτική Μάθηση σε Συστήματα Γνωστικών Ραδιοεπικοινωνιών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωστόπουλος Γ. Παναγιώτης

Επιβλέπων : Παναγιώτης Κωττής

Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2011

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Ενισχυτική Μάθηση σε Συστήματα Γνωστικών Ραδιοεπικοινωνιών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωστόπουλος Γ. Παναγιώτης

Επιβλέπων : Παναγιώτης Κωττής

Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 15-07-2011

.....	.....	.....
Π. Κωττής	Χ. Καψάλης	Γ.Φικιώρης
Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π	Επ. Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2011

.....

Κωστόπουλος Παναγιώτης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Στο πλαίσιο αυτής της διπλωματικής εργασίας θα ήθελα να εκφράσω τις πιο θερμές μου ευχαριστίες στον επιβλέποντα Καθηγητή Π. Κωττή για την πολύτιμη συμβολή του και την άριστη συνεργασία μας, καθώς και την Υποψήφια Διδάκτορα Αγγελική Κορδαλή για τις σημαντικές υποδείξεις και την ενδελεχή βοήθειά της κατά την εκπόνηση της διατριβής. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και το φιλικό μου περιβάλλον για την αμέριστη συμπαράστασή τους.

ΠΕΡΙΛΗΨΗ

Η αλματώδης αύξηση της ζήτησης για υπηρεσίες ασύρματης επικοινωνίας έχει οδηγήσει στην ελαχιστοποίηση του διαθέσιμου φάσματος συχνοτήτων, καθιστώντας το σπάνιο φυσικό πόρο. Συνεπώς, η διαχείριση και εκχώρηση φάσματος αποτελεί αντικείμενο με ιδιαίτερο οικονομικό και επιστημονικό αντίκτυπο. Σύμφωνα με πολλές σχετικές μελέτες, η χρησιμοποίηση του φάσματος δεν είναι αποδοτική εξαιτίας της στατικής εκχώρησής του αποκλειστικά σε αδειοδοτημένους χρήστες (πρωτεύοντες χρήστες). Ως επακόλουθο των στατικών πολιτικών πρόσβασης και της διαδικασίας αδειοδότησης φάσματος υπάρχει σημαντικό πλεόνασμα ανεκμετάλλευτων συχνοτήτων και κρίνεται επιτακτική η αλλαγή στην πολιτική εκχώρησης φάσματος. Προς το σκοπό αυτό έχει προταθεί η δυναμική πρόσβαση στο φάσμα η οποία γίνεται δυνατή μέσω των γνωστικών ραδιοεπικοινωνιών.

Σκοπός της εργασίας είναι η διερεύνηση της εφαρμογής τεχνικών ενισχυτικής μάθησης στη διαδικασία ανίχνευσης φάσματος από δευτερεύοντες χρήστες γνωστικών ραδιοεπικοινωνιών. Αρχικά, γίνεται επισκόπηση του αντικειμένου των γνωστικών ραδιοεπικοινωνιών και παρουσιάζεται το πρόβλημα της ενισχυτικής μάθησης, δίνοντας έμφαση στις τεχνικές που στηρίζονται σε διαδικασίες απόφασης Markov. Στη συνέχεια, ακολουθεί η παρουσίαση και τα συμπεράσματα από εφαρμογές ήδη υπαρχόντων μοντέλων ενισχυτικής μάθησης. Τέλος παρουσιάζονται τα εργαστηριακά αποτελέσματα της προσομοίωσης ενός σχήματος ενισχυτικής μάθησης για αποδοτικότερη πρόσβαση στο φάσμα με χρήση του εργαλείου MATLAB.

Λέξεις κλειδιά

Γνωστικές ραδιοεπικοινωνίες, δυναμική εκχώρηση φάσματος, ανίχνευση φάσματος, ενισχυτική μάθηση, adaptive load balancing.

ABSTRACT

The growth of demand for wireless communication services has led to the minimazation of the available frequency spectrum, rendering it a scarce natural resource. Consequently, the spectrum management and assignment has become a very important matter with severe financial and scientific impact. According to a lot of relevant researches, spectrum utilization is not efficient because of the static spectrum allocation only to primary users. As a result of static spectrum access, there is a lot of unexploited frequency bands, in this way the change of the spectrum access policy is necessary. As a solution to the problem, dynamic spectrum access was proposed to allow non-licensed users to access the spectrum opportunistically. This kind of access is achieved through the use of cognitive radios.

This thesis aims at examining the application of reinforcement learning techniques in the framework of spectrum sensing from secondary cognitive radio users. First, the principles of cognitive radio and reinforcement learning are presented extensively, focusing on Markov Decision Processes. Then, a specific application of reinforcement learning techniques is simulated and the corresponding results are presented.

Keywords

Cognitive radios, dynamic spectrum access, spectrum sensing, reinforcement learning, adaptive load balancing.

Περιεχόμενα

Κεφάλαιο 1: Γνωστικές ραδιοεπικοινωνίες 13

1.1. Χρησιμοποίηση του φάσματος ραδιοσυχνοτήτων 14
1.2. Δυναμική εκχώρηση φάσματος 18
1.2.1. Μοντέλο αποκλειστικης δυναμικής χρήσης (Dynamic exclusive use model) 18
1.2.2. Μοντέλο ανοικτης ανταλλαγής (Open sharing model) 19
1.2.3. Μοντέλο ιεραρχικής πρόσβασης (Hierarchical access model) 19
1.3. Γνωστικά συστήματα ραδιοεπικοινωνιών (Cognitive radios) 21
1.4. Αρχές λειτουργίας και επίπεδα πρωτοκόλλων σε συστήματα γνωστικών ραδιοεπικοινωνιών 24
1.5. Βασικές λειτουργίες συστημάτων γνωστικών ραδιοεπικοινωνιών 27
1.5.1. Ανίχνευση φάσματος (Spectrum sensing) 27
1.5.1.1. Ανίχνευση πομπού (Non-cooperative sensing) 28
1.5.1.2. Συνεργατική ανίχνευση (Cooperative sensing) 31
1.5.1.3. Ανίχνευση παρεμβολής 33
1.5.2. Διαχείριση φάσματος (Spectrum management) 35
1.5.3. Κινητικότητα φάσματος (Spectrum mobility) 36
1.6. Πρόσφατα ασύρματα πρότυπα 37
1.6.1. Πρότυπο 802.11 37
1.6.2. IEEE 802.11k 38
1.6.3. Bluetooth 39
1.6.4. Τεχνολογία IEEE και νέες εφαρμογές γνωστικών ραδιοεπικοινωνιών 39

Κεφάλαιο 2: Ενισχυτική Μάθηση 41

2.1. Προέλευση του όρου-ορισμός 41
2.1.1. Διαφορά ενισχυτικής μάθησης και επιβλεπόμενης μάθησης 42
2.1.2. Διαφορά ενισχυτικής μάθησης και δυναμικού προγραμματισμού 42
2.1.3. Το πρόβλημα των k-κουλοχέρηδων 43
2.2. Το μοντέλο της ενισχυτικής μάθησης 44
2.3. Στοιχεία ενισχυτικής μάθησης 46
2.4. Διαδικασία απόφασης Markov (MDP) 48
2.4.1. Εύρεση πολιτικής για δεδομένο μοντέλο Markov 49
2.4.2. Επανάληψη αξιών (Value iteration) 50
2.4.3. Επανάληψη πολιτικών (Policy iteration) 51
2.4.4. Μαθαίνοντας μια βέλτιστη πολιτική 52
2.4.5. Μέθοδοι βασισμένες σε κάποιο μοντέλο 53
2.4.6. Μέθοδοι χωρίς τη χρήση μοντέλου 53
2.4.7. Εισαγωγή στη μάθηση Q 56
2.4.8. Ο αλγόριθμος μάθησης Q 57
2.5. Ανίνευση πηγών φάσματος με εφαρμογή αλγορίθμου ενισχυτικής μάθησης Q 59
2.5.1. Περιγραφή του συστήματος 59

2.5.2. Σύστημα Ανίχνευσης Φάσματος 60
2.5.2.1. Διατύπωση προβλήματος ενισχυτικής μάθησης 62
2.5.2.2. Στρατηγική επίλυσης 65
2.5.3. Αποτελέσματα προσομοίωσης 68
2.5.4. Επίλογος 70

Κεφάλαιο 3: Εφαρμογές ενισχυτικής μάθησης 71

3.1. Σχέδιο κατανεμημένης εκχώρησης φάσματος (Distributed spectrum sharing scheme) 71
3.1.1. Συνάρτηση ανταμοιβής 71
3.1.2. Βήματα αλγορίθμου 72
3.1.3. Αποτελέσματα προσομοίωσης 75
3.1.4. Συμπεράσματα 79
3.2. Ενισχυτική μάθηση συνεργασίας πολλών πρακτόρων (Multi-agent reinforcement learning) 80
3.2.1. Ενισχυτική μάθηση συνεργασίας πολλών πρακτόρων για ισορροπημένη κατνομή υπηρεσιών (Adaptive load balancing-Multi – agent learning) 80
3.2.2. Η γενική θέση του συστήματος 80
3.2.3. Προσαρμοσμένοι κανόνες επιλογής πηγών (Selections rules) 81
3.2.4. Ετερογενείς πληθυσμοί πρακτόρων (Heterogenous populations) 82
3.2.5. Επικοινωνία μεταξύ των πρακτόρων (Communication among agents) 83
3.2.6. Συμπεράσματα 85
3.3. Ετερογενή τηλεπικοινωνιακά επίγεια συστήματα πολυεκπομπής 86
3.3.1. (Heterogeneous multicast terrestrial communication systems) 86
3.3.2. Το μοντέλο του σχήματος επιλογής καναλιών από τους επίγειους σταθμούς βάσης 86
3.3.3. Συμπεράσματα 92
3.4. Σύνοψη 92

Κεφάλαιο 4: Προσομοιωση 93

4.1. Μοντέλο χρήσης του υπάρχοντος φάσματος 93
4.2. Ανίχνευση και πρόσβαση στο φάσμα 93
4.3. Μελέτη συμπεριφοράς σε διαφορετικού τύπου μηνύματα 96
4.4. Αποτελέσματα προσομοίωσης 98
4.4.1. Σύκριση σταθερής κατανομής-απότομων μεταβολών 98
4.4.2. Πρόοδος ενισχυτικής μάθσησης 100
4.4.3. Συνδυασμός τυχαίων μεταβολών και σταθερής κατανομής πιθανοτήτων 102
4.5. Συμπεράσματα 105

ΚΕΦΑΛΑΙΟ 1^{ο}: ΓΝΩΣΤΙΚΕΣ ΡΑΔΙΟΕΠΙΚΟΙΝΩΝΙΕΣ

Οι γνωστικές ραδιοεπικοινωνίες (Cognitive radios) αποτελούν ένα νέο τρόπο φασματικής πρόσβασης, ικανό να προσφέρει την απαιτούμενη αποδοτικότητα, επιτυχία και αξιοπιστία στη σχεδίαση των τηλεπικοινωνιακών συστημάτων. Ο όρος Cognitive radios ορίστητκε το 1999 από τον Mitola [1]. Οι τεχνολογίες αυτού του είδους χρησιμοποιούνται στα τηλεπικοινωνιακά συστήματα που βασίζονται στην εκχώρηση φάσματος και αναμένεται να επικρατήσουν στον τηλεπικοινωνιακό χώρο στο άμεσο μέλλον εξαιτίας της αποδοτικότητας και της αξιοπιστίας που προσφέρουν. Υπόσχονται δυνατότητα φασματικής χρήσης στις διαστάσεις του χώρου του χρόνου και της συχνότητας που έως τώρα δεν ήταν διαθέσιμες. Δηλαδή για τη χρησιμοποίηση του φάσματος θα αξιοποιούνται στο έπακρο όλοι οι διαθέσιμοι πόροι.

1.1 Χρησιμοποίηση του φάσματος ραδιοσυχνοτήτων

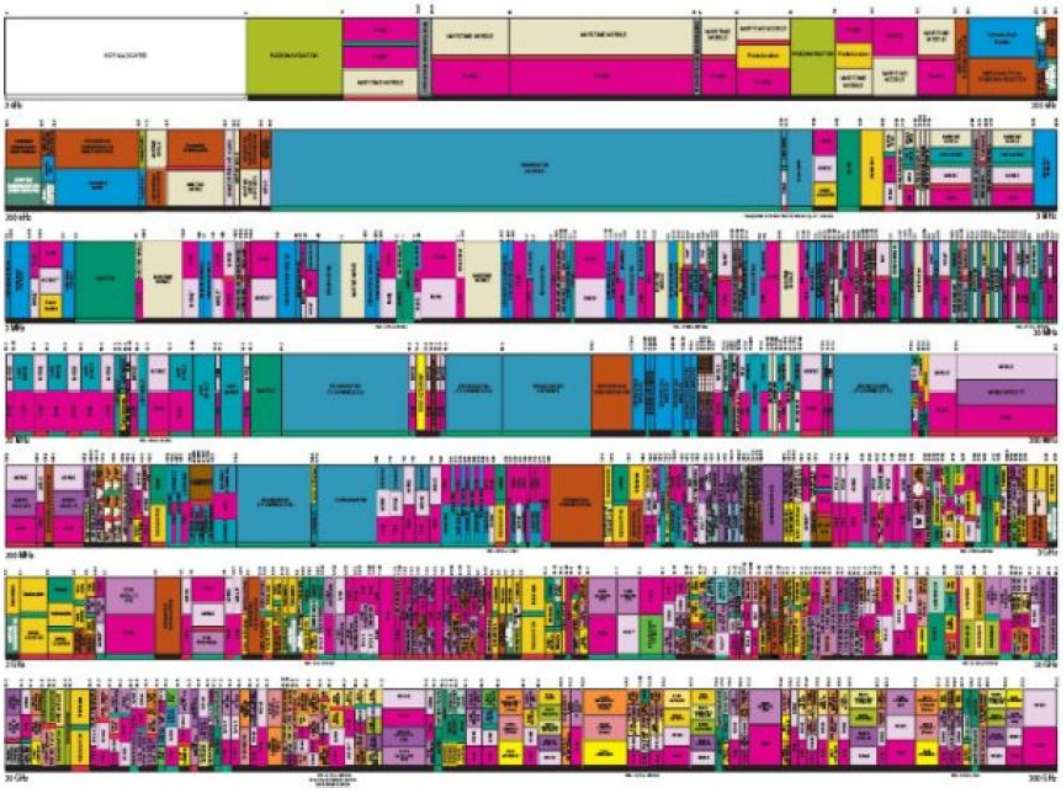
Το φάσμα ραδιοσυχνοτήτων αποτελεί σπάνιο φυσικό πόρο, απαραίτητο για τη λειτουργία των σύγχρονων Ηλεκτρονικών Επικοινωνιών αλλά και υπηρεσιών υψίστης σημασίας όπως η αεροναυτιλία και οι υπηρεσίες έκτακτης ανάγκης. Μια σειρά διεθνών οργάνων και ομάδων προτυποποίησης μεταξύ των οποίων συμπεριλαμβάνονται οι:

- Διεθνής Ένωση Τηλεπικοινωνιών (ITU)
- Ευρωπαϊκή Διάσκεψη των Διοικήσεων Ταχυδρομείων και Τηλεπικοινωνιών (CEPT)
- Ευρωπαϊκό Ινστιτούτο Τηλεπικοινωνιακών Προτύπων (ETSI)
- Διεθνής Ειδική Επιτροπή Ηλεκτρομαγνητικών Παρεμβολών (CISPR)

είναι υπεύθυνα για τους κανονισμούς και τις ρυθμιστικές συμφωνίες που αφορούν τη διαχείριση του φάσματος ραδιοσυχνοτήτων. Οι ομάδες αυτές έχουν κατανείμει στατικά το φάσμα ραδιοσυχνοτήτων στους κατωτέρω τρεις τύπους φασματικών περιοχών:

- Περιοχές όπου κανένας δεν μπορεί να εκπέμψει όπως, για παράδειγμα, οι συχνότητες ραδιοαστρονομίας
- Περιοχές όπου οποιοσδήποτε μπορεί να εκπέμψει, όπως για παράδειγμα οι ISM (Industrial, Scientific and Medical radio band) και ultra-wideband ζώνες καθώς και οι ζώνες οι οποίες έχουν παραχωρηθεί σε ραδιοερασιτέχνες.
- Περιοχές όπου μόνο αδειοδοτημένοι χρήστες μπορούν να εκπέμψουν όπως, για παράδειγμα, οι φασματικές ζώνες υπηρεσιών τηλεόρασης και κινητής τηλεφωνίας.

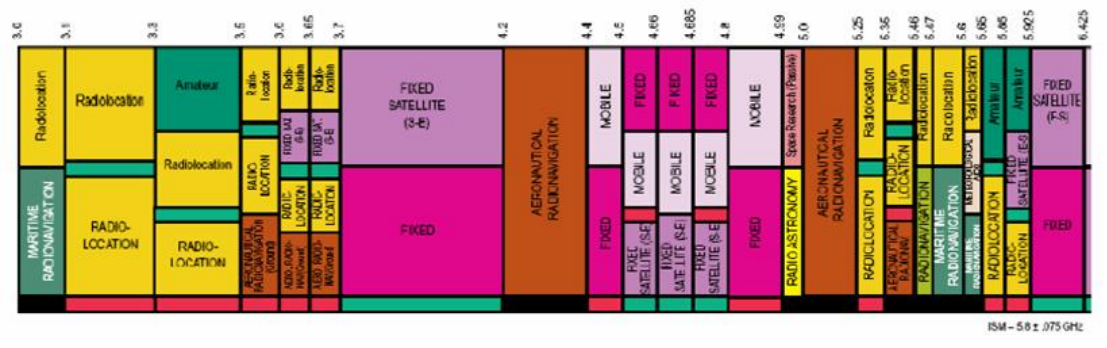
Κατανέμοντας το φάσμα στατικά, δηλαδή αδειοδοτώντας ορισμένες ζώνες συχνοτήτων για κάθε ασύρματη υπηρεσία εξασφαλίζεται με απλό τρόπο ο περιορισμός των παρεμβολών και η εύρυθμη συνύπαρξη διαφορετικών ασύρματων εφαρμογών. Ένα παράδειγμα στατικής κατανομής φαίνεται στο Σχ.1.1 όπου απεικονίζεται ο χάρτης φασματικής κατανομής των Ηνωμένων Πολιτειών Αμερικής.



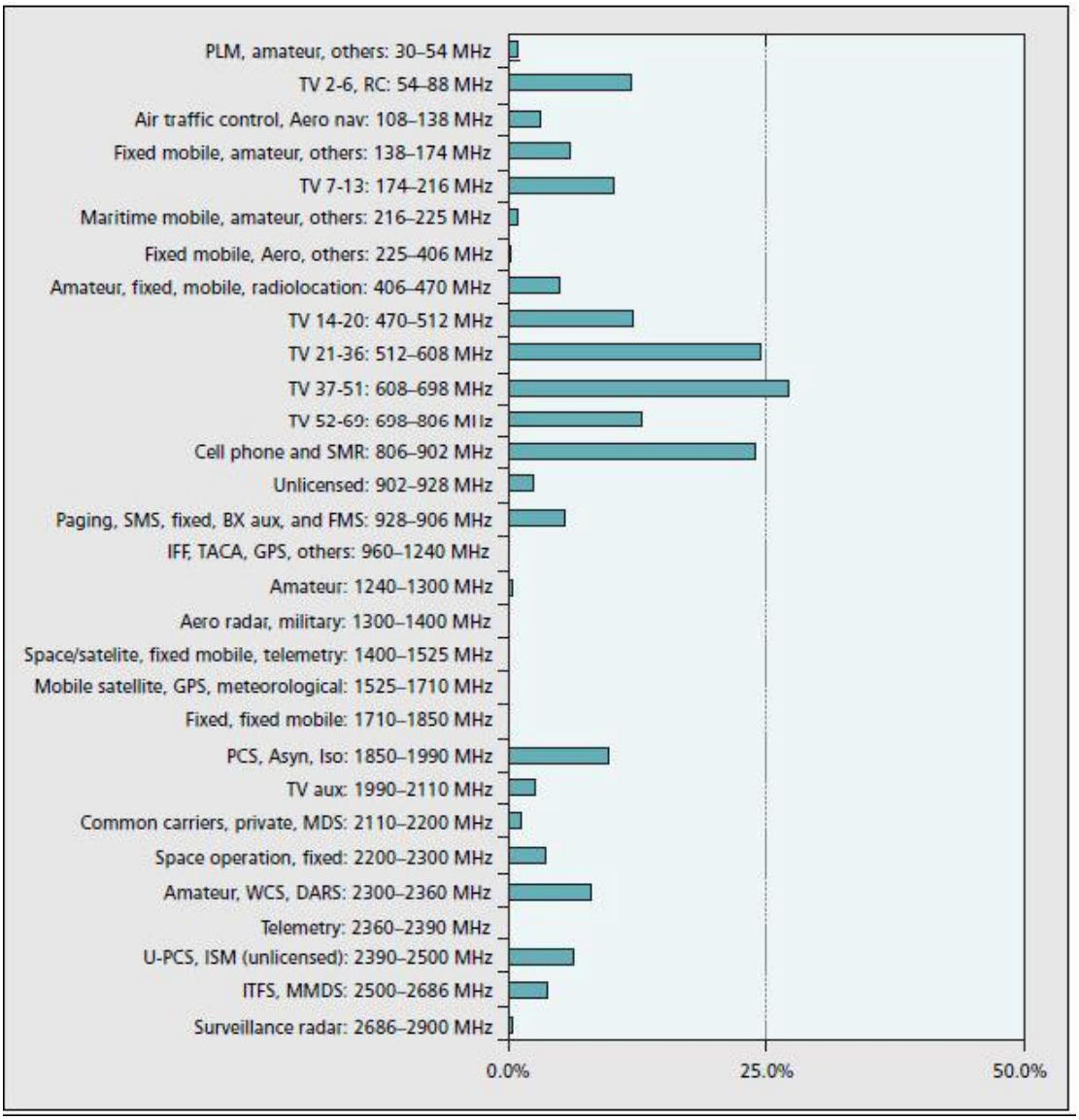
Σχήμα 1.1: Χάρτης φασματικής κατανομής των Ηνωμένων Πολιτειών Αμερικής

Ωστόσο, τις τελευταίες δεκαετίες, με τον πολλαπλασιασμό των τεχνολογιών ασύρματης επικοινωνίας, το μεγαλύτερο ποσοστό του ραδιοφάσματος έχει εκχωρηθεί. Αυτό έχει ως αποτέλεσμα την έλλειψη φάσματος, η οποία προκάλεί σημαντικό πρόβλημα για την μελλοντική ανάπτυξη της βιομηχανίας ασύρματων επικοινωνιών.

Από την άλλη πλευρά, προσεκτικές μελέτες του τρόπου χρησιμοποίησης του διαθέσιμου φάσματος αποκαλύπτουν ότι οι περισσότερες κατανεμημένες ζώνες (κανάλια) του ραδιοφάσματος έχουν χαμηλή χρησιμοποίηση. Πρόσφατες μετρήσεις της Ομοσπονδιακής Επιτροπής Τηλεπικοινωνιών των ΗΠΑ (FCC Federal Communications Commission) αποκαλύπτουν ότι το 70% από το κατανεμημένο φάσμα στις ΗΠΑ δεν χρησιμοποιείται καθόλου [2]. Αξιοσημείωτο είναι ότι αυτή η χαμηλή χρησιμοποίηση του φάσματος δεν είναι συνεπής με το διάγραμμα εκχώρησης συχνοτήτων της FCC το οποίο απεικονίζεται στο Σχ.1.2 από το οποίο προκύπτει ότι υπάρχουν πολλαπλές εκχωρήσεις σε όλες τις ζώνες συχνοτήτων για πληθώρα ασύρματων υπηρεσιών.



Σχήμα 1.2: Εκχώρηση συχνοτήτων σε ασύρματες υπηρεσίες στη ζώνη 3-6 GHz (FCC)

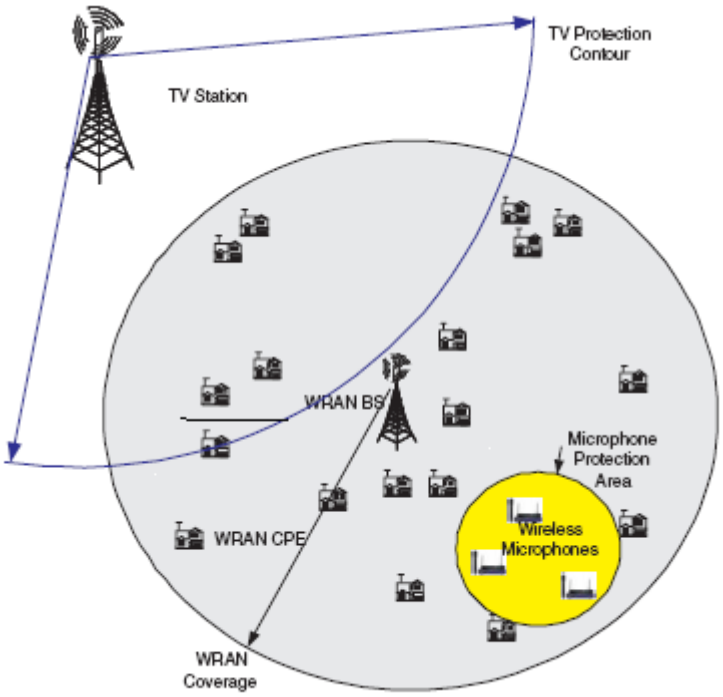


Σχήμα 1.3: Μέσος όρος απασχόλησης φάσματος σε έξι γεωγραφικές περιοχές που μελετήθηκαν για κάθε ζώνη συχνοτήτων χωριστά

Ως λύση στο προαναφερθέν πρόβλημα της χαμηλής φασματικής χρησιμοποίησης προτείνεται η εφαρμογή της αναχρησιμοποίησης συχνότητας. Μια νέα ευφυή εκδοχή της αναχρησιμοποποίησης συχνότητας αποτελούν οι γνωστικές ραδιοεπικοινωνίες, με θεμελιώδες συστατικό την ανίχνευση φάσματος (channel sensing). Με βάση αυτήν οι ασύρματες συσκευές θα μπορούν να αποκτούν γνώση του περιβάλλοντος διάδοσης εντός του οποίου έχουν λειτουργική εμβέλεια ώστε να ανιχνεύουν ζώνες συχνοτήτων οι οποίες δεν είναι κατειλημένες από πρωτεύοντες χρήστες (primary users). Το Δεκέμβριο του 2003 η FCC εξέδωσε ανακοίνωση από προτεινόμενους κανόνες οι οποίοι ανγνωρίζουν τις γνωστικές ραδιοεπικοινωνίες ως το βασικό υποψήφιο εργαλείο διαπραγματεύσιμης κατανομής του ραδιοφάσματος [2]. Σε απόκριση της ανακοίνωσης αυτής η IEEE έχει συγκροτήσει το 802.22 Working Group με σκοπό να αναπτύξει προδιαγραφές για

WRAN (Wireless Region Area Networks) βασιζόμενη στην τεχνολογία γνωστικών ραδιοεπικοινωνιών.

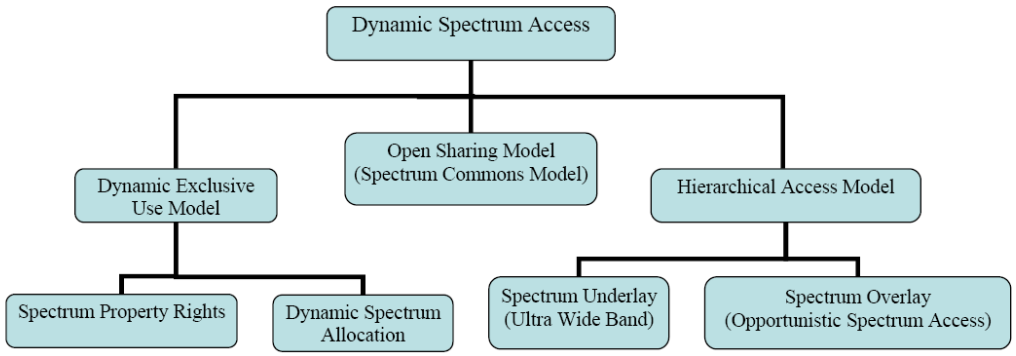
Το Σχ.1.4 που ακολουθεί απεικονίζει την τοπολογία ενός WRAN συστήματος στο οποίο φαίνονται οι πρωτεύοντες χρήστες (τηλεοπτικοί σταθμοί και ασύρματα μικρόφωνα) και οι δευτερεύοντες χρηστες (secondary users) που περιλαμβάνουν σταθμούς βάσης WRAN και WRAN CPEs (εξοπλισμός πελατών). Τα WRAN συστήματα είναι σχεδιασμένα να παρέχουν ασύρματη ευρυζωνική πρόσβαση σε αγροτικές και προαστιακές περιοχές με μέση ακτίνα κάλυψης 33 km. Τα WRAN συστήματα πρέπει να είναι ικανά να κάνουν ανίχνευση του ραδιοφάσματος ώστε να αναγνωρίζουν τα αχρησιμοποίητα από τους πρωτεύοντες χρήστες τηλεοπτικά κανάλια. Το θεμελιώδες αντικείμενο των WRAN συστημάτων είναι η μεγιστοποίηση της χρήσης των τηλεοπτικών καναλιών από δευτερεύοντες χρήστες όταν αυτά δεν χρησιμοποιούνται από τους πρωτεύοντες χρήστες.



Σχήμα 1.4: Ένα απλό γενικό IEEE 802.22 WRAN σενάριο

1.2 Δυναμική εκχώρηση φάσματος – Dynamic spectrum access (DSA)

Σε αντίθεση με τη μέχρι τώρα πολιτική στατικής εκχώρησης φάσματος, ο όρος δυναμική εκχώρηση φάσματος ο οποίος χρησιμοποιείται ως συνώνυμο του όρου γνωστικές ραδιοεπικοινωνίες περιλαμβάνει διάφορες προσεγγίσεις στον επανακαθορισμό της κατανομής του ραδιοφάσματος (φασματική μεταρρύθμιση). Οι βασικές στρατηγικές δυναμικής πρόσβασης φάσματος μπορούν να κατηγοριοποιηθούν στα ακόλουθα τρια μοντέλα που απεικονίζονται στο Σχ. 1.5



Σχημα 1.5: Στρατηγικές δυναμικής πρόσβασης στο φάσμα

1.2.1 Μοντέλο αποκλειστικής δυναμικής χρήσης (Dynamic exclusive use model)

Διατηρεί τη βασική δομή της υφιστάμενης πολιτικής ρύθμισης του φάσματος. Ζώνες συχνοτήτων αδειοδοτούνται σε υπηρεσίες για αποκλειστική χρήση. Η βασική ιδέα είναι η εισαγωγή ευελιξίας προς βελτίωση της φασματικής απόδοσης. Δύο προσεγγίσεις έχουν προταθεί σύμφωνα με το μοντέλο αποκλειστικής δυναμικής χρήσης: Η πρώτη προσέγγιση επιτρέπει στους δικαιούχους (licensees-primary users) να εμπορεύονται το φάσμα και να επιλέγουν ελεύθερα την τεχνολογία πρόσβασης σε αυτό. Η οικονομία και η αγορά, επομένως, θα διαδραματίσουν ένα σημαντικό ρόλο στην αποδοτικότερη χρησιμοποίηση του περιορισμένου αυτού πόρου.

Η δεύτερη προσέγγιση της δυναμικής κατανομής φάσματος διατυπώθηκε από το ευρωπαϊκό σχέδιο DRIVE. Στόχος της είναι η βελτίωση της αποτελεσματικής χρησιμοποίησης του φάσματος μέσω της δυναμικής εκχώρησης, αξιοποιώντας τα χωρικά και χρονικά στατιστικά κίνησης διαφόρων υπηρεσιών. Δηλαδή, σε συγκεκριμένη περιοχή και για συγκεκριμένο χρονικό διάστημα, το φάσμα διατίθεται σε υπηρεσίες για αποκλειστική χρήση. Ο τρόπος αυτός εκχώρησης διαφέρει από την τρέχουσα πολιτική καθόσον επιτυγχάνει ταχύτερα τις αλλαγές κατανομής συχνοτήτων όπου και όταν απαιτούνται.

Ωστόσο, με βάση μόνο το μοντέλο αποκλειστικής χρήσης δεν είναι δυνατό να αντιμετωπιστούν και να αξιοποιηθούν τα λευκά αχρησιμοποίητα φασματικά

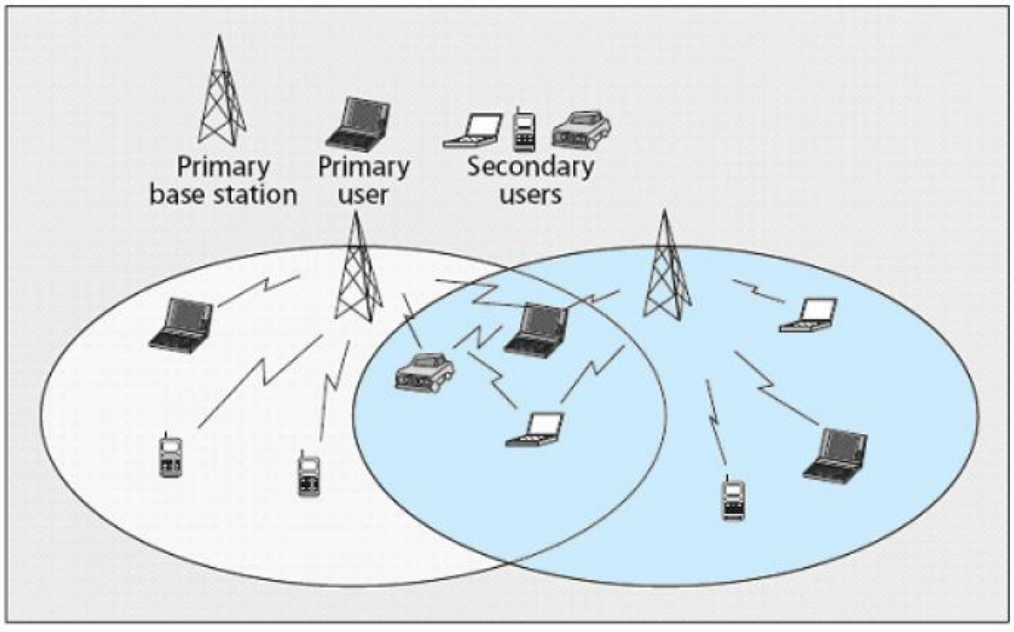
διαστήματα (white spaces) τα οποία προκύπτουν από την εκρηκτική (bursty) φύση της ασύρματης τηλεπικοινωνιακής κίνησης.

1.2.2 Μοντέλο ανοικτής ανταλλαγής (Open sharing model)

Αναφέρεται και ως μοντέλο κοινού φάσματος και χρησιμοποιεί την ανοικτή ανταλλαγή φάσματος μεταξύ ομότιμων χρηστών ως βάση για τη διαχείριση μιας φασματικής περιοχής. Οι υποστηρικτές του μοντέλου αυτού τάσσονται υπέρ της χρησιμοποίησής του εξαιτίας της αδιαμφισβήτητης επιτυχίας των ασύρματων υπηρεσιών (πχ WiFi) οι οποίες λειτουργούν στη μη αδειοδοτημένη βιομηχανική, επιστημονική και ιατρική ζώνη ραδιοσυχνοτήτων (ISM-Industrial, Scientific and Medical radio bands). Στο πλαίσιο του συγκεκριμένου μοντέλου διαχείρισης ραδιοφάσματος για την αντιμετώπιση των τεχνολογικών προκλήσεων, έχουν διερευνηθεί τόσο συγκεντρωτικές όσο και κατανεμημένες στρατηγικές κατανομής φάσματος.

1.2.3 Μοντέλο ιεραρχικής πρόσβασης (Hierarchical access model)

Το μοντέλο αυτό υιοθετεί μια ιεραρχική δομή πρόσβασης με πρωτεύοντες και δευτερεύοντες χρήστες.



Σχήμα 1.6: Απεικόνιση δικτύων τα οποία ακολουθούν το μοντέλο ιεραρχικής πρόσβασης

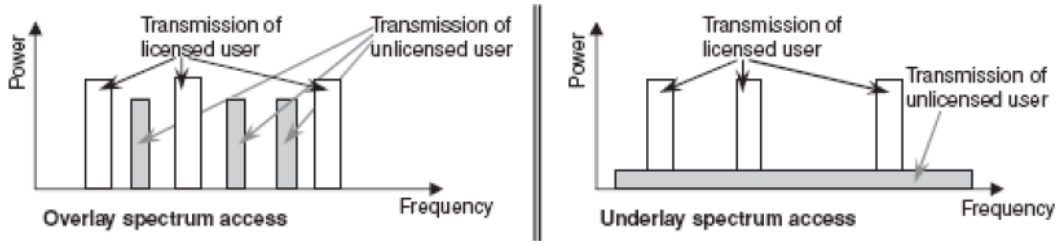
Η βασική ιδέα είναι η δυνατότητα χρησιμοποίησης αδειοδοτημένου φάσματος από δευτερεύοντες χρήστες εφόσον περιορίζονται επαρκώς οι παρεμβολές τις

οποίες αντιλαμβάνονται οι πρωτεύοντες χρήστες (βλ Σχ.1.6). Δύο προσεγγίσεις καταμερισμού του φάσματος μεταξύ πρωτευόντων και δευτερευόντων χρηστών έχουν εξεταστει: η τεχνική φασματικής υπόστρωσης (spectrum underlay) και η τεχνική φασματικής επίστρωσης (spectrum overlay).

Η *προσέγγιση φασματικού υποστρώματος* επιβάλλει αυστηρούς περιορισμούς σχετικά με τη ισχύ εκπομπής των δευτερευόντων χρηστών ώστε να λειτουργούν κάτω από το επίπεδο θορύβου των πρωτευόντων χρηστών. Μέσω της φασματικής εξάπλωσης των σημάτων σε μια ευρύτερη ζώνη ραδιοσυχνοτήτων (UWB), οι δευτερεύοντες χρήστες έχουν τη δυνατότητα να επιτύχουν υψηλούς ρυθμούς μετάδοσης δεδομένων με εξαιρετικά χαμηλή ισχύ εκπομπής.

Η *τεχνική της φασματικής επίστρωσης* αρχικά επινοήθηκε από το Mitola [4] και έγινε γνωστή υπό τον όρο συγκέντρωση φάσματος (spectrum pooling) και, στη συνέχεια, ερευνήθηκε από τον οργανισμό προηγμένων ερευνητικών έργων του υπουργείου άμυνας των Ηνωμένων Πολιτειών (DARPA) στο πλαίσιο του προγράμματος επόμενης γενιάς (XG) με την ονομασία της ευκαιριακής φασματικής πρόσβασης (opportunistic spectrum access). Η διαφορά από την προηγούμενη προσέγγιση είναι ότι δεν υποβάλλει αναγκαστικά αυστηρούς περιορισμούς στη ισχύ εκπομπής των δευτερευόντων χρηστών αλλά καθορίζει πότε και πού μπορούν να εκπέμψουν. Η προσέγγιση αυτή αποσκοπεί να αξιοποιήσει άμεσα τα χρονικά και χωρικά φασματικά κενά, επιτρέποντας στους δευτερεύοντες χρήστες να αναγνωρίζουν και να αξιοποιούν την περιστασιακή φασματική διαθεσιμότητα σε τοπική βάση αρκεί η δραστηριότητα αυτή να πραγματοποιείται με μη παρεμβατικό τρόπο. Δηλαδή σε καμία περίπτωση δεν πρέπει η δραστηριότητα που αναπτύσσουν οι δευτερεύοντες χρήστες να επηρεάζει την ποιότητα υπηρεσίας των πρωτευόντων χρηστών.

Συγκρινόμενο με τα μοντέλα δυναμικής αποκλειστικής χρήσης και ανοικτής ανταλλαγής, το μοντέλο ιεραρχικής πρόσβασης είναι ίσως η πλέον συμβατή προσέγγιση εφαρμογής της δυναμικής εκχώρησης συχνοτήτων, δεδομένης της τρέχουσας πολιτικής διαχείρισης του ραδιοφάσματος και της κληρονομιάς των ασύρματων συστημάτων. Επιπλέον, οι δύο διαφορετικές τεχνικές ιεραρχικής πρόσβασης ίσως στο μέλλον να έχουν τη δυνατότητα να εφαρμοστούν συνδυαστικά για την περαιτέρω βελτίωση της φασματικής απόδοσης.



Σχήμα 1.7: Απεικόνιση φασματικής επίστρωσης και φασματικής υπόστρωσης

1.3 Γνωστικά συστήματα ραδιοεπικοινωνιών (Cognitive radios)

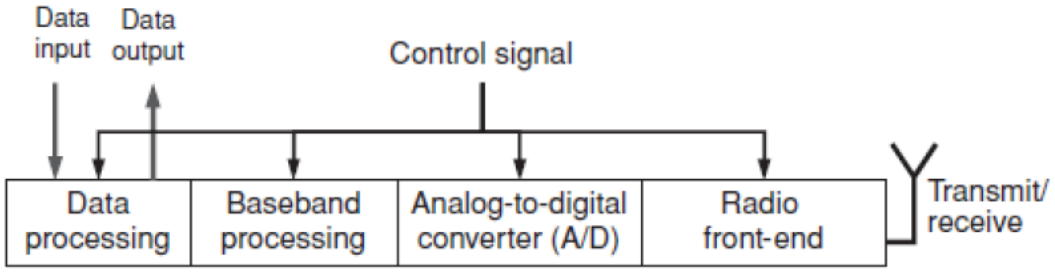
Αν και ο όρος γνωστικές ραδιοεπικοινωνίες προέκυψε σχετικά πρόσφατα, η ιδέα για κατασκευή έξυπνων συστημάτων ραδιοεπικοινωνιών δεν είναι νέα. Ήδη από τη δεκαετία του 1980 είχαν αναπτυχθεί δέκτες που διέθεταν αρκετά χαρακτηριστικά ευφυίας, όπως για παράδειγμα η αυτόματη αναγνώριση του είδους διαμόρφωσης των σημάτων λήψης.

Η ιδέα των συστημάτων γνωστικών ραδιοεπικοινωνιών στηρίζεται στις αρχές του Software Defined Radio (SDR). Οι συσκευές που λειτουργούν με βάση το SDR αναζητούν και χρησιμοποιούν δυναμικά διαθέσιμες φασματικές ζώνες οι οποίες καθορίζονται από λογισμικό. Το SDR βασίζεται ουσιαστικά σε πομπούς των οποίων οι παράμετροι λειτουργίας (συχνότητα λειτουργίας, ισχύς εκπομπής, διαμόρφωση, κωδικοποίηση, διάγραμμα ακτινοβολίας) μπορούν να μεταβάλλονται κατά προγραμματιζόμενο τρόπο χωρίς την αλλαγή του hardware του πομπού.

Το κύριο χαρακτηριστικό μιας SDR συσκευής είναι η ικανότητά της να επαναπροσαρμόζεται-επαναπρογραμματίζεται (reconfigurability). Ο όρος αυτός περιλαμβάνει:

- Την προσαρμογή της διεπαφής του ραδιοσυστήματος στις μεταβολές του περιβάλλοντος
- Την ενσωμάτωση νέων υπηρεσιών και εφαρμογών όπως για παράδειγμα υπηρεσίες κινητής τηλεφωνίας ή ασύρματου ευρυζωνικού Internet
- Την ενσωμάτωση των τελευταίων εξελίξεων στην τεχνολογία λογισμικού
- Την εκμετάλλευση των ευέλικτων ετερογενών υπηρεσιών που παρέχονται από τα δίκτυα ραδιοεπικοινωνιών

Η γενική δομή ενός SDR πομποδέκτη φαίνεται στο Σχ.1.8



Σχήμα 1.8: Πομποδέκτης SDR

Ο όρος γνωστικές ραδιοεπικοινωνίες αποτελεί εξέλιξη του Software Define Radio (SDR). Στηρίχθηκε στις αρχές του SDR και τις επεξέτεινε με άμεση στόχο την αξιοποίηση του υποχρησιμοποιούμενου φάσματος. Όλα τα βήματα της εξέλιξης του SDR οδήγησαν αργά και σταθερά στις γνωστικές ραδιοεπικοινωνίες. Η ιδέα του των γνωστικών ραδιοεπικοινωνιών παρουσιάστηκε επίσημα για πρώτη φορά στην εργασία [1], όπου εισήχθησαν οι όροι aware (ενήμερος), adaptive (προσαρμοστική), και ideal Cognitive radio για να περιγράψουν τα διαφορετικά επίπεδα ικανότητας των γνωστικών ραδιοεπικοινωνιών. Επιπλέον προτάθηκε ο ακόλουθος ορισμός «*ο όρος γνωστικές ραδιοεπικοινωνίες ταυτίζεται με το σημείο όπου τα ασύρματα PDAs (personal digital assistants, πρωσοπικοί ψηφιακοί βοηθοί) και τα σχετικά με αυτά δίκτυα διαθέτουν αρκετή υπολογιστική νοημοσύνη ως προς την αξιοποίηση των ραδιοπόρων και τη σχετική επικοινωνία μεταξύ υπολογιστών, ώστε να ανιχνεύουν τις τηλεπικοινωνιακές ανάγκες των χρηστών ως συνάρτηση του περιβάλλοντος χρήσης και για να παρέχουν ραδιοπόρους και ασύρματες επικοινωνίες κατάλληλες να ικανοποιήσουν τις ανάγκες αυτές*» Με αυτό τον τρόπο οι γνωστικές ραδιοεπικοινωνίες είναι ικανές να επιλέξουν αυτόματα την καλύτερη υπηρεσία και να καθυστερήσουν ή να προωθήσουν άμεσα ορισμένες ασύρματες μεταδόσεις ανάλογα με τους διαθέσιμους ή προβλεπόμενους πόρους.

Η FCC έδωσε τον ακόλουθο ορισμό για το Cognitive radio [5]: «*Cognitive radio είναι ένα σύστημα ραδιοεπικοινωνιών το οποίο μπορεί να αλλάξει τις παραμέτρους μετάδοσής του, βασιζόμενο στην αλληλεπίδραση με το περιβάλλον το οποίο λειτουργεί*»

Η FCC αναγνώρισε τα ακόλουθα χαρακτηριστικά που πρέπει να διαθέτει ένα CR σύστημα για να επιτύχει την αποτελεσματική χρήση του φάσματος.

- *Ευελιξία συχνότητας (Frequency Agility):* Το σύστημα πρέπει να μπορεί να μεταβάλλει τη συχνότητα λειτουργίας του ώστε να μπορεί να προσαρμόζεται στο περιβάλλον του.
- *Δυναμική επιλογή συχνότητας (Dynamic Frequency Selection DFS):* Ο μηχανισμός αυτός επιτρέπει τη χρήση των υποχρησιμοποιούμενων

αδειοδοτημένων ζωνών συχνοτήτων, υπό την προϋπόθεση ότι δεν προκαλούνται παρεμβολές στα αδειοδοτημένα συστήματα

- *Έλεγχος Ισχύος Εκπομπής (Transmit Power Control):* Ο TPC είναι ένας μηχανισμός ο οποίος προσαρμόζει την ισχύ ενός σταθμού στις συνθήκες που επικρατούν στο δίαυλο μετάδοσης.
- *Προσαρμοστική Διαμόρφωση (Adaptive Modulation):* Μια διάταξη CR πρέπει να είναι ικανή να μεταβάλλει δυναμικά το σχήμα διαμόρφωσης που χρησιμοποιεί κατα την εκπομπή σημάτων.
- *Επίγνωση της θέσης (Location Awareness):* Ένα CR σύστημα πρέπει να είναι ικανό να αναγνωρίζει τη θέση του καθώς και τη θέση των υπόλοιπων συσκευών που χρησιμοποιούν την ίδια ζώνη συχνοτήτων.
- *Διαπραγμάτευση χρήσης φάσματος (Negotiated Use):* Ένα CR σύστημα πρέπει να διαθέτει αλγορίθμους που επιτρέπουν τον καταμερισμό του φάσματος με βάση προσυμφωνημένους κανόνες.

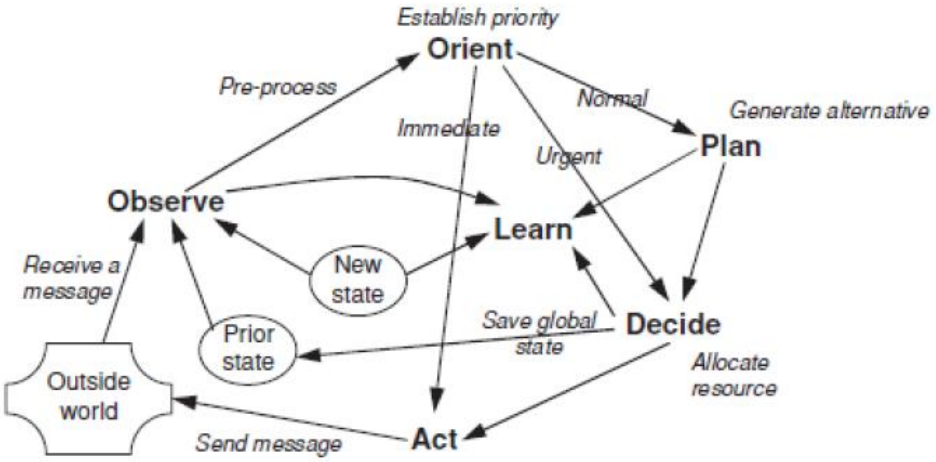
Επίσης, παρατίθεται και ο ορισμός του Simon Haykin όπως προκύπτει από μια δημοσίευσή του [6]: *"Σύστημα γνωστικών ραδιοεπικοινωνιών είναι ένα έξυπνο ασύρματο σύστημα επικοινωνιών που έχει γνώση για το περιβάλλον του (δηλαδή για τον εξωτερικό κόσμο) και χρησιμοποιεί τη μέθοδο «κατανοώ οικοδομώντας» (Understanding by building) για να μάθει από το περιβάλλον του και να προσαρμόσει τις εσωτερικές του καταστάσεις στις στατιστικές μεταβολές των εισερχομένων RF ερεθισμάτων, κάνοντας αντίστοιχες αλλαγές σε συγκεκριμένες λειτουργικές παραμέτρους(π.χ. στην ισχύ μετάδοσης, τη συχνότητα των φερόντων και το σχήμα διαμόρφωσης) σε πραγματικό χρόνο και έχοντας τους ακόλουθους δύο κύριους στόχους:*

- *Επικοινωνίες υψηλής αξιοπιστίας οποτεδήποτε και οπουδήποτε χρειαστεί*
- *Αποδοτική χρήση του φάσματος ραδιοσυχνοτήτων*

Τέλος, το 2007 η ομάδα εργασίας IEEE 1900.1, η οποία δημιουργήθηκε για τον καθορισμό ορολογιών και εννοιών, πρότεινε τον ακόλουθο ορισμό [7] για τα συστήματα γνωστικών ραδιοεπικοινωνιών: *"ένα είδος ασύρματου συστήματος ραδιοεπικοινωνιών το οποίο μπορεί να ανιχνεύσει και να κρίνει αυτόνομα το περιβάλλον του και προσαρμόζεται ανάλογα σε αυτό. Για το ασύρματο αυτό σύστημα επικοινωνιών θα μπορούσε να χρησιμοποιεί απόκτηση γνώσης, αυτοματοποιημένη κρίση και μηχανισμό μάθησης για την εγκατάσταση, διεξαγωγή και τερματισμό επικοινωνίας με άλλα ασύρματα συστήματα επικοινωνιών. Τα γνωστικά συστήματα ραδιοεπικοξνωνιών μπορούν δυναμικά και αυτόνομα να προσαρμόζουν τις λειτουργικές τους παραμέτρους" .*

1.4 Αρχές λειτουργίας και επίπεδα πρωτοκόλλων σε συστήματα γνωστικών ραδιοεπικοινωνιών

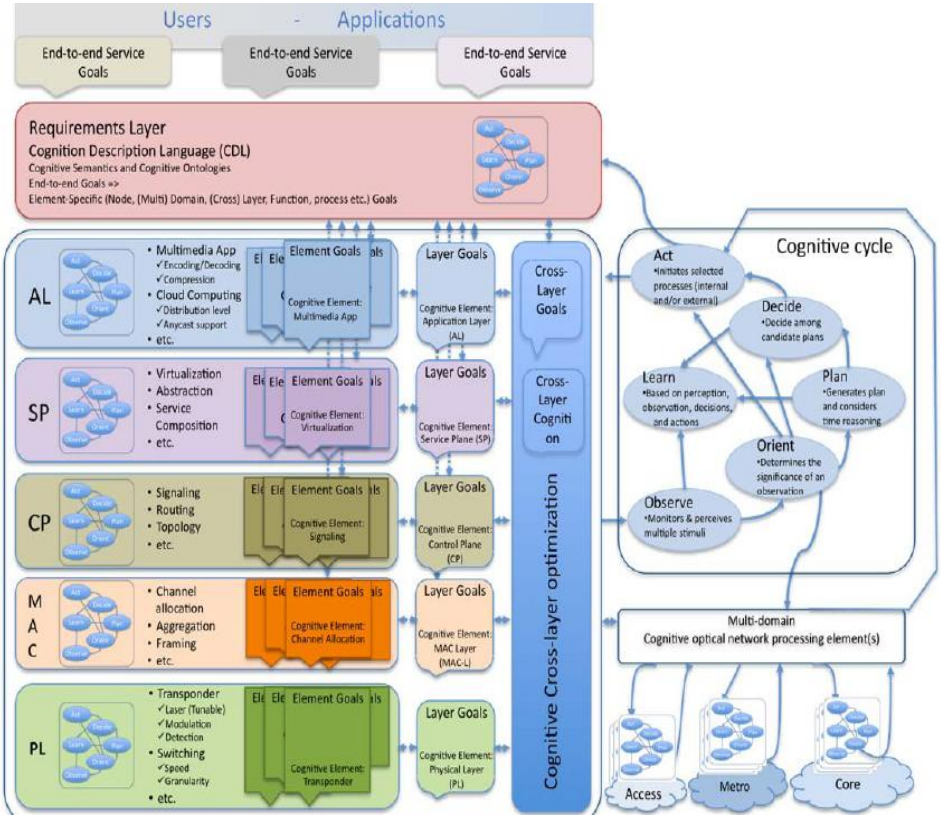
Η λειτουργία ενός συστήματος γνωστικών ραδιοεπικοινωνιών βασίζεται στον κύκλο γνώσης που απεικονίζεται στο Σχ.1.9 όπως αυτός παρουσιάστηκε από τον Mitola [1]. Το τηλεπικοινωνιακό σύστημα αποκτά πληροφορίες σχετικά με το εξωτερικό περιβάλλον (Outside world) μέσω άμεσων παρατηρήσεων (Observe) ή μέσω σηματοδοσίας. Έπειτα, οι πληροφορίες αυτές εκτιμώνται (Orient) ώστε να αποτιμηθεί η σπουδαιότητά τους. Με την αποτίμηση των πληροφοριών πραγματοποιείται ο προγραμματισμός των εναλλακτικών σχεδιασμών (Plan) και, στη συνέχεια, με κριτήριο τη βελτιστοποίηση του τελικού στόχου, επιλέγεται (Decide) και εκτελείται (Act) η κατάλληλη ενέργεια. Τα αποτελέσματα των ενεργειών αυτών εξαντικρύζονται στις επιδόσεις του συστήματος και στις ενδεχόμενες ανεπιθύμητες παρεμβολές που παρουσιάζονται στο εξωτερικό περιβάλλον. Το ασύρματο σύστημα επικοινωνιών, τέλος χρησιμοποιεί τα αποτελέσματα και τις αποφάσεις που λαμβάνονται για βελτίωση της διαδικασίας μάθησης (Learn), με σκοπό τη βελτίωση της λειτουργίας του, δημιουργώντας νέες μονελοποιημένες καταστάσεις και παράγοντας νέες εναλλακτικές στρατηγικές που προσαρμόζονται καλύτερα στις συνθήκες του περιβάλλοντος λειτουργίας.



Σχήμα 1.9: Γνωστικός κύκλος

Ο τελικός στόχος των συστήματων γνωστικών ραδιοεπικονωνιών είναι να χρησιμοποιήσουν με αποδοτικότερο τρόπο την υποδομή του δικτύου και τις υπηρεσίες του. Σε ένα δίκτυο γνωστικών ραδιοεπικοινωνιών το περιβάλλον δεν είναι ποτέ στατικό. Τα πρωτόκολλα, οι μηχανισμοί, οι αλγόριθμοι και τα συστήματα μαθαίνουν συνεχώς και προσαρμόζονται στις μεταβαλλόμενες συνθήκες περιβάλλοντος, με σκοπό να επιτύχουν την καλύτερη δυνατή αξιοποίηση των πόρων του.

Η αρχιτεκτονική ενός δικτύου γνωστικών ραδιοεπικοινωνιών και οι λειτουργίες τις οποίες επιτελεί κάθε στρώμα παρατίθενται στην ακόλουθη γραφική παράσταση. Τα ανώτερα επίπεδα σχετίζονται με ειδικότερες εφαρμογές και βελτιστοποίηση άλλων παραμέτρων του δικτύου . Σημασιολογίες (τρόποι λειτουργίας) και οντότητες (π.χ ένας δρομολογητής) του δικτύου μπορούν να δρουν όπως ορίζουν οι γνωστικές ραδιοεπικοινωνίες χρησιμοποιώντας το γνωστικό κύκλο (cognitive cycle)→(Observe-Orient-Plan-Decide-Act and Learn) ώστε να μάθουν να σχεδιάζουν και να ενεργούν βασιζόμενες σε προηγούμενες εμπειρίες. Ενδεικτικά, αναφέρεται ότι στο στρώμα εφαρμογής (Application Layer, AL), τα στοιχεία του γνωστικού κύκλου μπορούν να αναπτυχθούν ώστε να δημιουργήσουν γνωστικά στοιχεία encoding/decoding και compression για εφαρμογές πολυμέσων.



Σχήμα 1.10: Cognitive Optical Network Architecture

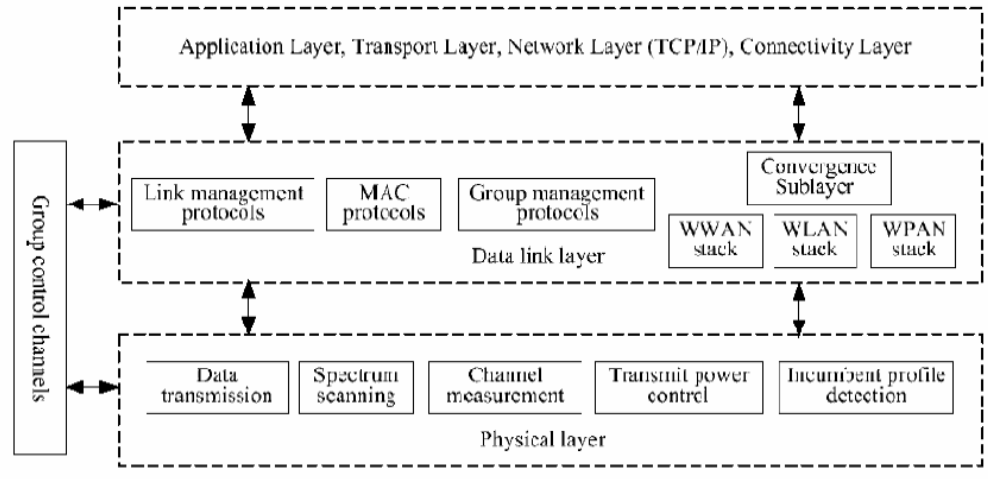
Βεβαίως η διαχείριση του φάσματος σχετίζεται πρωτίστως με τα δύο κατώτερα επίπεδα πρωτοκόλλων: το φυσικό επίπεδο (Physical Layer) και το επίπεδο ζεύξης δεδομένων (Link Layer).

Το φυσικό στρώμα περιλαμβάνει τις ακόλουθες διαδικασίες.

- *Μετάδοση των δεδομένων (Data Transmission)* η οποία περιλαμβάνει την επιλογή σχήματος διαμόρφωσης και κωδικοποίησης, την επιλογή ρυθμού μετάδοσης κλπ.
- *Σάρωση του φάσματος (Spectrum Scanning)* για την ανίχνευση φασματικών κενών τα οποία μπορούν να χρησιμοποιηθούν.
- *Μετρήσεις του διαύλου (Channel Measurement)* ώστε να προσδιοριστούν οι συνθήκες μετάδοσης και να καθοριστούν οι παράμετροι μετάδοσης.
- *Έλεγχος ισχύος εκπομπής (TCP-Transmit Power Protocol)* για την αποφυγή παρεμβολών.
- *Επικείμενη ανίχνευση προφίλ (incumbent profile detection)*

Όσον αφορά το στρώμα ζεύξης δεδομένων περιλαμβάνονται:

- *Πρωτόκολλα Διαχείρισης Ζεύξης (Link Management Protocols)* τα οποία είναι υπεύθυνα για τη ζεύξη μεταξύ δύο χρηστών τεχνολογίας “Cognitive radio”
- *Πρωτόκολλα MAC* τα οποία συμβάλλουν στην επιλογή κατάλληλων συχνοτήτων για την ομαλή επικοινωνία των κόμβων του δικτύου και την επίλυση προβλημάτων κρυμμένου ή αλλιώς εκτεθειμένου τερματικού
- *Πρωτόκολλα Διαχείρισης Ομάδων (Group Management Protocols)* για το συντονισμό των χρηστών που ανήκουν στην ίδια ομάδα (υποδίκτυο).
- *Υπόστρωμα Σύγκλισης (Convergence Sublayer)* το οποίο παρέχει τη δυνατότητα στο γνωστικό σύστημα να λειτουργεί σε εντελώς διαφορετικά ασύρματα περιβάλλοντα όπως για παράδειγμα Ασύρματα Τοπικά Δίκτυα (WLANs), Ασύρματα Προσωπικά Δίκτυα (WPANs) και Ασύρματα Δίκτυα Ευρείας Περιοχής (WWANs).

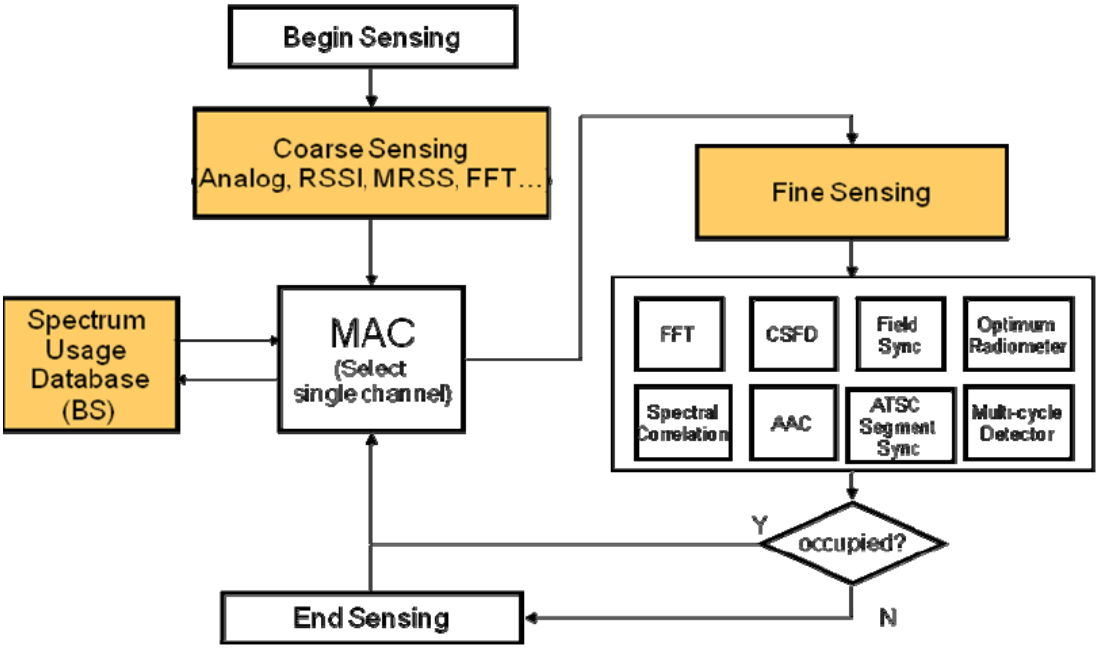


Σχήμα 1.11: Ιεραρχία πρωτοκόλλων λειτουργίας σε Cognitive Radio

1.5 Βασικές λειτουργίες συστημάτων γνωστικών ραδιοεπικοινωνιών

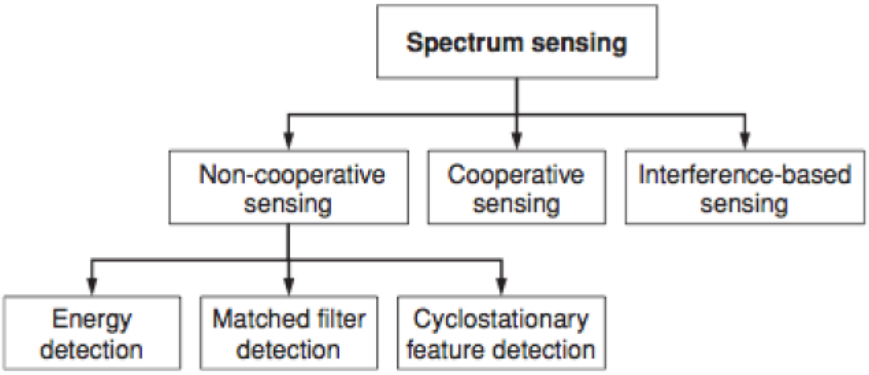
Οι βασικές λειτουργίες ενός συστήματος γνωστικών ραδιοεπικοινωνιών είναι οι ακόλουθες:

1.5.1 Ανίχνευση φάσματος (Spectrum sensing): Ο κύριος ρόλος της λειτουργίας αυτής είναι η ανίχνευση του αναξιοποίητου ραδιοφάσματος και ο διαμοιρασμός του σε άλλους χρήστες χωρίς να προκαλούνται επιζήμιες παρεμβολές στα δίκτυα πρωτευόντων χρηστών. Η από κοινού ανίχνευση του φάσματος των συστημάτων πρωτευόντων χρηστών, τα οποία αποτελεί σημαντική απαίτηση των δικτύων γνωστικών ραδιοεπικοινωνιών είναι ο πλέον αποτελεσματικός τρόπος για την ανίχνευση κενών στο ραδιοφάσμα. Στην εικόνα που ακολουθεί απεικονίζεται η στρατηγική της ανίχνευσης φάσματος.

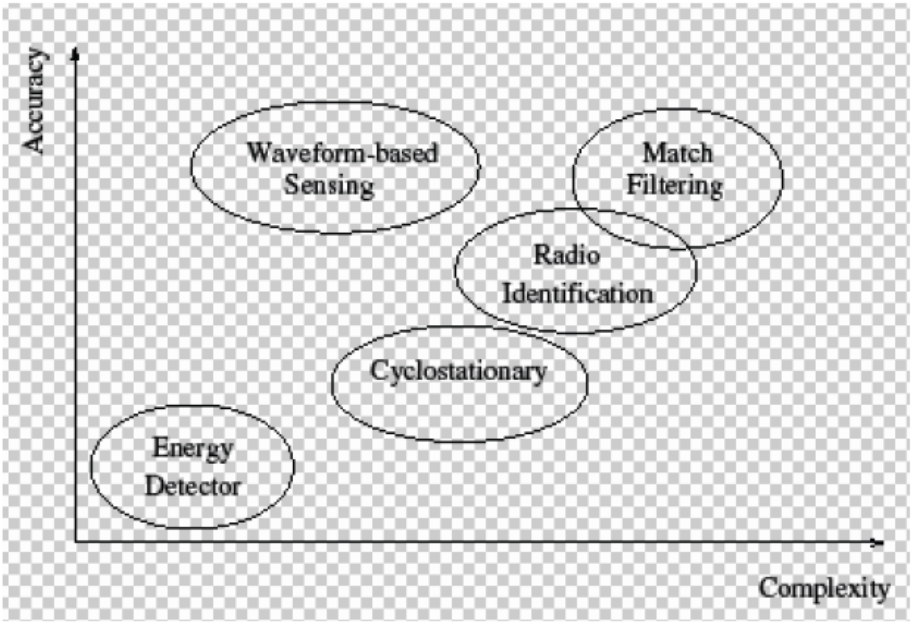


Σχήμα 1.12: Στρατηγική και συστήματα για ανίχνευση φάσματος

Ειδικότερα οι τεχνικές για ανίχνευση φάσματος μπορούν να ταξινομηθούν στις ακόλουθες τρεις κατηγορίες: ανίχνευση πομπού, συνεργατική ανίχνευση, ανίχνευση παρεμβολής.



Σχήμα 1.13: Κατηγορίες τεχνικών ανίχνευσης φάσματος



Σχήμα 1.14: Κύριες μέθοδοι ανίχνευσης φάσματος ως συνάρτηση της ακρίβειας (accuracy) στην ανίχνευση και της πολυπλοκότητας (complexity) στην τεχνική ανίχνευσης.

1.5.1.1. Ανίχνευση πομπού (Non-cooperative sensing): Ένα σύστημα γνωστικών ραδιοεπικοινωνιών πρέπει να είναι σε θέση να διαχωρίζει τις φασματικές ζώνες που χρησιμοποιούνται από αυτές που δεν χρησιμοποιούνται. Για το λόγο αυτό πρέπει να μπορεί να αποφασίζει εάν ένα σήμα από κάποιον πρωτεύοντα πομπό είναι τοπικά παρόν σε συγκεκριμένη ζώνη του φάσματος. Η προσέγγιση της ανίχνευσης πομπού βασίζεται στην ανίχνευση του αδύναμου σήματος που εκπέμπει ένας πομπός μέσω τοπικών παρατηρήσεων των χρηστών του συστήματος γνωστικών ραδιοεπικοινωνιών. Ένα βασικό θεωρητικό μοντέλο για την ανίχνευση πομπού μπορεί να οριστεί μέσω της σχέσης:

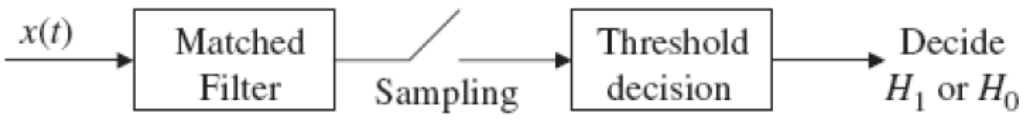
$$x(t) = \begin{cases} n(t) & H_o \\ hs(t)+n(t) & H_1 \end{cases} \quad (1.1)$$

όπου x(t) είναι το σήμα που λαμβάνεται από κάποιο χρήστη του συστήματος γνωστικών ραδιοεπικοινωνιών, s(t) είναι το σήμα που εκπέμπει ο πρωτεύων χρήστης, n(t) είναι ο θόρυβος AWGN και h είναι το κέρδος του διαύλου. H_0 είναι η υπόθεση ότι εδώ βρίσκεται σήμα μη εξουσιοδοτημένου χρήστη (δευτερεύοντα). H_1 είναι η εναλλακτική υπόθεση που δείχνει ότι υπάρχει σήμα πρωτεύοντα χρήστη.

Υπάρχουν τρεις βασικές τεχνικές ανίχνευσης πομπού: η ανίχνευση προσαρμοσμένου φίλτρου, η ανίχνευση ενέργειας, και η κυκλοστατική ανίχνευση.

α) Ανίχνευση με προσαρμοσμένο φίλτρο (Matced filter detection):

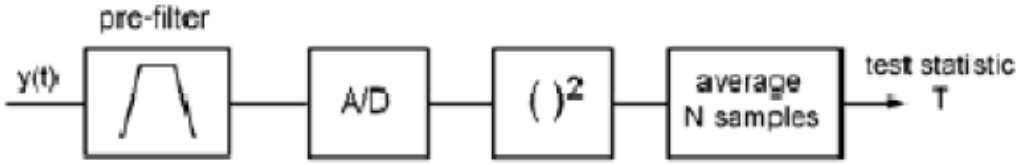
Ο βέλτιστος τρόπος για ανίχνευση κάθε σήματος είναι το προσαρμοσμένο φίλτρο, καθώς μεγιστοποιεί τον σηματοθορυβικό λόγο λήψης. Ωστόσο, το προσαρμοσμένο φίλτρο απαιτεί την αποδιαμόρφωση του πρωτεύοντος σήματος. Αυτό σημαίνει ότι μια συσκευή γνωστικών ραδιοεπικοινωνιών πρέπει να γνωρίζει εκ των προτέρων το σήμα του πρωτεύοντος σήματος στο φυσικό στρώμα και στο στρώμα MAC π.χ είδος διαμόρφωσης, μορφοποιητικό παλμό, μορφή πακέτου. Αυτή η πληροφορία πρέπει να είναι προαποθηκευμένη στη μνήμη της δευτερεύουσας συσκευής, όμως για την αποδιαμόρφωση πρέπει να επιτευχθεί συγχρονισμός σε επίπεδο χρόνου, φέροντος, ακόμα και ισοστάθμιση καναλιού. Αυτό είναι ακόμη δυνατό αφού οι περισσότεροι πρωτεύοντες χρήστες έχουν σήματα πιλότους, λέξεις συγχρονισμού ή κώδικες διασποράς για πιλότους ή κανάλια συγχρονσμού, όπως π.χ τα πακέτα OFDM που έχουν προοίμια για τη λήψη πακέτου. Το κύριο πλεονέκτημα του προσαρμοσμένου φίλτρου είναι ότι λόγω της σύμφωνης αποδιαμόρφωσης απαιτεί λιγότερο χρόνο για να επιτευχθεί ο περιορισμός της πιθανότητας ανίχνευσης, καθώς ο απαιτούμενος αριθμός δειγμάτων αυξάνει ως $O(SNR)^{-1}$. Ωστόσο, ένα σημαντικό μειονέκτημα του προσαρμοσμένου φίλτρου είναι το κόστος και κυρίως το ότι η συσκευή γνωστικών ραδιοεπικοινωνιών πρέπει να έχει έναν δέκτη για κάθε κατηγορία πρωτεύοντος χρήστη, που έχει και το μειονέκτημα της μεγάλης κατανάλωσης ισχύος.



Σχήμα 1.15: Υλοποίηση ανιχνευτή με προσαρμοσμένο φίλτρο

β) Ενεργειακή ανίχνευση (Energy detection)

Ο ανιχνευτής ενέργειας είναι ο συνηθέστερος ανιχνευτής φάσματος καθώς δεν απαιτεί καμία γνώση για το πρωτεύον σύστημα [9]. Ένας τυπικός ανιχνευτής ενέργειας αποτελείται από ένα βαθυπερατό φίλτρο (αφού έχει προηγηθεί το τμήμα ενδιάμεσης συχνότητας IF με ζωνοπερατό φίλτρο για να απορριφθεί ο εκτός ζώνης θόρυβος καθώς και η παρεμβολή γειτονικού διαύλου). Ένα μετατροπέα A/D με δειγματοληψία Nyquist, κύκώμα τετραγωνικού νόμου και τέλος έναν ολοκληρωτή.



Σχήμα 1.16: Υλοποίηση ανιχνευτή ενέργειας με βαθυπερατό φίλτρο

Το σήμα εξόδου από το φίλτρο εύρους W υψώνεται στο τετράγωνο και ολοκλήρώνεται στο διάστημα παρατήρησης. Ένας αλγόριθμος παρατήρησης συγκρίνει την έξοδο του ολοκληρωτή ,Τ,με ένα κατώφλι λ προκειμένου να διαπιστώσει αν υπάρχει ή όχι αδειοδοτημένος χρήστης. Αν Τ>λ υπάρχει πρωτεύοντας χρήστης, άλλως δεν υπάρχει. Εφόσον η ενεργειακή ανίχνευση εφαρμοστεί σε περιβάλλον χωρίς εξασθένηση όπου h είναι το κέρδος πλάτους του καναλιού όπως φαίνεται στην (1.1), τότε η πιθανότητα ανίχνευσης πρωτεύοντα χρήστη είναι η P_d. Ο συναγερμός λάνθασμένης ανίχνευσης πρωτεύοντα χρήστη δίνεται από την πιθανότητα P_f. Οι τύποι από τους οποίους υπολογίζονται οι δύο πιθανότητες είναι:

$$P_d = P\{Y > \lambda \mid \mathrm{H}_1\} = Q_m(\sqrt{2\gamma}, \sqrt{\lambda}) \tag{1.2}$$

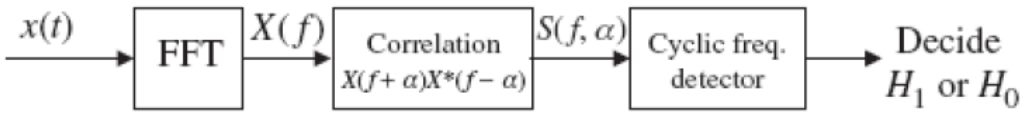
$$P_f = P\{Y > \lambda \mid \mathrm{H}_0\} = \frac{\Gamma(m, \lambda/2)}{\Gamma(m)} \tag{1.3}$$

όπου γ είναι ο σηματοθορυβικός λόγος SNR, u = TW είναι το γινόμενο του διαστήματος παρατήρησης επί το εύρος του φίλτρου, Γ(.) και Γ(.,.) είναι πλήρεις ή μη πλήρεις γάμα συναρτήσεις are και $Q_m(\ )$ είναι η γενικευμένη Marcum Q-συνάρτηση. Όπως προκύπτει από τις παραπάνω συναρτήσεις χαμηλές τιμές της πιθανότητας P_d οδηγούν σε αυξημένη παρεμβολή στον πρωτεύοντα χρήστη καθώς η πιθανότητα να μην γίνει αντιληπτός από ένα δευτερεύοντα χρήστη είναι μεγάλη. Αντίθετα υψηλές τιμές P_f έχουν ως αποτέλεσμα τη χαμηλή χρησιμοποίηση του φάσματος καθώς οι λανθασμένοι συναγερμοί ύπαρξης πρωτεύοντος σήματος αυξάνουν το πλήθος των χαμένων ευκαιριών πρόσβασης.

Ο ανιχνευτής ενέργειας είναι εύκολος στην υλοποίηση αλλά χαρακτηρίζεται από αρκετά μειονεκτήματα. Η απόδοσή του επηρεάζεται από τη διακύμανση στην ισχύ θορύβου. Για να επιλυθεί αυτό το πρόβλημα, χρησιμοποιείται ένας πιλοτικός τόνος από τον πρωτεύοντα πομπό για τη βελτίωση της ακρίβειας του ενεργειακού ανιχνευτή. Ένα άλλο μειονέκτημα είναι ότι ο ενεργειακός ανιχνευτής δεν μπορεί να διαφοροποιήσει διαφορετικούς τύπους σημάτων παρά μόνο να αποφασίσει για την ύπαρξη του σήματος. Επομένως, ο ενεργειακός ανιχνευτής είναι επιρρεπής σε λανθασμένη ανίχνευση που προκαλείται από αδιάφορα σήματα .

γ) Ανίχνευση στοιχείου κυκλοστατικότητας (Cyclostationary feature detection)

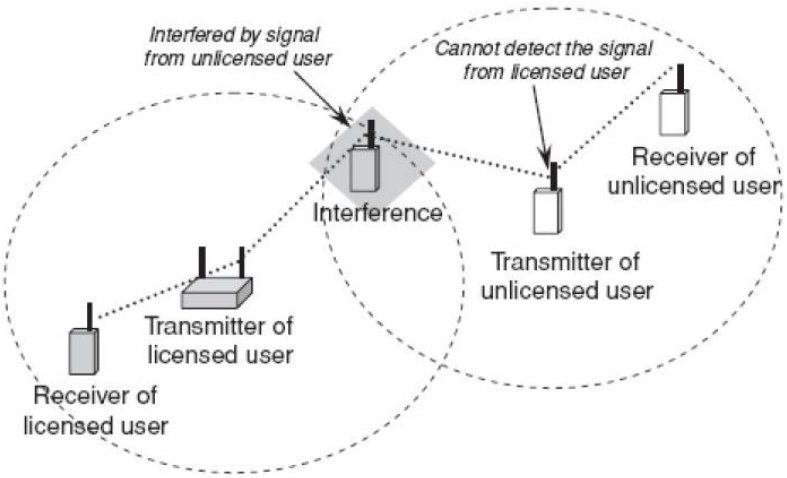
Το σήμα που εκπέμπει ένας αδειοδοτημένος χρήστης παρουσιάζει, στη γενική περίπτωση, περιοδικότητα που αναφέρεται ως κυκλοστατικότητα και μπορεί να χρησιμοιποιηθεί για τον εντοπισμό της παρουσίας πρωτευόντων χρηστών [8]. Ένα σήμα θεωρείται κυκλοστατικό όταν η συνάρτηση αυτοσυσχέτισής του είναι περιοδική. Με βάση την περιοδικότητα αυτή το σήμα εκπομπής πρωτεύοντος χρήστη μπορεί να διαχωριστεί από το θόρυβο, ο οποίος είναι στατικός και δεν παρουσιάζει συσχέτιση. Γενικά, η κυκλοστατική ανίχνευση προσφέρει ένα αρκετά ακριβές αποτέλεσμα ανίχνευσης και δεν επηρεάζεται από μεταβολές της ισχύος του θορύβου. Ωστόσο, η διαδικασία ανίχνευσης είναι περίπλοκη και απαιτεί μακράς διάρκειας περιόδους παρατήρησης για την απόκτηση του αποτελέσματος ανίχνευσης. Η υλοποίηση ενός συστήματος ανίχνευσης στοιχείου κυκλοστατικότητας φαίνεται στο Σχ.1.16 που ακολουθεί.



Σχήμα 1.17: Υλοποίηση ανιχνευτή στοιχείου κυκλοστατικότητας

1.5.1.2. Συνεργατική ανίχνευση (Cooperative sensing)

Η συνεργασία προτείνεται ως λύση για προβλήματα που προκύπτουν κατά την ανίχνευση φάσματος λόγω της τυχαίας διακύμανσης θορύβου, διαλείψεων και σκίασης. Η συνεργατική ανίχνευση μειώνει αισθητά την πιθανότητα λανθασμένης ανίχνευσης και την πιθανότητα ψευδούς συναγερμού (false alarm). Επιπλέον, η συνεργασία δίνει τη δυνατότητα λύσης του προβλήματος του κρυμμένου πρωτεύοντος τερματικού. Το πρόβλημα αυτό συνίσταται στην αδυναμία του πομπού και του δέκτη του δευτερεύοντος δικτύου να εντοπίσουν το σήμα ενός πρωτεύοντος πομπού καθώς βρίσκονται έξω από την περιοχή κάλυψής του. Συνεπώς όταν ο δευτερεύων πομπός εκπέμψει θα παρεμβάλλει τον πρωτεύοντα δέκτη.

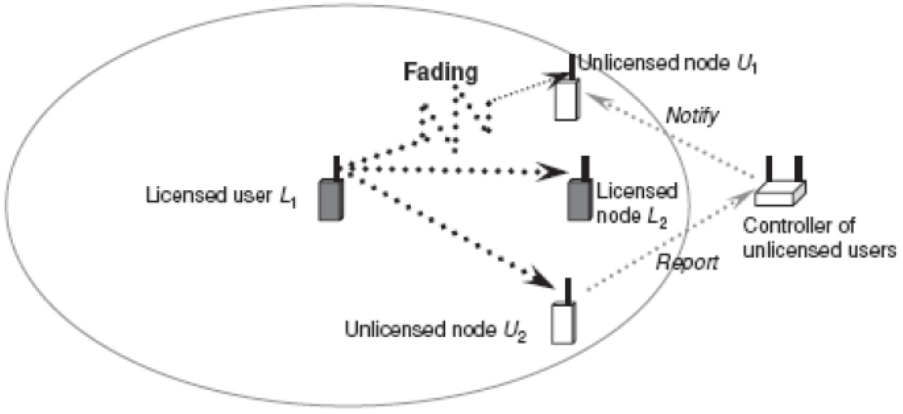


Σχήμα 1.18: Πρόβλημα Κρυμμένου Τερματικού

Η συνεργατική ανίχνευση μπορεί να είναι κεντρικά ελεγχόμενη ή κατανεμημένη.

α) Κεντρικά ελεγχόμενη συνεργατική ανίχνευση

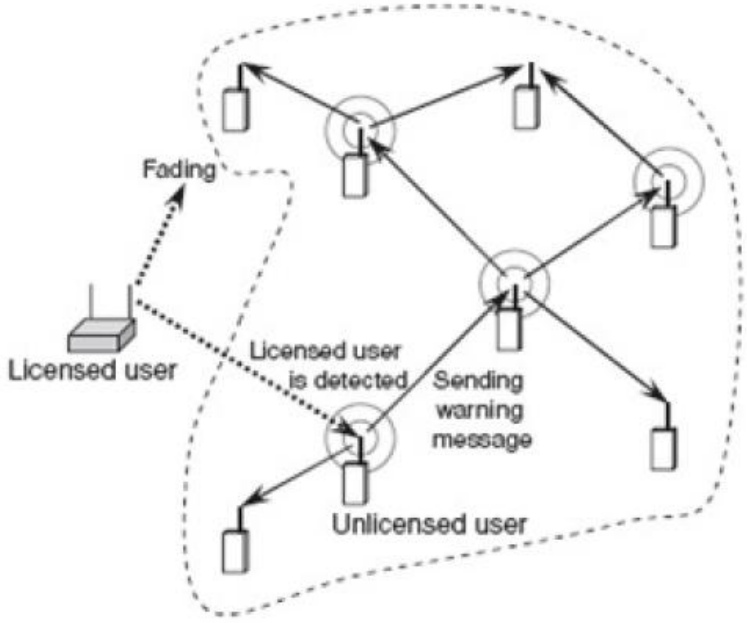
Σύμφωνα με αυτό τον τρόπο ανίχνευσης, μια κεντρική μονάδα συλλέγει πληροφορίες ανίχνευσης από συσκευές γνωστικών ραδιοεπικοινωνιών, αναγνωρίζει το διαθέσιμο φάσμα και κάνει ευρυεκπομπή της πληροφορίας αυτής σε άλλες δευτερεύουσες συσκευές ή ελέγχει απευθείας τη μεταδιδόμενη κίνηση μεταξύ των δευτερευουσών συσκευών. Το σημείο στο οποίο συγκεντρώνονται τα αποτελέσματα της ανίχνευσης είναι το σημείο πρόσβασης (Access Point). Στόχος είναι η ελάττωση των επιπτώσεων των διαλείψεων του διαύλου και η αύξηση της αξιοπιστίας και της ταχύτητας της ανίχνευσης.



Σχήμα 1.19: Παράδειγμα συνεργατικής κεντρικά ελεγχόμενης ανίχνευσης

β) Κατανεμημένη συνεργατική ανίχνευση

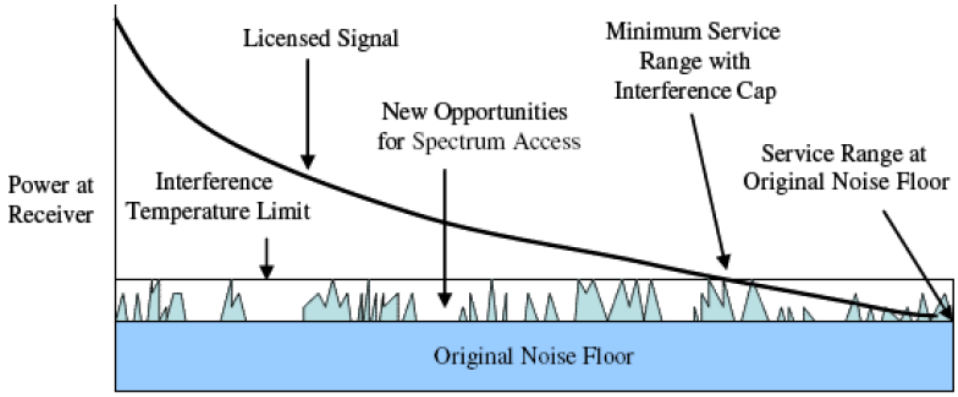
Κατά την κατανεμημένη συνεργατική ανίχνευση οι συσκευές γνωστικών ραδιοεπικοινωνιών ανταλλάσουν πληροφορίες μεταξύ τους αλλά λαμβάνουν αυτοτελώς την απόφαση για τη ζώνη του φάσματος στην οποία τελικά θα εκπέμψουν. Αυτός ο τύπος συνεργατικής ανίχνευσης πλεονεκτεί διότι δεν υπάρχει η ανάγκη για υποστηρικτική (backbone) υποδομή.



Σχήμα 1.20: Παράδειγμα κατανεμημένης συνεργατικής ανίχνευσης

1.5.1.3. Ανίχνευση παρεμβολής

Πρόκειται για νέα μέθοδο η οποία προτάθηκε από την FCC [10] σύμφωνα με την οποία ο αλγόριθμος ανίχνευσης μετρά τα επίπεδα παρεμβολής από όλες τις πηγές σημάτων στο δέκτη του πρωτεύοντος χρήστη. Η πληροφορία αυτή χρησιμοποιείται από ένα δευτεύοντα χρήστη ώστε να ελέγχει την πρόσβασή του στο φάσμα χωρίς να παραβιάζεται το επίπεδο της θερμοκρασίας παρεμβολής. Η έννοια της θερμοκρασίας παρεμβολής είναι I_T είναι παρόμοια με αυτή της θερμοκρασίας θορύβου, αλλά περιλαμβάνει τόσο το θόρυβο όσο και το γνωστό (υπολογίσιμο κατά ντετερμινιστικό τρόπο) επίπεδο παρεμβολής από άλλες πηγές σημάτων.

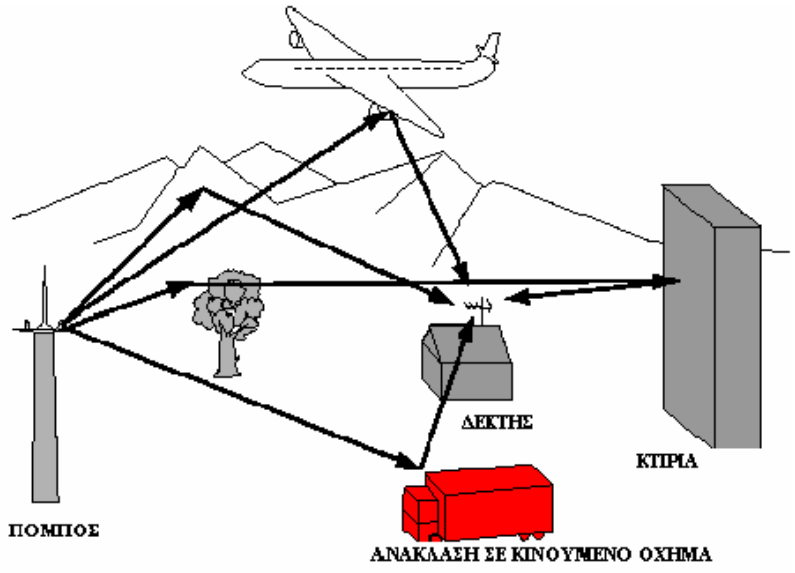


Σχήμα 1.21: Μοντέλο θερμοκρασίας παρεμβολής

Η θερμοκρασία παρεμβολής προσφέρει ακριβή και αποτελεσματική πληροφορία σχετικά με την παρεμβολή σε ζώνη συχνοτήτων εύρους W περί κεντρική συχνότητα f_c, ώστε ένας δευτερεύων χρήστης να μπορεί να αξιολογήσει κατά πόσο είναι εφικτή η πρόσβαση σε αυτή. Η θερμοκρασία παρεμβολής μετράται σε βαθμούς Kelvin και εκφράζεται μέσω της σχέσης:

$$I_T(f_c,W)=\frac{P_I(f_c,W)}{KW} \tag{1.5}$$

όπου $P_I(f_c,W)$ είναι η μέση ισχύς πηγής σήματος σε εύρος W περί τη συχνότητα f_c και K η σταθερά Boltzmann ($K=1,38\times10^{-23}$ W/Hz/^{0}K)



Σχήμα 1.22: Παράδειγμα παρεμβολών από κανάλια πολλαπλών διαδρομών

1.5.2 Διαχείριση φάσματος (Spectrum management)

Από τη στιγμή όπου σε δίκτυα γνωστικών ραδιοεπικοινωνιών γίνει η επιλογή της ζώνης φάσματος που έχει τη δυνατότητα να καλύψει τις απαιτήσεις ποιότητας υπηρεσίας, απαιτούνται νέες μέθοδοι διαχείρισης φάσματος. Λαμβάνοντας υπόψη τη δυναμική συμπεριφορά των νέων χαρακτηριστικών που εισάγουν οι γνωστικές ραδιοεπικοινωνίες στο φάσμα, οι λειτουργίες αυτές ταξινομούνται ως ανίχνευση φάσματος, ανάλυση φάσματος, και απόφαση φάσματος. Σε αυτή την ενότητα, εξετάζονται η ανάλυση φάσματος και η απόφαση φάσματος.

α) Ανάλυση φάσματος

Στα δίκτυα γνωστικών ραδιοεπικοινωνιών οι διαθέσιμες τρύπες φάσματος παρουσιάζουν χαρακτηριστικά τα οποία ποικίλλουν στο πεδίο του χρόνου. Η ανάλυση φάσματος επιτρέπει το χαρακτηρισμό των διαφορετικών ζωνών φάσματος. Ο χαρακτηρισμός αυτός μπορεί να χρησιμοποιηθεί για να εντοπίσει τη ζώνη φάσματος που είναι κατάλληλη για τις απαιτήσεις του δευτερεύοντος χρήστη. Προκειμένου να περιγραφεί η δυναμική φύση των δικτύων γνωστικών ραδιοεπικοινωνιών, κάθε φάσματική ευκαιρία πρέπει να χαρακτηριστεί εξετάζοντας όχι μόνο το μεταβαλλόμενο χρονικά περιβάλλον αλλά και τη δραστηριότητα των πρωτευόντων χρηστών και την πληροφορία των χαρακτηριστικών των ζωνών φάσματος, όπως η συχνότητα λειτουργίας και το εύρος ζώνης. Συνεπώς, πρέπει να οριστούν παράμετροι, όπως το επίπεδο παρεμβολής, το ποσοστό λαθών, καθυστέρηση στρώματος σύνδεσης και χρόνος εκμετάλλευσης. Με βάση τις προαναφερθείσες παραμέτρους μπορεί να ποσοτικοποιηθεί η ποιότητα υπηρεσίας μιας συγκεκριμένης περιοχής φάσματος.

Η χωρητικότητα του καναλιού, η οποία προσδιορίζεται με βάση τις προαναφερθείσες παραμέτρους είναι ο σημαντικότερος παράγοντας για το χαρακτηρισμό φάσματος. Συνήθως χρησιμοποιείται το SNR λήψης για το χαρακτηρισμό των ζωνών φάσματος. Εντούτοις, δεδομένου ότι το SNR λαμβάνει υπόψη μόνο τοπικές παρατηρήσεις των χρηστών, και δεν υπάρχει πλήρης εικόνα του φάσματος δεν είναι αρκετό για να αποφευχθεί η παρεμβολή στους πρωτεύοντες χρήστες. Κατά συνέπεια, ο χαρακτηρισμός του φάσματος στρέφεται στην εκτίμηση χωρητικότητας με βάση την παρεμβολή στους πρωτεύοντες δέκτες. Μπορεί να χρησιμποποιηθεί για αυτή την προσέγγιση το πρότυπο θερμοκρασίας παρεμβολής. Το όριο θερμοκρασίας παρεμβολής υποδεικνύει ένα ανώτατο όριο στην πρόσθετη ενέργεια RF που θα μπορούσε να εισαχθεί στη ζώνη. Συνεπώς, χρησιμοποιώντας το μέγεθος της επιτρεπόμενης παρεμβολής μπορεί να καθοριστεί η μέγιστη επιτρεπτή ισχύς εκπομπής ενός χρήστη γνωστικών ραδιοεπικοινωνιών.

Μια συνήθης μέθοδος εκτίμησης της χωρητικότητας του φάσματος λαμβάνει υπόψη της το εύρος ζώνης και την επιτρεπτή ισχύ εκπομπής. Σύμφωνα λοιπόν με τη μέθοδο αυτή η χωρητικότητα C υπολογίζεται από τη γνωστή σχέση:

$$C = B\log_2(1+\frac{S}{N+l}) \quad (1.6)$$

όπου B το εύρος ζώνης, S η ισχύς λήψης σήματος από τον χρήστη γνωστικών ραδιοεπιοινωνιών, N η ισχύς του θορύβου, και I είναι η ισχύς παρεμβολής στο δέκτη γνωστικών ραδιοεπικοινωνιών λόγω της παρουσίας του αρχικού πρωτεύοντος εκπομπού.

Ο υπολογισμός της χωρητικότητας φάσματος ερευνάται επίσης και στο πλαίσιο εφαρμογής της OFDM διαμόρφωσης σε συστήματα γνωστικών ραδιοεπικοινωνιών. Σύμφωνα με τη λογική αυτή, η χωρητικότητα φάσματος των δικτύων αυτών εκφράζεται μέσω της σχέσης:

$$C = \int_{\Omega}\frac{1}{2}\log_2(1+\frac{G(f)S_0}{N_0})df \quad (1.7)$$

όπου Ω το σύνολο όλων των τμημάτων του αχρησιμοποίητου φάσματος, G(f) είναι το κέρδος διαύλου στην συχνότητα f, S_0 και N_0 η ισχύς σήματος και θορύβου ανα μονάδα συχνότητας, αντίστοιχα. Προκειμένου να αποδοθεί το κατάλληλο φάσμα στις διάφορες εφαρμογές, είναι επιθυμητό και αντικείμενο έρευνας να προσδιοριστούν οι ζώνες φάσματος που συνδυάζουν όλες τις παραμέτρους χαρακτηρισμού που περιγράφηκαν προηγουμένως.

β) Απόφαση για πρόσβαση στο φάσμα (Spectrum decision)

Αφού χαρακτηριστούν όλες οι διαθέσιμες ζώνες φάσματος, πρέπει να γίνει η επιλογή της κατάλληλης ζώνης φάσματος για την τρέχουσα μετάδοση από το δευτερεύοντα χρήστη, λαμβάνοντας φυσικά υπόψη τις απιτήσεις QoS και τα χαρακτηριστικά του φάσματος. Με βάση τις απαιτήσεις των χρηστών μπορούν να καθοριστούν ο ρυθμός μετάδοσης, το αποδεκτό ποσοστό λαθών καθώς και το εύρος ζώνης της μετάδοσης. Υπεύθυνο για την πρόσβαση στο φάσμα είναι ένα γνωστικό πρωτόκολλο πρόσβασης μέσου (cognitive MAC protocol), που αποσκοπεί στην αποφυγή συγκρούσεων (παρεμβολή του ενός στον άλλο) τόσο με τους πρωτεύοντες όσο και με τους δευτερεύοντες χρήστες.

1.5.3 Κινητικότητα φάσματος (Spectrum mobility)

Ως κινητικότητα φάσματος ορίζεται η διαδικασία κατα την οποία ένας δευτερεύοντας χρήστης αλλάζει τη συχνότητα λειτουργίας του με σκοπό να προδιορίσει το καλύτερο διαθέσιμο κανάλι για την ποιότητα της υπηρεσίας που θέλει να εξασφαλίσει. Όταν ένας πρωτεύων χρήστης ξεκινά πρόσβαση σε ένα

κανάλι που χρησιμοποιείται από ένα δευτερεύοντα χρήστη, τότε ο δευτερεύων χρήστης μεταπηδά σε άλλο κανάλι που εκείνη τη στιγμή δεν χρησιμοποιείται. Ο σκοπός της διαχείρισης κινητικότητας φάσματος στα δίκτυα γνωστικών ραδιοεπικοινωνιών είναι η ανάγκη εξασφάλισης ότι οι μεταβάσεις αυτού του τύπου πραγματοποιούνται ομαλά και πολύ ταχέως ώστε κατά τη διάρκεια της μεταγωγής φάσματος (spectrum handoff) του δευτερεύοντος χρήστη να υποστούν την ελάχιστη δυνατή υποβάθμιση απόδοσης.

Συνεπώς, απαιτούνται πρωτόκολλα πολυστρωματικής διαχείρισης κινητικότητας για να υλοποιήσουν τις λειτουργίες φασματικής κινητικότητας. Αυτά τα πρωτόκολλα υποστηρίζουν διαχείριση κινητικότητας προσαρμόσιμη στους διαφορετικούς τύπους εφαρμογών. Για παράδειγμα, μια σύνδεση TCP μπορεί να τεθεί σε κατάσταση αναμονής μέχρι να ολοκληρωθεί η φασματική μεταγωγή. Επιπλέον, δεδομένου ότι οι παράμετροι TCP θα αλλάξουν μετά από μια φασματική μεταγωγή , είναι απαραίτητο να μάθουν τις νέες παραμέτρους και να εξασφαλίσουν ότι η μετάβαση από τις παλαιές παραμέτρους στις νέες παραμέτρους θα πραγματοποιηθεί ταχύτατα. Προκειμένου περί μετάδοσης δεδομένων π.χ., FTP, τα πρωτόκολλα διαχείρισης κινητικότητας πρέπει να εφαρμόσουν τους μηχανισμούς για να αποθηκεύονται τα πακέτα που μεταδίδονται κατά τη διάρκεια μιας φασματικής μεταγωγής.

1.6 Πρόσφατα ασύρματα πρότυπα

Τα τελευταία ασύρματα πρότυπα έχουν αρχίσει να περιλαμβάνουν γνωστικά χαρακτηριστικά. Τα τερματικά γνωστικών ραδιοεπικοινωνιών αναμένεται να προκαλέσουν επανάσταση στον τρόπο που οι χρήστες θα χρησιμοποιούν το φάσμα στο άμεσο μέλλον. Στην παρούσα ενότητα γίνεται αναφορά με χρονολογική σειρά σε πρότυπα τα οποία έχουν υιοθετήσει τις αρχές των γνωστικών ραδιοεπικοινωνιών

1.6.1 Πρότυπο 802.11

Η ανάπτυξη των ασύρματων τοπικών δικτύων (WLANs) οδήγησε άμεσα σε πληθώρα διαφορετικών τύπων WLAN. Το πρόβλημα ήταν ότι κανένα από αυτά τα δίκτυα WLAN δεν ήταν συμβατό με τα υπόλοιπα. Τελικά, η βιομηχανία διαπίστωσε την ανάγκη ύπαρξης ενός κοινού προτύπου, τη σχεδίαση του οποίου ανέλαβε ειδική επιτροπή του IEEE η οποία δημοσίευσε το πρώτο πρότυπο IEEE 802.11 το 1997 [11]. Το ασύρματο LAN που περιγραφόταν λειτουργούσε με ρυθμούς μετάδοσης διαύλου 1 ή 2 Mbps και υποστήριζε τεχνολογίες διασποράς φάσματος (DSSS και FHSS) καθώς και υπέρυθρης ακτινοβολίας (DFIR). Λειτουργούσε στα 2.4 GHz και είχε μεγάλη εμβέλεια. Μετά τις διαμαρτυρίες για χαμηλές ταχύτητες ξεκίνησαν εργασίες για ταχύτερα πρότυπα. Ακολούθησαν αρκετές νέες εκδόσεις του προτύπου IEEE 802.11 με διαφορετικό χαρακτηριστικό γράμμα στο τέλος του ονόματός τους. Το 1999 δημοσιεύτηκαν δύο νέες εκδόσεις, οι IEEE 802.11a και IEEE

802.11b ενώ ακολούθησε η έκδοση IEEE 802.11g το 2003. Στις νέες αυτές εκδόσεις προδιαγράφονται και νέοι τύποι φυσικού στρώματος. Ο πίνακας 1.6 παρουσιάζει τα πρότυπα της σειράς IEEE 802.11 με τα κύρια χαρακτηριστικά τους.

Πρότυπο	Χρονολογία	Συχνότητα Λειτουργίας	Ρυθμός Μετάδοσης στο φυσικό στρώμα
802.11	1997	2.4-2.5 GHz	2 Mbit/sec
802.11a	1999	5.15-5.25/ 5.25-5.35/ 5.49-5.725/ 5.725-5.85 GHz	54 Mbit/sec
802.11b	1999	2.4-2.5 GHz	11 Mbit/sec
802.11g	2003	2.4-2.5 GHz	54 Mbit/sec
802.11n	Μάρτιος 2009	2.4 GHz και/ή 5 GHz	248 Mbit/sec

Πίνακας 1.6: Πρότυπα της σειράς IEEE 802.11

1.6.2 IEEE 802.11k

Μια προτεινόμενη επέκταση του 802.11 είναι το 802.11k το οποίο καθορίζει διάφορους τύπους μετρήσεων. Μερικές από τις μετρήσεις περιλαμβάνουν αναφορά κίνησης καναλιού, ιστόγραμμα θορύβου και στατιστικά σταθμού [12]. Το ιστόγραμμα θορύβου περιέχει μεθόδους για τη μέτρηση του επιπέδου της παρεμβολής που προέρχεται από όλη την ηλεκτρομαγνητική ακτινοβολία που δεν σχετίζεται με 802.11. Το σημείο πρόσβασης (Access Point) συλλέγει την πληροφορία από κάθε τερματικό και πραγματοποιεί τις δικές του μετρήσεις τις οποίες χρησιμοποιεί για να ρυθμίζει την πρόσβαση σε ένα συγκεκριμένο κανάλι. Στο 802.11k όταν ένα σημείο πρόσβασης με το δυνατότερο σήμα φορτώνεται στην πλήρη χωρητικότητα, κάθε νεοεισερχόμενη συσκευή ανατίθεται σε ένα υποχρησιμοποιούμενο σημείο πρόσβασης. Παρά το γεγονός ότι το σήμα λήψης είναι ασθενέστερο, ο συνολικός ρυθμός μετάδοσης του συστήματος είναι μεγαλύτερος λόγω αποδοτικότερης εκμετάλλευσης των πόρων του δικτύου.

1.6.3 Bluetooth

Μια νέα μέθοδος που ονμάζεται Προσαρμοστική Αναπήδηση Συχνότητας (Adaptive Frequency Hopping AFH), έχει εισαχθεί στο πρότυπο του Bluetooth με σκοπό να μειωθεί η παρεμβολή μεταξύ των ασύρματων συστημάτων που λειτουργούν στη μη αδειοδοτημένη ζώνη των 2.4GHz. Σε αυτή τη ζώνη, συσκευές 802.11b/g, ασύρματα τηλέφωνα, φούρνοι μικροκυμάτων χρησιμοποιούν τις ίδιες συχνότητες με το Bluetooth. Η AFH αναγνωρίζει τις αλλότριες μεταδόσεις στην ISM ζώνη και αποφεύγει τις συχνότητές τους. Έτσι η παρεμβολή στενής ζώνης μπορεί να αποφευχθεί και να επιτευχθεί καλύτερη απόδοση ως προς το ποσοστό λαθών, όπως επίσης και μείωση της ισχύος εκπομπής των τερματικών.

Η AFH χρειάζεται έναν αλγόριθμο ανίχνευσης για να προσδιορίσει αν λειτουργούν άλλες συσκευές στην ISM ζώνη και αν πρέπει να τις αποφύγει ή όχι. Ο αλγόριθμος ανίχνευσης βασίζεται σε στατιστικά που λαμβάνονται για να προσδιοριστεί ποιά κανάλια είναι κατειλλημένα και ποιά όχι. Τα στατιστικά των καναλιών μπορεί να είναι το ποσοστό λανθασμένων πακέτων (Packet Error Rate, PER), το ποσοστό λανθασμένων ψηφίων (Bit Error Rate, PER), η ένδειξη ισχύος του σήματος λήψης (Received Signal Strength Indicator, RSSI), ο λόγος φέροντος προς παρεμβολή(Carrier to Interference Noise Ratio, CINR) ή άλλα μεγέθη.

1.6.4 Τεχνολογία IEEE και νέες εφαρμογές γνωστικών ραδιοεπικοινωνιών

Τα τερματικά γνωστικών ραδιοεπικοινωνιών σε συνδυασμό με τους αλγορίθμους δυναμικής πρόσβασης φάσματος που ανιχνεύουν το διαθέσιμο φάσμα και διαχειρίζονται τις εκχωρήσεις συχνοτήτων σε ένα ασύρματο δίκτυο αποτελούν την εξέλιξη της τεχνολογίας γνωστικών ραδιοεπικοινωνιών. Επιπλέον, στις αγορές τηλεπικοινωνιών 3G και 4G υπάρχουν αρκετές στρατιωτικές και βοηθητικές αγορές οι οποίες θα επωφεληθούν από το πλεονέκτημα της νέας τεχνολογίας. Ορισμένες από αυτές τις αγορές είναι:

- Wireless Handset Games
- Municipality Frequency Use
- Internet Wireless Connectivity
- Cyber Attack Detection and Prevention
- Automobile and Transportation Applications
- Wireless Security
- Military Wireless Tactical Networks

Τα κριτήρια της IEEE στην τεχνολογία γνωστικών ραδιοεπικοινωνιών αναμένεται να αυξήσουν τη διαλειτουργικότητα μεταξύ διαφορετικών τεχνολογιών. Ειδικότερα τα IEEE 802 και SCC41 πρότυπα αναμένεται να έχουν σημαντική συμβολή στα μελλοντικά δίκτυα γνωστικών ραδιοεπικοινωνιών. Αυτή η σειρά προτύπων είναι το

πρώτο παράδειγμα ενεργειών της IEEE που προβλέπουν προσαρμοστικότητα, και ευφυία για τα τερματικά γνωστικών ραδιοεπικοινωνιών με σκοπό τη βελτιστοποιημένη χρήση των εφαρμογών τους.

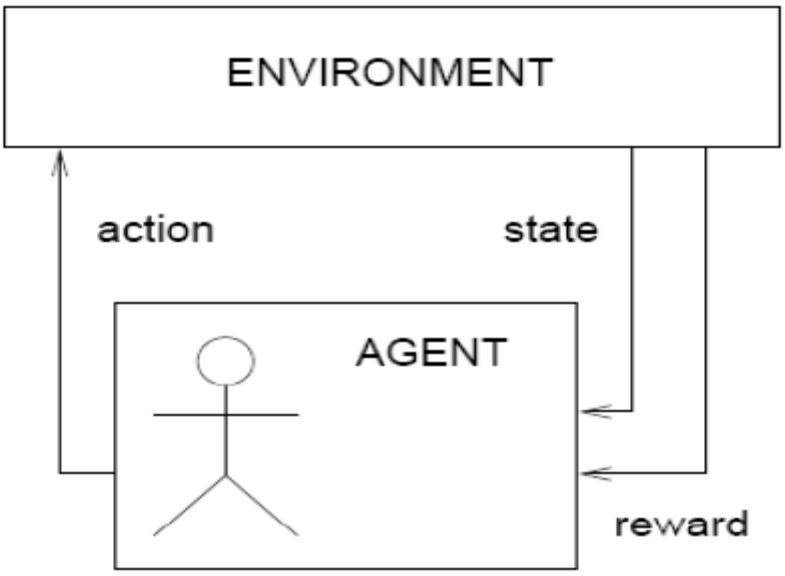
Τα τερματικά στο Δίκτυα πρόσβασης γνωστικών ραδιοεπικοινωνιών (Radio Access Networks) θα καθιστούν ικανή τη δυναμική πρόσβαση φάσματος (DSA) μέσω του προτύπου 802.22 για WRANs ορίζοντας τη συνάρτηση ανίχνευσης φάσματος. Το P1900.4 πρότυπο διαχειρίζεται τα τερματικά γνωστικών ραδιοεπικοινωνιών και τα RANs. Το πρότυπο P1900.3 επιτρέπει τις τεχνικές εκτίμησης συμφόρησης.Τα τερματικά γνωστικών ραδιοεπικοινωνιών και τα δίκτυα διαχείρισής τους αναμένεται να περιέχονται στις WiFi και Wimax αγορές το 2009-2011. Η εφαρμογή των προτύπων αυτών θα έχει ως αποτέλεσμα εφαρμογές υψηλής απόδοσης καθώς και χαμηλότερο κόστος από όσο κάποιος θα εκτιμούσε όταν γινόταν η εισαγωγή του όρου των γνωστικών ραδιοεπικοινωνιών στις υπηρεσίες ασύρματης επικοινωνίας.

ΚΕΦΑΛΑΙΟ 2^{ο}: ΕΝΙΣΧΥΤΙΚΗ ΜΑΘΗΣΗ

2.1 Προέλευση όρου-ορισμός

Ο όρος ενισχυτική μάθηση (reinforcement learning) αναφέρθηκε πρώτη φορά στην τεχνητή νοημοσύνη από τον Minski το 1961 και ανεξάρτητα στη θεωρία ελέγχου από τους Waltz και Fu το ίδιο έτος. Έχει τις αρχές του στις πρώτες μέρες ανάπτυξης του κυβερνοχώρου με τις έρευνες στη στατιστική, ψυχολογία, νευροεπιστήμη και επιστήμη των υπολογιστών. Τα τελευταία δέκα χρόνια έχει προσελκύσει ραγδαία αυξανόμενο ενδιαφέρον στον τομέα της μηχανικής μάθησης και της τεχνητής νοημοσύνης. Η υπόσχεση την οποία δίνει η ενισχυτική μάθηση είναι δελεαστική. Πρόκειται για ένα τρόπο προγραμματισμού πρακτόρων (agents) με ανταμοιβή ή τιμωρία σε κάθε ενέργειά τους, χωρίς να χρειάζεται να προσδιοριστεί ο τρόπος που πρέπει να επιτευχθεί ο στόχος.

Ως ενισχυτική μάθηση ορίζεται το πρόβλημα που αντιμετωπίζει ένας λογικός πράκτορας προκειμένου να εκπαιδευτεί σε συγκεκριμένη συμπεριφορά μέσω αλληλεπιδράσεων δοκιμής λάθους με ένα δυναμικό περιβάλλον. Το γενικό σχήμα της ενισχυτικής μάθησης παρουσιάζεται ως ακολούθως.



Σχήμα 2.1: Γενικό σχήμα ενισχυτικής μάθησης

Υπάρχουν δύο βασικές στρατηγικές για την αντιμετώπιση προβλημάτων ενισχυτικής μάθησης. Η πρώτη είναι να διευκρινήσει ο πράκτορας το χώρο των συμπεριφορών όπου έχει οριστεί στο σύστημα και να προσδιορίσει μια συμπεριφορά η οποία λειτουργεί καλά (με αποδοτικό τρόπο) μέσα στο περιβάλλον. Η προσέγγιση έχει βρει πρόσφορο έδαφος στους γενετικούς αλγορίθμους και στο γενετικό προγραμματισμό. Η δεύτερη στρατηγική είναι να γίνει χρήση στατιστικών

τεχνικών και μεθόδων δυναμικού προγραμματισμού ώστε να υπολογιστεί η χρησιμότητα της επιλογής όλων των δράσεων στις καταστάσεις του περιβάλλοντος όπως αυτές έχουν οριστεί.

2.1.1 Διαφορά ενισχυτικής και επιβλεπόμενης μάθησης

Η ενισχυτική μάθηση διαφέρει από το πλέον συχνό και μελετημένο πρόβλημα της επιβλεπόμενης μάθησης σε αρκετά σημεία. Η σημαντικότερη διαφορά είναι ότι στην ενισχύτική μάθηση δεν υπάρχουν ζεύγη εισόδου-εξόδου. Αντίθετα, ο πράκτορας, αφού επιλέξει μια δράση, ενημερώνεται για την άμεση ανταμοιβή του και την επόμενη κατάσταση αλλά δεν ενημερώνεται για το ποιά είναι η καλύτερη δράση όσο αφορά το μακροπρόθεσμο ενδιαφέρον του στο πρόβλημα. Είναι πολύ σημαντικό για τον πράκτορα να συγκεντρώσει χρήσιμη εμπειρία για τις πιθανές καταστάσεις του συστήματος, τον τρόπο μετάβασης μεταξύ αυτών και τις ενδεχόμενες ανταμοιβές με σκοπό να λειτουργήσει κατά το βέλτιστο δυνατό τρόπο. Ακόμα μια διαφορά από την επιβλεπόμενη μάθηση είναι ότι η εντός λειτουργίας επίδοση (on-line performance) είναι σημαντική. Σε ορισμένες περιπτώσεις η επιβεβαίωση του συστήματος γίνεται ταυτόχρονα με τη μάθηση.

2.1.2 Διαφορά ενισχυτικής μάθησης και δυναμικού προγραμματισμού

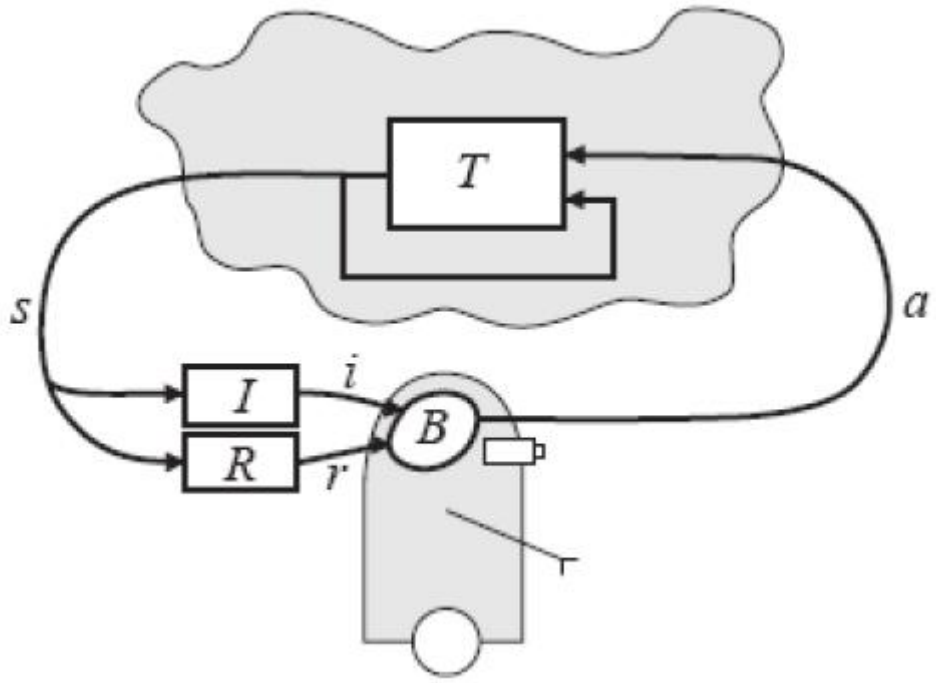
Η ενισχυτική μάθηση και ο δυναμικός προγραμματισμός είναι στενά συνδεδεμένες έννοιες, αφού και οι δύο προσεγγίσεις χρησιμοποιούνται για τη επίλυση αλυσίδων Markov. Η βασική ιδέα και των δύο είναι να διατυπωθούν οι συναρτήσεις ανταμοιβής, οι οποίες μπορούν να χρησιμοποιηθούν για την εύρεση της βέλτιστης πολιτικής. Παρά τη στενή αυτή σχέση, υπάρχει μια βασική διαφορά μεταξύ ενισχυτικής μάθησης και δυναμικού προγραμματισμού. Στην πρώτη, ο πράκτορας δεν γνωρίζει τη συνάρτηση ανταμοιβής και τη συνάρτηση μετάβασης καταστάσεων. Τόσο η αναταμοιβή όσο και η επιλογή μιας δράσης, καθορίζονται από το περιβάλλον και οι συνέπειες μιας δράσης προκύπτουν από την αλληλεπίδραση με το περιβάλλον. Δηλαδή οι πράκτορες της ενισχυτικής μάθησης δεν χρειάζεται να γνωρίζουν ένα μοντέλο του δικού τους περιβάλλοντος. Το γεγονός αυτό διαχωρίζει την ενισχυτική μάθηση από το δυναμικό προγραμματισμό, στον οποίο είναι απαραίτητες η πλήρης γνώση τόσο της συνάρτησης ανταμοιβής όσο και της συνάρτησης μετάβασης καταστάσεων. Για τη σύγκλιση στη βέλτιστη πολιτική απαιτείται ένας πεπερασμένος αριθμός επαναλήψεων και όχι μετάβαση στο πραγματικό περιβάλλον και παρατήρηση των αποτελεσμάτων.

2.1.3 Το πρόβλημα των k-κουλοχέρηδων

Η απλούστερη περίπτωση προβλήματος ενισχυτικής μάθησης είναι το γνωστό ως πρόβλημα των <k κουλοχέρηδων> (τα γνωστά μηχανήματα τζόγου), το οποίο έχει τροφοδοτήσει με σημαντικά στοιχεία την έρευνα της στατιστικής και των εφαρμοσμένων μαθηματικών. Ο πράκτορας βρίσκεται σε ένα δωμάτιο με ένα σύνολο k-κουλοχέρηδων, όπου έχει το δικαίωμα να τραβήξει το μοχλό μόνο σε h μηχανήματα. Ο όρος <δράση> θα χρησιμοποιηθεί για την επιλογή μοχλού από τον πράκτορα. Οποιοδήποτε μηχάνημα μπορεί να επιλεγεί κάθε φορά. Το μοναδικό κόστος του πράκτορα είναι η σπατάλη μιας από τις h προσπάθειες εφόσον επιλέξει κάποιο μη-βέλτιστο μηχάνημα. Όταν επιλεγεί το μηχάνημα i, αυτό δίνει ως επιστροφή 1 ή 0 με κάποια πιθανότητα p_i, με τα ενδεχόμενα των επιστροφών να είναι ασυμβίβαστα. Ζητούμενη είναι η πολιτική που πρέπει να ακολουθήσει ο πράκτορας για να επιτύχει τα μεγαλύτερα δυνατά κέρδη. Η απάντηση στο πρόβλημα αυτό ουσιαστικά εκφράζει τη βασική ανάγκη ισορρόπησης μεταξύ εκμετάλλευσης και εξερεύνησης. Ο πράκτορας μπορεί να πιστεύει ότι κάποιο μηχάνημα έχει σημαντικά υψηλή πιθανότητα επιστροφής. Γεννάται το ερώτημα, αν πρέπει να επιλέγει αυτό το μηχάνημα συνεχώς ή πρέπει να δοκιμάσει κάποιο άλλο για το οποίο έχει λιγότερες πληροφορίες; Η απάντηση σε αυτό το ερώτημα εξαρτάται από το πόσο χρόνο αναμένεται ο πράκτορας να παραμείνει στο παίγνιο. Όσο περισσότερη είναι η διάρκεια του παιγνίου τόσο χειρότερες είναι οι συνέπειες προσκόλλησης σε κάποιο μη βέλτιστο μηχάνημα και τόσο περισσότερο πρέπει να εξερευνήσει ο πράκτορας. Υπάρχουν αρκετές στρατηγικές για την επίλυση του συκεκριμένου προβλήματος. Καταρχήν, υπάρχουν λύσεις των οποίων τα αποτελέσματα είναι τυπικώς ορθά. Τέτοιες είναι ο δυναμικός προγραμματισμός, αυτόματα μάθησης, προσέγγιση Gittins με ευρετήρια κατανομής. Παρά το γεγονός ότι οι λύσεις αυτές μπορούν να επεκταθούν σε προβλήματα με πραγματικές τιμές ανταμοιβών, δεν μπορούν να εφαρμοστούν άμεσα στα προβλήματα ενισχυτικής μάθησης με πολλές καταστάσεις. Άλλες λύσεις με μεγαλύτερη χρήση αλλά με επιφυλάξεις ως προς την επιτυχία της αποτελεσματικότητάς τους είναι <η άπληστη τεχνική>, <τυχαία επιλογή βάσει πιθανοτήτων>, <τεχνικές βασισμένες στα διαστήματα>.

2.2 Το μοντέλο της ενισχυτικής μάθησης

Σύμφωνα με το κλασικό μοντέλο ενισχυτικής μάθησης ένας πράκτορας συνδέεται στο περιβάλλον μέσω της αντίληψης και της δράσης, όπως απεικονίζεται στο σχήμα που ακολουθεί.



Σχήμα 2.2: Κλασικό μοντέλο ενισχυτικής μάθησης

Σε κάθε βήμα της αλληλεπίδρασης με το περιβάλλον ο πάκτορας δέχεται ως είσοδο i κάποια ένδειξη της παρούσας κατάστασης s του περιβάλλοντος. Έπειτα, ο πράκτορας επιλέγει μια δράση α ώστε να δημιουργήσει έξοδο προς το περιβάλλον. Η δράση αυτή αλλάζει την κατάσταση του περιβάλλοντος και η αποτίμηση αυτής της αλλαγής κατάστασης επιστρέφεται στον πράκτορα ως ενίσχυση r. Ο πράκτορας Β πρέπει να επιλέγει δράσεις οι οποίες τείνουν να αυξάνουν το μακροπρόθρεσμο άθροισμα των τιμών των ενισχυτικών σημάτων. Κάτι τέτοιο γίνεται εφικτό έπειτα από συστηματικές προσπάθειες καθοδηγούμενες από μεγάλη ποικιλία αλγορίθμων.

Τυπικά, το μοντέλο αποτελείται από:

- Ένα διακριτό σύνολο καταστάσεων περιβάλλοντος, S
- Ένα διακριτό σύνολο ενεργειών του πράκτορα, A
- Ένα σύνολο ενισχυτικών σημάτων που συνήθως είναι το σύνολο (0,1) ή το σύνολο πραγματικών αριθμών

Στο Σχ.2.2 φαίνεται η συνάρτηση εισόδου I , η οποία καθορίζει πώς ο πράκτορας αντιλαμβάνεται την εκάστοτε κατάσταση του περιβάλλοντος. Συνήθως, θεωρείται

ότι αυτή είναι και η ταυτοτική συνάρτηση , ότι δηλαδή ο πράκτορας έχει αντίληψη της ακριβούς κατάστασης του περιβάλλοντος. Ένας διαισθητικός τρόπος για την κατανόηση της σχέσης μεταξύ περιβάλλοντος και πράκτορα, είναι ο ακόλουθος.

Περιβάλλον:Είμαι στην κατάσταση 65,Έχεις 4 πιθανές δράσεις.

Πράκτορας:Θα διαλέξω τη δράση 2.

Περιβάλλον:Έλαβες ενίσχυση 7 μονάδων.Τώρα είσαι στην κατάσταση 15 και έχεις 2 πιθανές δράσεις.

Πράκτορας:Θα διαλέξω τη δράση 1

Περιβάλλον:Έλαβες ενίσχυση -4 μονάδων .Τώρα είσαι στην κατάσταση 65 και έχεις 4 πιθανές δράσεις.

Πράκτορας:Θα διαλέξω τη δράση 2

Περιβάλλον:Έλαβες ενίσχυση 5 μονάδων.Τώρα είσαι στην κατάσταση 44 και έχεις 5 πιθανές δράσεις.

...

Σχήμα 2.3:Υποθετικός διάλογος μεταξύ πράκτορα και περιβάλλοντος

Ο στόχος του πράκτορα είναι, αντιστοιχίζοντας καταστάσεις σε δράσεις, να προσδιορίσει την πολιτική π που μεγιστοποιεί το μακροπρόθεσμο μέγεθος της ενίσχυσης. Ακριβώς για το λόγο αυτό, ο πράκτορας δύναται να προτιμήσει αντί για άμεση ενίσχυση να λάβει μεγαλύτερη ανταμοιβή αργότερα. Γενικά, το περιβάλλον αναμένεται να είναι τυχαίο, δηλαδή η επιλογή της ίδιας δράσης στην ίδια κατάσταση σε δύο διαφορετικές περιστάσεις μπορεί να οδηγήσει σε διαφορετικές επόμενες καταστάσεις ή και διαφορετικές τιμές του σήματος ενίσχυσης. Αυτό παρατηρείται και στο διάλογο του Σχ.2.3 όπου από την κατάσταση 65 και επιλέγοντας την δράση 2 ανακύπτουν δύο διαφορετικές περιπτώσεις επόμενων καταστάσεων και σημάτων ενίσχυσης. Ωστόσο, θεωρείται στατικό περιβάλλον, δηλαδή οι πιθανότητες μετάβασης κατάστασης ή λήψης σημάτων ενίσχυσης δεν μεταβάλλονται με το χρόνο.

2.3 Στοιχεία ενισχυτικής μάθησης

Πέντε είναι τα βασικά στοιχεία ενός συστήματος ενισχυτικής μάθησης (reinforcement learning):

1. *Ο πράκτορας (agent):* Είναι αυτός ο οποίος μαθαίνει μέσω του συστήματος ενισχυτικής μάθησης. Ανάλογα με την κατάσταση του περιβάλλοντος στην οποία βρίσκεται, ο πράκτορας μπορεί να εκτελεί κάποιες ενέργειας $\alpha \in A$, όπου το A είναι το επιτρεπτό σύνολο ενεργειών.

2. *Το μοντέλο του περιβάλλοντος (model of the environment):* Μιμείται τη συμπεριφορά του περιβάλλοντος, δηλαδή, δοθείσας μιας κατάστασης και μιας ενέργειας, καθορίζει ποιά θα είναι η επόμενη ενέργεια. Το περιβάλλον περιγράφεται από ένα σύνολο καταστάσεων $s \in S$, όπου S το σύνολο των καταστάσεων του περιβάλλοντος.

3. *Η στρατηγική-πολιτική (Policy):* Πολιτική είναι ο τρόπος που καθορίζεται η επιλογή ενέργειας ή άλλως της τακτικής όπου ακολουθείται. Συχνά, χρησιμοποιούνται τρείς πολιτικές για επιλογή ενέργειας με στόχο την εξισορρόπηση μεταξύ της εκμετάλλευσης της προηγούμενης γνώσης και της εξερεύνσης νέων ενεργειών. Οι πολιτικές αυτές είναι:

1) ε-greedy, ή άλλως άπληστη στρατηγική, κατα την οποία τις περισσότερες φορές επιλέγεται η ενέργεια με τη μεγαλύτερη εκτιμώμενη ανταμοιβή, η οποία ενέργεια ονομάζεται ως η πλέον άπληστη ενέργεια. Με μια μικρή πιθανότητα ε, μια ενέργεια επιλέγεται τυχαία. Η ενέργεια επιλέγεται ομοιόμορφα, ανεξάρτητα από τις εκτιμήσεις ενέργειας-αξίας. Αυτή η μέθοδος εξασφαλίζει ότι εφόσον εκτελεστούν πολλές επαναλήψεις κάθε ενέργεια θα δοκιμαστεί αρκετές φορές. Συνεπώς η στρατηγική αυτή εγγυάται ότι θα ανακαλυφθούν βέλτιστες ενέργειες.

2) ε-soft, η οποία είναι παρόμοια με την άπληστη πολιτική. Η καλύτερη ενέργεια επιλέγεται με πιθανότητα 1-ε και τον υπόλοιπο χρόνο επιλέγεται ομοιόμορφα μια τυχαία ενέργεια. Μειονέκτημα των πολιτικών ε-greedy και ε-soft αποτελεί το ότι επιλέγουν τυχαίες ενέργειες ομοιόμορφα με αποτέλεσμα η χειρότερη ενέργεια και η δεύτερη καλύτερη μπορούν να επιλεγούν με την ίδια πιθανότητα.

3) Softmax ή τυχαιοποιημένη στρατηγική η οποία δεν έχει το μειονέκτημα των δύο προηγούμενων αφού ταξινομεί κάθε ενέργεια ανάλογα με την εκτίμηση ενέργειας-αξίας που της αποδίδεται. Μια τυχαία ενέργεια επιλέγεται έχοντας υπόψη τη βαρύτητα που προσάπτεται σε κάθε ενέργεια, κάτι το οποίο σημαίνει ότι υπάρχει πολύ μικρή πιθανότητα να επιλεγούν οι χειρότερες ενέργειες. Η πλέον συνήθης softmax μέθοδος επιλέγει την ενέργεια τη χρονική στιγμή t ακολουθώντας την κατανομή Boltzmann:

$$P(a) = \frac{e^{Q_t(a)/\tau}}{\sum_{b=1}^{n} e^{Q_t(b)/\tau}}, \qquad (2.1)$$

όπου τ θετική παράμετρος η οποία ονομάζεται θερμοκρασία, $Q_t(a)$ η εκτιμημένη αξία της ενέργειας α τη χρονική στιγμή t και n είναι ο αριθμός των διαθέσιμων ενεργειών.

4. *Η συνάρτηση ανταμοιβής (reward function):* Ορίζει το στόχο του προβλήματος ενισχυτικής μάθησης. Αντιστοιχίζει κάθε ζεύγος κατάστασης-ενέργειας σε μια τιμή ανταμοιβής η οποία ποσοτικοποιεί την επίπτωση της συγκεκριμένης επιλογής. Ο πράκτορας μαθαίνει να εκτελεί εκείνες τις ενέργειες οι οποίες μακροπρόθεσμα μεγιστοποιούν το άθροισμα των ανταμοιβών. Η συνάρτηση ανταμοιβής καθορίζει κατα κάποιο τρόπο ποιές από τις καταστάσεις στις οποίες αναμένεται να βρεθεί ο πράκτορας (agent) είναι καλές. Αυτό μπορεί να αποτελέσει βάση για αλλαγή της τακτικής όπου ακολουθείται.

5. *Η συνάρτηση χρησιμότητας (value function):* Καθορίζει ποιές κινήσεις είναι καλές μακροπρόθεσμα. Συγκεκριμένα, η τιμή μιας κατάστασης υπολογίζεται ως το άθροισμα των ανταμοιβών που αναμένεται να συγκεντρώσει ο πράκτορας μέχρι το τέλος του παιγνίου. Σε αντίθεση με τις τιμές της συνάρτησης ανταμοιβής οι οποίες εκφράζουν την προσωρινή αξία μιας κατάστασης του περιβάλλοντος, οι τιμές της συνάρτησης χρησιμότητας εκφράζουν την μακροπρόθεσμη αξία μιας κατάστασης λαμβάνοντας υπόψη τις καταστάσεις που ενδέχεται να προκύψουν στη συνέχεια. Για παράδειγμα, μια κατάσταση μπορεί να αποδίδει χαμηλή τιμή ανταμοιβής αλλά να έχει υψηλή τιμή χρησιμότητας καθώς ακολουθείται από καταστάσεις οι οποίες αποδίδουν υψηλές ανταμοιβές ή και αντιστρόφως.

Ένα βασικό ερώτημα που τίθεται κατά τον υπολογισμό της χρησιμότητας είναι αν υπάρχει πεπερασμένος ορίζοντας (finite horizon) ή άπειρος ορίζοντας (infinite horizon) για τη λήψη αποφάσεων. Πεπερασμένος ορίζοντας σημαίνει ότι έχει τεθεί σταθερός χρόνος N, μετά τον οποίο ο πράκτορας δεν λειτουργεί. Συνεπώς $V([s_0,s_{1,\dots,}s_{N+k}])=V([s_0,s_1, \dots,S_N])$ για κάθε k>0. Στην περίπτωση προβλημάτων πεπερασμένου ορίζοντα η βέλτιστη ενέργεια υπό δεδομένη κατάσταση μπορεί να μεταβάλλεται με το χρόνο, σε αντίθεση με τις πολιτικές απείρου ορίζοντα όπου δεν υπάρχει διαφορετική συμπεριφορά στην ίδια κατάσταση σε διαφορετικούς χρόνους. Υποθέτοντας ότι οι προτιμήσεις ενός πράκτορα ως προς τις ακολουθίες καταστάσεων είναι στάσιμες, αποδεικνύεται ότι υπάρχουν μόνο δύο τρόποι απόδοσης χρησιμοτήτων στις ακολουθίες καταστάσεων.

- Προσθετικές ανταμοιβές (additive rewards): Η χρησιμότητα μιας ακολουθίας καταστάσεων είναι
  $V([s_0,s_1,\dots s_2,\dots])=R(s_0) + R(s_1) + R(s_2) +\dots$ (2.2)

- Προεξοφλημένες ανταμοιβές (discounted rewards): Η χρησιμότητα μιας ακολουθίας καταστάσεων είναι

$$V([s_0,s_1,s_2,\ldots])=R(s_0)+\gamma R(s_1)+\gamma^2 R(S_2)+\ldots \quad (2.3)$$

όπου ο παράγοντας ελάττωσης (discount factor) γ είναι ένας αριθμός μεταξύ του 0 και του 1. Ο παράγοντας προεξόφλησης περιγράφει την προτίμηση του πράκτορα για τρέχουσες ανταμοιβές έναντι μελλοντικών ανταμοιβών. Ειδικά, όταν το γ τείνει στο 0, οι ανταμοιβές του μακρινού μέλλοντος θεωρούνται αμελητέες.

2.4 Διαδικασία απόφασης Markov (MDP)

Στο πρόβλημα της ενισχυτικής μάθησης, ο πράκτορας οφείλει να μαθαίνει από καθυστερημένη ενίσχυση (delayed reinforcement). Δηλαδή υπάρχει το ενδεχόμενο να χρειαστεί μια μακρά ακολουθία ενεργειών, από τις οποίες θα λάβει ασήμαντη ενίσχυση, μέχρι να φθάσει τελικά σε κατάσταση υψηλής ενίσχυσης. Τέτοιου είδους προβλήματα καθυστερημένης ενίσχυσης για ένα πλήρως παρατηρήσιμο περιβάλλον περιγράφονται αποτελεσματικά ως διαδικασίες απόφασης Markov (Markov Decision process-MDP). Μια διαδικασία απόφασης Markov ορίζεται από τις ακόλουθες συνιστώσες.

- σύνολο καταστάσεων S
- σύνολο ενεργειών A
- συνάρτηση ανταμοιβής $R : S \times A \rightarrow \mathfrak{M}$
- συνάρτηση μετάβασης καταστάσεων $T : S \times A \rightarrow \Pi(\Sigma)$

Η πιθανότητα μετάβασης από μια κατάσταση s στην κατάσταση s΄ εκτελώντας μια ενέργεια α ορίζεται ως $T(s,a,s')$. Η συνάρτηση ανταμοιβής προσδιορίζει την άμεση αναμενόμενη ανταμοιβή ως συνάρτηση της τρέχουσας κατάστασης και της ενέργειας που επιλέγεται. Το μοντέλο ονομάζεται Markov εφόσον οι πιθανότητες είναι ανεξάρτητες από τις προηγούμενες καταστάσεις του περιβάλλοντος και τις προηγούμενες ενέργειες των πρακτόρων

2.4.1 Εύρεση πολιτικής για δεδομένο μοντέλο Markov

Ξεκινώντας από μια αυθαίρετη αρχική κατάσταση s_t η αναμενόμενη τιμή της συσσωρευμένης ενίσχυσης που επιτυγχάνεται από μια αυθαίρετη πολιτική π, δίδεται από τη σχέση:

$$V^{\pi}(s_t)=E[\sum_{i=1}^{\infty}\gamma^{i}r_{t+1}] \tag{2.4}$$

όπου το r_{t+1} είναι είναι η ανταμοιβή η οποία επιτυγχάνεται επιλέγοντας μια δράση από την κατάσταση s_{t+1} χρησιμοποιώντας την πολιτική π και $\gamma \in [0,1)$. Η αναμενόμενη τιμή (συμβολίζεται με Ε) είναι απαραίτητη διότι οι ανταμοιβές μπορεί να είναι μη ντετερμινιστικές. Αν γ=0, μόνο οι άμεσες ανταμοιβές λαμβάνονται υπόψη. Όσο το γ πλησιάζει στο 1 αποδίδεται έμφαση στις μελλοντικές ανταμοιβές. Η συνάρτηση V^{π} ονομάζεται συνάρτηση χρησιμότητας (utility function) για την πολιτική π.

Η διαδικασία μάθησης, όπως αντιμετωπίζεται από ένα πράκτορα ενισχυτικής μάθησης, συνήθως θεωρείται ως μια αλυσίδα αποφάσεων Markov . Έτσι η ανταμοιβή και η νέα κατάσταση ορίζονται ως εξής:

$$\begin{aligned} R_t &= R(s_t, a_t) \\ s_{t+1} &= \delta(s_t, a_t) \end{aligned} \tag{2.5}$$

όπου η συνάρτηση ανταμοιβής R και η συνάρτηση μετάβασης καταστάσεων δ μπορεί να είναι μη ντετερμινιστικές. Αν ο πράκτορας γνώριζε τη συνάρτηση βέλτιστης χρησιμότητας V*, τις πιθανότητες μετάβασης μεταξύ καταστάσεων και τις αναμενόμενες ανταμοιβές, θα μπορούσε εύκολα να ορίσει τη βέλτιστη δράση εφαρμόζοντας την αρχή της μέγιστης αναμενόμενης χρησιμότητας, δηλαδή μεγιστοποιώντας το άθροισμα της αναμενόμενης άμεσης ανταμοιβής και της αναμενόμενης τιμής της επόμενης κατάστασης, η οποία δείχνει τις αναμενόμενες ανταμοιβές από το σημείο αυτό και μετά.

$$\begin{aligned} &\pi^*(s) = \delta(s_t, a_t)\pi^*(s) = \operatorname{argmax}_{\alpha \in A} E[R(s,a) + \gamma V^*(\delta(s,a))] = \\ &\operatorname{argmax}_{\alpha \in A}(E[R(s,a)] + \gamma \sum_{s' \in S}^{n} T(s,a,s')V^*(s')), \forall s \in S \end{aligned} \tag{2.6}$$

όπου το T(s,a,s΄) συμβολίζει την πιθανότητα μετάβασης από την κατάσταση s στη κατάσταση s΄ επιλέγοντας τη δράση α.

Οι χρησιμότητες της μιας κατάστασης και της επόμενής της ορίζονται ως:

$$V^*(s) = E[R(s,\pi^*(s)) + \gamma V^*(\delta(s,a))] = \max_a (E[R(s,a)] + \gamma \sum_{s' \in S}^{n} T(s,a,s') V^*(s')), \forall s \in S \quad (2.7)$$

Οι τελευταίες εξισώσεις είναι γνωστές ως εξισώσεις Bellman. Μέσω της επίλυσης αυτών προκύπτει μοναδική χρησιμότητα για κάθε κατάσταση. Λόγω της παρουσίας του συντελεστή <max> οι εξισώσεις είναι μη γραμμικές και επομένως δύσκολο να λυθούν. Συνήθως επιλύονται με εφαρμογή τεχνικών δυναμικού προγραμματισμού όπως επανάληψη αξιών (value iteration) και επανάληψη πολιτικής (policy iteration).

2.4.2 Επανάληψη αξιών (value iteration)

Ένας τρόπος εύρεσης της βέλτιστης πολιτικής είναι ο υπολογισμός της βέλτιστης συνάρτησης χρησιμότητας με τη βοήθεια του αλγορίθμου επανάληψης αξιών. Η βασική ιδέα είναι να υπολογιστεί η χρησιμότητα κάθε κατάστασης με βάση την εξίσωση Bellman (2.7) και να χρησιμοποιηθούν οι χρησιμότητες αυτές για την επιλογή της βέλτιστης ενέργειας σε κάθε κατάσταση. Αν υπάρχουν n δυνατές καταστάσεις, υπάρχουν n εξισώσεις Bellman με n αγνώστους. Το σύστημα των εξισώσεων που προκύπτουν μπορεί να λυθεί δοκιμάζοντας μια επαναληπτική διαδικασία προσέγγισης όταν επιτευχθεί ισορροπία στις τιμές δύο διαδοχικών συναρτήσεων χρησιμότητας. Η περιγραφή του αλγορίθμου με ψευδοκώδικα είναι η ακόλουθη.

function Value-Iteration (mdp, ε) **returns** μια συνάρτηση χρησιμότητας

inputs: mdp, MDP με καταστάσεις S, μοντέλο μετάβασης T, συνάρτηση ανταμοιβής R, προεξόφληση γ, ε: το μέγιστο σφάλμα το οποίο επιτρέπεται για την χρησιμότητα κάθε κατάστασης

local variables: V , V' , διανύσματα χρησιμοτήτων για τις καταστάσεις στο S (αρχικά μηδενικά), δ: η μέγιστη μεταβολή στη χρησιμότητα οποιασδήποτε κατάστασης σε μια επανάληψη

repeat

V ← V':δ ← 0

for each κατάσταση s in S **do {**

for each ενέργεια α ∈ A **do**

$$Q(s,a) := R(s,a) + \gamma \sum_{s' \in S} T(s,a,s')\ V(s')$$

V'(s):=maxQ(s,a) }

if |V*[s]-V[s]|>δ **then** δ ← |V'[s]-V[s]|

until δ<ε(1-γ)/γ

return V

Το κριτήριο τερματισμού του αλγορίθμου προέρχεται από την εργασία των Williams και Baird. Σύμφωνα με το κριτήριο αυτό, εφόσον η μέγιστη διαφορά μεταξύ δύο διαδοχικών συναρτήσεων χρησιμότητας είναι μικρότερη από ε, η χρησιμότητα της άπληστης πολιτικής, δηλαδή η επιλογή για κάθε κατάσταση της ενέργειας η οποία μεγιστοποιεί την αναμενόμενη προεξοφλημένη ανταμοιβή, διαφέρει από τη βέλτιστη χρησιμότητα λιγότερο από 2ε(1-γ)/γ για κάθε κατάσταση. Ο απαιτούμενος αριθμός επεναλήψεων ώστε να προκύψει η βέλτιστη συνάρτηση χρησιμότητας είναι πολυωνυμικός ως προς το πλήθος των καταστάσεων και το μέγεθος της μέγιστης ανταμοιβής όταν ο παράγοντας προεξόφλησης παραμένει σταθερός. Ωστόσο, στη χειρότερη περίπτωση ο αριθμός των επαναλήψεων αυξάνεται πολυωνυμικά σε 1/(1-γ) και ο ρυθμός σύγκλισης μειώνεται αισθητά καθώς ο παράγοντας προεξόφλησης πλησιάζει το 1.

2.4.3 Επανάληψη πολιτικών (policy iteration)

Ένας εναλλακτικός τρόπος εύρεσης βέλτιστων πολιτικών είναι ο αλγόριθμος επανάληψης πολιτικών που επαναλαμβάνει τα ακόλουθα δύο βήματα ξεκινώντας από κάποια αρχική πολιτική π_0 :

- Αξιολόγηση πολιτικής (policy evaluation): με δεδομένη μια πολιτική π_i, υπολογίζεται το $V_i=V^{\pi i}$, δηλαδή η χρησιμότητα για κάθε κατάσταση υπό την υπόθεση ότι θα εκτελεστεί η πολιτική π_i.
- Βελτίωση πολιτικής: υπολογίζεται μια νέα πολιτική π_{i+1} , χρησιμοποιώντας προεξέταση ενός βήματος με βάση την V_i.

Ο αλγόριθμος τερματίζεται όταν το βήμα βελτίωσης της πολιτικής δεν προκαλεί μεταβολή των χρησιμοτήτων. Η περιγραφή του αλγορίθμου με ψευδοκώδικα είναι η ακόλουθη.

function Policy-Iteration (mdp) **returns** μια τακτική

inputs: mdp, MDP με καταστάσεις S, μοντέλο μετάβασης T

local variables: V , V′ , διανύσματα χρησιμοτήτων για τις καταστάσεις στο S (αρχικά μηδενικά), π : ένα διάνυσμα πολιτικής το οποίο δεικτοδοτείται από τις καταστάσεις (αρχικά τυχαίο)

repeat

V ← Policy_Evaluation(π,V,mdp)

ΥΠΑΡΞΗ_ΒΕΛΤΙΩΣΗΣ $\leftarrow$ ΨΕΥΔΕΣ

for each κατάσταση s in S **do**

if $\max \sum_{s' \in S}^{n} T(s,a,s')V(s') > \sum_{s' \in S}^{n} T(s,\pi[s],s')V(s')$ **then**

{ $\pi[s] \leftarrow \arg\max \sum_{s' \in S} T(s,a,s')V(s')$

ΥΠΑΡΞΗ_ ΒΕΛΤΙΩΣΗΣ $\leftarrow$ ΑΛΗΘΕΣ }

until ΥΠΑΡΞΗ_ ΒΕΛΤΙΩΣΗΣ==ΨΕΥΔΕΣ

return π

Πρακτικά, το trade off μεταξύ των δύο αλγορίθμων είναι το εξής: Ο αλγόριθμος επανάληψης αξιών εκτελεί πολύ συντομότερες επαναλήψεις αλλά πολλές σε πλήθος, ενώ ο αλγόριθμος επανάληψης πολιτικών απαιτεί λιγότερες επαναλήψεις αλλά μεγαλύτερης διάρκειας.

2.4.4 Μαθαίνοντας μια βέλτιστη πολιτική

Προηγουμένως αναφέρθηκαν οι μέθοδοι για την εύρεση μιας βέλτιστης πολιτικής για μια αλυσίδα markov, υποθέτοντας ότι υπάρχει το μοντέλο. Το μοντέλο ουσιαστικά περιλαμβάνει τη συνάρτηση πιθανότητας μετάβασης T(s,a,s΄) και τη συνάρτηση ανταμοιβής R(s,a). Η ενισχυτική μάθηση ασχολείται κατά κύριο λόγο με το πώς θα προσδιοριστεί η βέλτιστη πολιτική χωρίς να υπάρχει γνώση του μοντέλου αυτού. Ο πράκτορας πρέπει να αλληλεπιδράσει άμεσα με το περιβάλλον ώστε να αποκτήσει πληροφορίες οι οποίες με τη βοήθεια κατάλληλου αλγορίθμου μπορούν να χρησιμοποιηθούν για την εύρεση της βέλτιστης πολιτικής. Με βάση την ύπαρξη ή όχι μοντέλου, υπάρχουν δύο μέθοδοι για την επίτευξη του στόχου:

- Χωρίς μοντέλο: όπου ο πράκτορας εκπαιδεύεται χωρίς να διαθέτει κάποιο μοντέλο
- Με τη χρήση μοντέλου: όπου πρώτα διατυπώνεται ένα μοντέλο και, στη συνέχεια, αυτό χρησιμοποιείται για την εκπαίδευση του πράκτορα

Ποιά προσέγγιση είναι η καλύτερη; Αυτό το ερώτημα έχει δημιουργήσει έντονη διαμάχη στην κοινότητα που ερευνά την ενισχυτική μάθηση. Ένας σημαντικός αριθμός αλγορίθμων έχει προταθεί και από τις δύο πλευρές.

2.4.5 Μέθοδοι βασισμένες σε κάποιο μοντέλο

Παραδείγματα αλγορίθμων της κατηγορίας αυτής είναι:

Μέθοδος ισοδύναμης αβεβαιότητας: Πρώτα προσδιορίζονται οι συναρτήσεις T και R εξερευνώντας το περιβάλλον και διατηρώντας στατιστικά στοιχεία για το αποτέλεσμα κάθε δράσης. Έπειτα υπολογίζεται η βέλτιστη πολιτική με κάποια από τις μεθόδους επανάληψης πολιτικών ή επανάληψης αξιών. Οι ενστάσεις που έχουν τεθεί για αυτή τη μέθοδο είναι αρκετές:

- Πραγματοποιεί αυθαίρετη διάκριση μεταξύ της φάσης μάθησης και της φάσης δράσης
- Τίθεται το ερώτημα, πώς θα συγκεντρώσει αρχικά δεδομένα για το περιβάλλον; Η τυχαία εξερεύνηση μπορεί να είναι επικίνδυνη για εξαγωγή εσφαλμένων στοιχείων και σε μερικά περιβάλλοντα αποτελεί υπερβολικά αναξιόπιστη μέθοδο συλλογής δεδομένων

Μια παραλλαγή της μεθόδου αυτής περιλαμβάνει τη συνεχή μάθηση του μοντέλου σε όλη τη διάρκεια της ζωής του πράκτορα και σε κάθε βήμα. Το μοντέλο αυτό χρησιμοποιείται για τον υπολογισμό της βέλτιστης πολιτικής και της συνάρτησης ανταμοιβής. Αυτή η μέθοδος κάνει αρκετά αποτελεσματική χρήση των δεδομένων του περιβάλλοντος για την εύρεση της βέλτιστης πολιτικής αλλά εξακολουθεί να αγνοεί την ερώτηση της εξερεύνησης και είναι ιδιαίτερα απαιτητική υπολογιστικά, ακόμα και για μικρούς χώρους καταστάσεων.

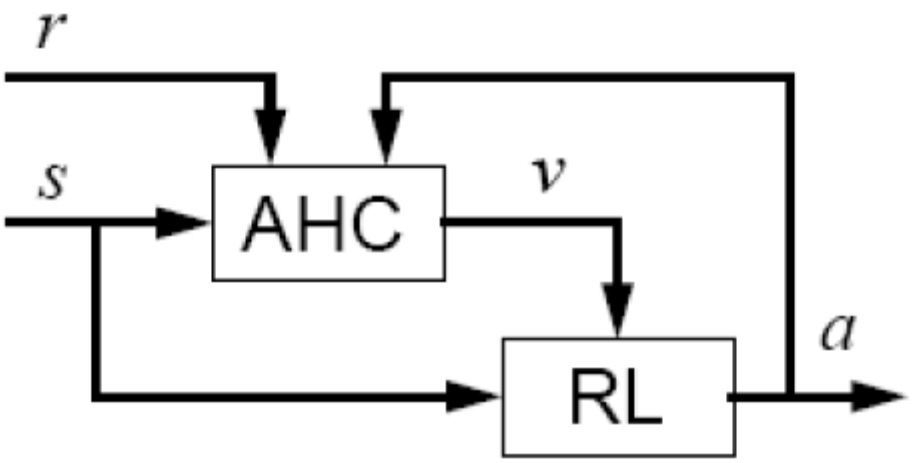
Μέθοδος Dyna: Η αρχιτεκτονική του Dyna χρησιμοποιεί στρατηγικές οι οποίες είναι και αποτελεσματικότερες σε σχέση με τη μάθηση χωρίς μοντέλο αλλά και αποδοτικότερες υπολογιστικά σε σχέση με τον προηγούμενο αλγόριθμο της μεθόδου ισοδύναμης αβεβαιότητας. Χρησιμποποιεί την εμπειρία τόσο για να διατυπώσει το μοντέλο (συναρτήσεις T και R) όσο και για να προσαρμόσει την πολιτική. Πέρα των ανωτέρω μεθόδων, έχουν προταθεί και άλλες βελτιωμένες μέθοδοι (βασισμένες σε κάποιο μοντέλο) όπως η ουρά-Dyna, η εκκαθάριση με προτεραιότητα κλπ.

2.4.6 Μέθοδοι χωρίς τη χρήση μοντέλου

Το σημαντικότερο πρόβλημα το οποίο αντιμετωπίζεται κατά την ενισχυτική μάθηση είναι η ανάθεση προσωρινής ανταμοιβής. Πώς μπορεί να γίνει γνωστό αν μια δράση που επιλέχθηκε θα αποφέρει καλά μακροπρόθεσμα αποτελέσματα; Μια στρατηγική είναι η αναμονή ως το τέλος ώστε να ανταμειφθούν οι δράσεις ανάλογα με το αν το αποτέλεσμα ήταν καλό ή κακό. Σε διάφορες διεργασίες όμως είναι δύσκολο να γίνει γνωστό πότε έρχεται το τέλος και κάτι τέτοιο ίσως απαιτεί μεγάλο ποσοστό μνήμης. Έτσι χρησιμοποιείται η γνώση από την επανάληψη αξιών ώστε να

προσαρμοστεί η αναμενόμενη χρησιμότητα μιας κατάστασης Οι αλγόριθμοι αυτοί είναι γνωστοί ως μέθοδοι χρονικής διαφοράς (temporal difference methods, TD).

Προσαρμοστική Ευριστική Κριτική (Adaptive Heuristic Critic, AHC) και TD(λ): Ο αλγόριθμος προσαρμοστικής ευριστικής κριτικής είναι μια προσαρμοσμένη έκδοση της επανάληψης πολιτικών στην οποία ο υπολογισμός της συνάρτησης ανταμοιβής υπολογίζεται μέσω αλγορίθμου που ονομάζεται TD(0), ένα λειτουργικό διάγραμμα του οποίου φαίνεται στο Σχ.2.4:



Σχήμα 2.4: Αρχιτεκτονική αλγορίθμου προσαρμοστικής ευρIστικής κριτικής

Ουσιαστικά υπάρχουν δύο λειτουργικές δράσεις: η μία είναι ο κριτής, (συμβολίζεται με AHC) και η άλλη είναι η συνιστώσα της ενισχυτικής μάθησης (συμβολίζεται με RL). Η συνιστώσα RL λειτουργεί όχι για να μεγιστοποιήσει τη στιγμιαία ανταμοιβή αλλά για να μεγιστοποιήσει την ευριστική χρησιμότητα ν η οποία υπολογίζεται από τον κριτή. Ο κριτής χρησιμποποιεί το πραγματικό εξωτερικό σχήμα της ενίσχυσης για να μάθει να αντιστοιχίζει τις καταστάσεις με τις αναμενόμενες χρησιμότητες λαμβάνοντας υπόψη τον παράγοντα προεξόφλησης, με δεδομένο ότι η πολιτική η οποία ακολουθείται είναι αυτή που υπάρχει τη δεδομένη χρονική στιγμή στη συνιστώσα RL.

Η αναλογία με την προσαρμοσμένη επανάληψη πολιτικών φαίνεται στις δύο συνιστώσες οι οποίες λειτουργούν εναλλακτικά. Η πολιτική π, που υλοποιείται από την RL, τροποποιείται και ο κριτής μαθαίνει τη συνάρτηση ανταμοιβής V_π για αυτή την πολιτική. Στη συνέχεια, τροποποιείται ο κριτής και η συνιστώσα RL μαθαίνει τη νέα πολιτική π΄ η οποία μεγιστοποιεί τη νέα συνάρτηση χρησιμότητας κ.ο.κ. Στις περισσότερες υλοποιήσεις, οι δύο λειτουργικές δράσεις λειτουργούν παράλληλα. Μόνο όμως η εναλλακτική υλοποίηση μπορεί να εγγυηθεί τη σύγκλιση στη βέλτιστη πολιτική, κάτω από τις κατάλληλες συνθήκες.

Απομένει να εξηγηθεί με ποιό τρόπο ο κριτής μπορεί να μάθει τη χρησιμότητα μιας πολιτικής. Ορίζεται ως (s,a,r,s΄) μια πλειάδα εμπειρίας η οποία περικλείει μια απλή μετάβαση στο περιβάλλον, όπου s είναι η κατάσταση του πράκτορα πριν τη μετάβαση, a είναι η επιλογή δράσης, r η άμεση ανταμοιβή και s΄ η νέα κατάσταση. Η ανταμοιβή μιας πολιτικής προσδιορίζεται με τη χρήση του αλγορίθμου TD(0) του Sutton που χρησιμοποιεί τον εξής κανόνα ενημέρωσης:

$$V(s):=V(s) + a(r+\gamma V(s')-V(s)) \qquad (2.8)$$

Όποτε ο πράκτορας βρεθεί σε μια κατάσταση s, η αναμενόμενη τιμή της (2.8) ενημερώνεται ώστε να είναι εγγύτερα στο r+γV(s'), αφού το r είναι η άμεση ανταμοιβή και το V(s') η αναμενόμενη χρησιμότητα της επόμενης κατάστασης. Αυτή η διαδικασία είναι παρόμοια με τον απλό κανόνα επανάληψης αξιών με τη μόνη διαφορά ότι η γνώση προέρχεται από τον πραγματικό κόσμο και όχι από κάποιο γνωστό μοντέλο. Η ιδέα-κλειδί είναι ότι το r+γV(s') είναι ένα δείγμα της τιμής V(s) και ότι είναι αρκετά πιθανό να είναι σωστή τιμή διότι ενσωματώνει την πραγματική ενίσχυση r. Αν ο ρυθμός μάθησης a προσαρμόζεται κατάλληλα (πρέπει να μειώνεται αργά) και η πολιτική τροποποιείται κατάλληλα, ο αλγόριθμος TD(0) εγγυάται σύγκλιση στη βέλτιστη συνάρτηση χρησιμότητας.

Ο αλγόριθμος TD(0) αποτελεί ουσιαστικά παραλλαγή μιάς γενικής τάξης αλγορίθμων που καλούνται TD(λ) για λ=0. Ο TD(0) παρατηρεί μόνο ένα βήμα μπροστά όταν προσδιορίζει τις αναμενόμενες χρησιμότητες. Αν και λογικά θα καταλήξει στη σωστή απάντηση, υπάρχει το ενδεχόμενο να απαιτήσει αρκετό χρόνο. Ο γενικότερος κανόνας TD(λ) είναι ο ακόλουθος:

$$V(u):=V(u) + a(r+\gamma V(s')-V(s))e(u) \qquad (2.9)$$

αλλά εφαρμόζεται σε κάθε κατάσταση ανάλογα με την καταλληλότητα της e(u). Κατάλληλος ορισμός της οποίας θα μπορούσε να είναι ο ακόλουθος:

$$e(s)=\sum_{k=1}^{t}(\lambda\gamma)^{t-k}\delta_{s,s_k}\ ,\quad \text{όπου}\ \ \delta_{s,s_k}=\begin{cases}1, s=s_k\\ 0, \alpha\lambda\lambda ού\end{cases} \qquad (2.10)$$

Η καταλληλότητα της κατάστασης s προσδιορίζεται από τον αριθμό των φορών όπου έχει βρεθεί σε αυτή ο πράκτορας στο πρόσφατο παρελθόν. ΄Οταν μια ενίσχυση ληφθεί, τ χρησιμοποιείται για να ενημερωθούν όλες οι καταστάσεις από τις οποίες έχει περάσει ο πράκτορας πρόσφατα, σύμφωνα με την καταλληλότητά τους. Όταν λ=0 προκύπτει ο αλγόριθμος TD(0). Όταν λ=1, ενημερώνονται όλες οι καταστάσεις ανάλογα με τον αριθμό των φορών που έχει βρεθεί ο πράκτορας σε αυτές στο τέλος ενός τρεξίματος του αλγορίθμου. Η ενημέρωση της καταλληλότητας μπορεί να γίνει εκατά τη λειτουργία του αλγορίθμου (on line) ως ακολούθως:

$$e(s) := \begin{cases} \gamma\lambda e(s) + 1, s = \pi\alpha\rho ούσα \\ \gamma\lambda e(s), \alpha\lambda\lambda ιώς \end{cases} \quad (2.11)$$

Είναι υπολογιστικά περισσότερο δαπανηρό να εκτελεστεί ο TD(λ), αν και συχνά συγκλίνει πολύ ταχύτερα για μεγάλα λ. Οι λειτουργίες των δύο λειτουργικών συνιστωσών του αλγορίθμου μπορούν να εκτελεστούν με ενοποιημένο τρόπο από τον αλγόριθμο της μάθησης Q του Watkins.

2.4.7 Εισαγωγή στη μάθηση Q

Η μάθηση Q (Q learning) είναι ένας αλγόριθμος ενισχυτικής μάθησης που μαθαίνει τις τιμές μιας συνάρτησης Q(s,a) ώστε να προσδιορίσει τη βέλτιστη πολιτική π υπό συγκεκριμένο κριτήριο. Οι τιμές της συνάρτησης Q(s,a) αποδίδουν το όφελος από την επιλογή μιας συγκεκριμένη δράσης σε μια συγκεκριμένη κατάσταση. Η συνάρτηση Q(s,a) ορίζεται ως η ανταμοιβή που δίνεται άμεσα εκτελώντας τη δράση α από την κατάσταση s επαυξημένη κατά την τιμή των προεξοφλημένων ανταμοιβών που θα ληφθούν στο μέλλον αν ακολουθηθεί η βέλτιστη πολιτική.

Για κάθε χρονική στιγμή t ορίζεται ως Q*(s,a) η αναμενόμενη προεξοφλημένη ανταμοιβή μιας ενέργειας a σε μια κατάσταση s. Θεωρώντας ότι η βέλτιστη ενέργεια έχει πραγματοποιηθεί αρχικά προκύπτει ότι $V^*(s)=\max_a Q^*(s,a)$. Συνεπώς, η Q*(s,a) γράφεται υπό την αναδρομική μορφή

$$Q^*(s,a) = R(s,a) + \gamma \sum_{s \in S}^{n} T(s,a,s') \max_a Q(s',a') \quad (2.12)$$

Κάθε πράκτορας ακολουθεί άπληστη πολιτική με βάση τις εκτιμήσεις του. Αυτό έχει ως αποτέλεσμα ότι, αν τη χρονική στιγμή t ο πράκτορας βρίσκεται στην κατάσταση s, η επόμενη ενέργεια που θα επιλέξει θα είναι η

$$a' := \operatorname{argmax}_{a \in A} Q^*(s,a) \quad (2.13)$$

Πρέπει, επίσης, να σημειωθεί ότι, εφόσον ισχύει $V^*(s)=\max_a Q^*(s,a)$, η βέλτιστη πολιτική προκύπτει από τη σχέση

$$\pi^*(s) = \operatorname{argmax}_a Q^*(s,a) \quad (2.14)$$

2.4.8 Ο αλγόριθμος μάθησης Q

Ο αλγόριθμος μάθησης Q βασίζεται στον ορισμό της συνάρτησης Q. Ένας πράκτορας υπολογίζει επαναληπτικά τις τιμές της συνάρτησης Q. Σε κάθε επανάληψη του αλγορίθμου ο πράκτορας παρατηρεί την παρούσα κατάσταση s, επιλέγει μια δράση a και έπειτα παρατηρεί την ανταμοιβή r=r(s,a) και τη νέα κατάσταση s΄=δ(s,a). Στη συνέχεια, ενημερώνει την αναμενόμενη τιμή της συνάρτησης Q σύμφωνα με τον ακόλουθο κανόνα εκπαίδευσης

$$Q(s_t,a_t):=Q(s_t,a_t)\ (\ 1-\ a_t\ (s_t,a_t))\ +\ a_t\ (s_t,a_t)[r_{t+1}\ +\ \gamma max_{a'}\ Q(s_{t+1},a)] \qquad (2.15)$$

Τα βήματα του αλγορίθμου είναι τα ακόλουθα:

Για όλες τις καταστάσεις s στο S και για όλες τις δράσεις α στο A

Απόδωσε αρχική τιμή στη συνάρτηση Q(s,a)

***ΕΠΑΝΑΛΑΒΕ** (για κάθε προσπάθεια)*

Αρχικοποίησε την παρούσα κατάσταση s

***ΕΠΑΝΑΛΑΒΕ** (για κάθε βήμα της προσπάθειας)*

Παρατήρησε την παρούσα κατάσταση s

Διάλεξε μια δράση a ακολουθώντας μια πολιτική π

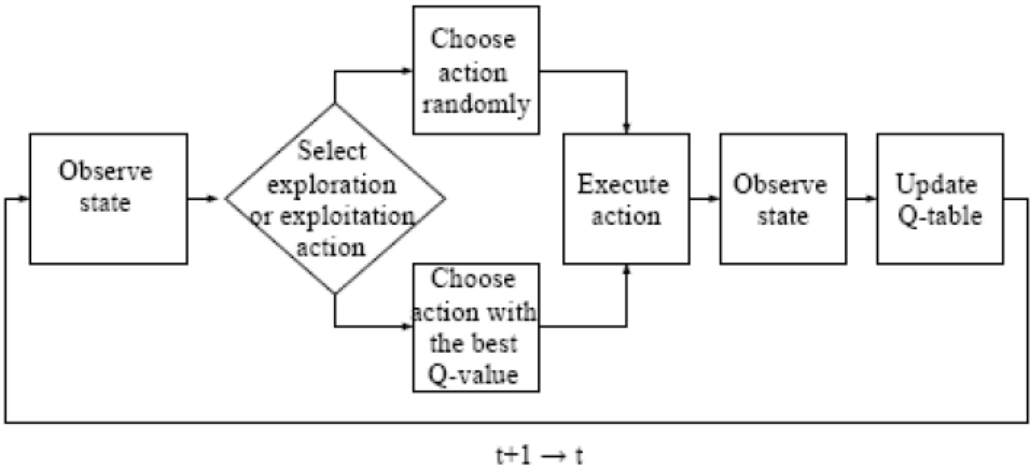
Εκτέλεσε τη δράση a

Λάβε μια άμεση ανταμοιβή r

Παρατήρησε τη νέα κατάσταση s΄

Ανανέωσε την Q(s,a) σύμφωνα με τη σχέση (2.15)

***ΜΕΧΡΙ** η s να είναι τελική κατάσταση*



Σχήμα: 2.6 Διάγραμμα λειτουργίας εκμάθησης Q

Οι Watkins και Danyan απέδειξαν ότι οι αναμενόμενες τιμές Q ενός πράκτορα θα συγκλίνουν στις πραγματικές τιμές με πιθανότητα 1 υπό με τις εξής προϋποθέσεις:

- το περιβάλλον είναι μια σταθερή αλυσίδα Markov με δεδομένες ανταμοιβές r(s,a)
- οι αναμενόμενες τιμές της συνάρτησης Q αποθηκεύονται σε ένα πίνακα αναζήτησης και αρχικοποιούνται σε πεπερασμένες αυθαίρετες τιμές
- κάθε δράση εκτελείται σε μια κατάσταση άπειρες φορές
- $\gamma \in [0,1)$, $a \in [0,1)$ και το a μειώνεται σταδιακά ως το 0 καθώς περνάει ο χρόνος

Η μάθηση Q είναι ανεπηρέαστη από την εξερεύνηση: Οι τιμές της Q θα συγκλίνουν στις βέλτιστες τιμές ανεξάρτητα από το πώς ο πράκτορας συμπεριφέρεται όταν συλλέγονται τα δεδομένα (εφόσον βέβαια όλα τα ζεύγη καταστάσεων-δράσεων δοκιμαστούν αρκετά συχνά). Αυτό σημαίνει ότι, αν και η σχέση αξιοποίησης εκμετάλλευσης-εξερεύνησης πρέπει να διευθετηθεί κατά τη μάθηση Q, οι λεπτομέρειες της πολιτικής εξερεύνησης δεν θα επηρεάσουν τη σύγκλιση του αλγορίθμου μάθησης. Για αυτούς τους λόγους η μάθηση Q είναι η πλέον δημοφιλής και φαίνεται ως ο αποτελεσματικότερος αλγόριθμος για μάθηση με καθυστερημένη ενίσχυση για συστήματα για τα οποία δεν υπάρχει διαθέσιμο μοντέλο το οποίο περιγράφει την εύρεση της βέλτιστης πολιτικής. Το μειονέκτημά της είναι ότι δεν διευθετεί θέματα που αφορούν τη γενίκευση σε μεγάλους χώρους καταστάσεων ή και δράσεων ενώ μπορεί να συγκλίνει αργά προς μια βέλτιστη πολιτική.

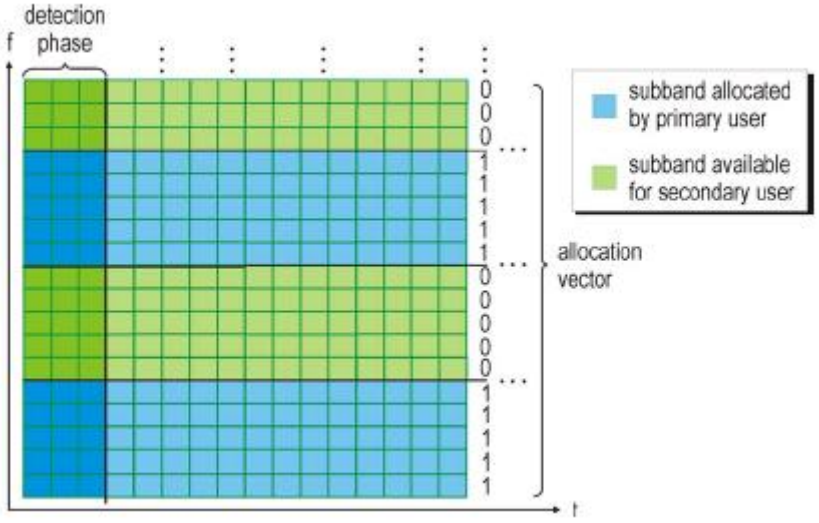
2.5 Ανίχνευση πηγών φάσματος με εφαρμογή αλγορίθμου ενισχυτικής μάθησης Q

2.5.1 Περιγραφή του συστήματος

Γίνεται η υπόθεση σεναρίου στο οποίο ένα σύστημα πρωτευόντων χρηστών (primary users system) και ένα σύστημα δευτερευόντων χρηστών (secondary users system) λειτουργούν στην ίδια ζώνη συχνοτήτων (frequency band). Για το σύστημα πρωτευόντων χρηστών γίνεται η παραδοχή ότι έχει προτεραιότητα και δεν πρέπει να επηρεάζεται από τη λειτουργία του συστήματος δευτερευόντων χρηστών. Έτσι, όλες οι απαραίτητες διαδικασίες επεξεργασίας σήματος (signal processing) και συντονισμού (coordination) από τη συνύπαρξη των δύο συστημάτων πρέπει να εκτελούνται στο σύστημα δευτερευόντων χρηστών. Οι κύριες επιδιώξεις (main goals) είναι να ελαχιστοποιηθεί η αμοιβαία παρεμβολή μεταξύ των συστημάτων πρωτευόντων και δευτερευόντων χρηστών και να γίνει η αποδοτικότερη δυνατή χρησιμοποίηση των διαθέσιμων πόρων φάσματος.

Στο σύστημα πρωτευόντων χρηστών απαιτείται ο συνδυασμός διαίρεσης χρόνου και πολλαπλής πρόσβασης συχνότητας (TDMA/FDMA). Γίνεται η υπόθεση ότι οι πρωτεύοντες χρήστες μπορούν να λειτουργήσουν σε κάθε ζώνη συχνοτήτων με διαφορετικά χαρακτηριστικά λειτουργίας. Κάθε μία από τις ζώνες συχνοτήτων αυτές διαθέτει διαφορετικό πλήθος από φασματικά κενά και, επομένως, διαθέσιμο φάσμα το οποίο θα μπορέσουν να εκμεταλλευτούν οι δευτερεύοντες χρήστες.

Για το σύστημα δευτερευόντων χρηστών χρησιμοποιείται ορθογωνική πολυπλεξία διαίρεσης συχνότητας (Orthogonal Frequency Division Multiplexing, OFDM) η οποία αποτελεί τεχνική διαμόρφωσης για ευέλικτα συστήματα δυναμικής εκχώρησης φάσματος, με δυνατότητες επίτευξης υψηλών ρυθμών μετάδοσης μέσω της συλλογικής χρήσης μεγάλου αριθμού υποφερουσών. Ένα βασικό πλεονέκτημα του συστήματος που μελετάται είναι ότι κάθε υποφέρουσα συχνότητα έχει τη δυνατότητα να είναι ενεργή (on) ή ανενεργή (off) με βάση την πρόσφατη κατανομή του συστήματος πρωτευόντων χρηστών. Στο Σχ.2.7 φαίνεται ότι χρησιμοποιούνται πέντε υποφέρουσες συχνότητες για κάθε κανάλι.



Σχήμα 2.7: Κατανομή πρωτευόντων χρηστών στο πλάνο χρόνου-συχνότητας

Το σύστημα των δευτερευόντων χρηστών πρέπει να εκτελέσει περιοδικά μετρήσεις για να προσδιορίσει την κατανομή του συστήματος πρωτευόντων χρηστών στο φάσμα. Αυτή η μέτρηση κατανομής μπορεί να πραγματοποιηθεί αποδοτικά χωρίς πρόσθετα κόστη υλοποίησης (hardware) χρησιμοποιώντας τα ήδη υπάρχοντα συστήματα ταχέος μετασχηματισμού Fourier (FFT). Το αποτέλεσμα του προσδιορισμού της κατανομής των πρωτευόντων χρηστών από τους δευτερεύοντες χρήστες είναι το διάνυσμα κατανομής (allocation vector AV) που υποδεικνύει για κάθε φέρουσα αν κάποιος πρωτεύων χρήστης ήταν ενεργός ‘1’ ή ανενεργός ‘0’.

2.5.2 Σύστημα ανίχνευσης φάσματος

Σε ένα σενάριο όπου υπάρχουν πολλές φασματικές ζώνες (multi-band) όπως αυτό που μελετάται πρέπει να γίνει διαχωρισμός σε δύο τύπους ανίχνευσης φάσματος. Από τη μια πλευρά, σε ανίχνευση που πρέπει να πραγματοποιηθεί περιοδικά στην πλέον πρόσφατα χρησιμοποιούμενη για μετάδοση ζώνη συχνοτήτων ώστε να αποφευχθεί η σύγκρουση των δευτερευόντων χρηστών με τους πρωτεύοντες χρήστες. Από την άλλη πλευρά, το σύστημα δευτερευόντων χρηστών πρέπει να παρατηρεί τις άλλες ζώνες συχνοτήτων για την ύπαρξη πιθανών φασματικών ευκαιριών. Σύμφωνα με τα δύο αυτά βασικά εργαλεία η διαδικασία ανίχνευσης χωρίζεται σε δύο φάσεις ως ακολούθως.

1) *Ανίχνευση της κατανομής φάσματος των πρωτευόντων χρηστών*: Το τμήμα αυτό είναι υπεύθυνο για την ακριβή και λεπτομερή ανίχνευση των διαθέσιμων πηγών φάσματος στην τρέχουσα ζώνη συχνοτήτων. Τα βασικά χαρακτηριστικά της φάσης αυτής είναι:

- η διαδικασία ανίχνευσης εκτελείται αρκετά συχνά για να αποφευχθούν συγκρούσεις με το σύστημα πρωτευόντων χρηστών της ζώνης συχνοτήτων που εξετάζεται
- η διαδικασία ανίχνευσης εκτελείται για όλα τα υποκανάλια στον ίδιο χρόνο χρησιμοποιώντας FFT
- τα αποτελέσματα της ανίχνευσης χρησιμοποιούνται για τον ακριβή προσδιορισμό των διαθέσιμων φασματικών κενών στη ζώνη συχνοτήτων που πραγματοποιείται η ανίχνευση

2) Ανίχνευση των πόρων φάσματος: Ο σκοπός αυτού του σταδίου της διαδικασίας ανίχνευσης φάσματος είναι να υπάρξει μια ικανοποιητική προσέγγιση στο μέσο όρο των διαθέσιμων πόρων σε όλες τις φασματικές ζώνες (ζώνες συχνοτήτων) του συστήματος. Αυτή η πληροφορία χρησιμοποιείται ώστε να γίνει περισσότερο καθολική η στρατηγική αποφάσεων για τις δραστηριότητες των δευτερευόντων χρηστών στο άμεσο μέλλον. Τα βασικά χαρακτηριστικά του σταδίου αυτού είναι:

- Η διαδικασία ανίχνευσης πραγματοποιείται μία φορά ώστε να γίνει εκτίμηση για την τρέχουσα κατάσταση κατανομής πόρων φάσματος στις άλλες ζώνες συχνοτήτων όπου λειτουργούν πρωτεύοντες χρήστες. Σε αντίθεση με τα χαμηλά επίπεδα ανίχνευσης ο πραγματικός σκοπός (goal) δεν είναι να αποφευχθούν συγκρούσεις με τους πρωτεύοντες χρήστες αλλά να υπάρξει επισκόπηση για την κατάσταση των φασματικών ευκαιριών σε όλες τις ζώνες συχνοτήτων του διαθέσιμου φάσματος. Κατα τη διάρκεια του σταδίου αυτού οι δευτερεύοντες χρήστες δεν μπορούν να πραγματοποιήσουν μετάδοση δεδομένων. Έτσι, πρέπει να επιτευχθεί το καλύτερο δυνατό trade-off μεταξύ της σπατάλης χρόνου για μετάδοση δεδομένων και του χρόνου που δαπανάται για την εκτίμηση της φασματικής κατάστασης στις άλλες ζώνες συχνοτήτων. Αυτό το trade-off θέτει ουσιαστικά το πρόβλημα της ενισχυτικής μάθησης στο εξεταζόμενο σύστημα. Ένας πράκτορας (το σύστημα δευτερευόντων χρηστών) πρέπει να μαθαίνει ποιές αποφάσεις θα έχουν ως αποτέλεσμα τα καλύτερα οφέλη για το σύστημα
- Η διαδικασία ανίχνευσης εκτελείται χρησιμοποιώντας FFT αλλά υπό τον περιορισμό ότι μόνο μια ζώνη συχνοτήτων μπορεί να ανιχνεύεται κάθε φορά. Όταν πραγματοποιείται ανίχνευση σε διαφορετική ζώνη συχνοτήτων για το επόμενο διάστημα ανανέωσης της κατάστασης των δευτερευόντων χρηστών, η ζώνη αυτή δεν μπορεί να χρησιμοποιηθεί για μετάδοση δεδομένων
- Τα αποτελέσματα της ανίχνευσης του σταδίου αυτού χρησιμοποιούνται για μια ενδιάμεση εκτίμηση και πρόβλεψη για την κατανομή των φασματικών ευκαιριών για τους δευτερεύοντες χρήστες ώστε να ληφθούν αποφάσεις στρατηγικής κατά την εξέλιξη του σεναρίου

2.5.2.1 Διατύπωση προβλήματος ενισχυτικής μάθησης

Στο σύστημα που μελετάται πραγματοποιούνται δύο προσεγγίσεις του τρόπου ενισχυτικής μάθησης των δευτερευόντων χρηστών.

1) Απλή διατύπωση του προβλήματος: Σε αυτή την προσέγγιση γίνεται η υπόθεση ότι δεν υπάρχουν πρόσθετα κόστη για τους δευτερεύοντες χρήστες καθώς πραγματοποιούν μετάβαση από κάποια ζώνη συχνοτήτων σε κάποια άλλη. Κατα την εκτέλεση του απλού προβλήματος ενισχυτικής μάθησης υπάρχει ένα μόνο στάδιο κατά το οποίο το σύστημα δευτερευόντων χρηστών εκτελεί αρκετές δράσεις $\alpha \in A$ με διαφορετικές ανταμοιβές $Q^*(\alpha)$, οι οποίες καλούνται τιμές της δράσης α. Στο γενικό πλαίσιο της ανίχνευσης των πηγών φάματος κάθε δράση του συστήματος δευτερευόντων χρηστών αντιστοιχεί σε ένα κύκλο επιλογής μιας ζώνης συχνοτήτων, εκτελώντας μια φάση ανίχνευσης και μια φάση μετάδοσης δεδομένων. Ο αριθμός των bits που μεταδίδει ο δευτερεύων χρήστης σε αυτή την περίπτωση είναι η άμεση ανταμοιβή (reward) του προβλήματος ενισχυτικής μάθησης. Πρέπει να τονιστεί ότι εφόσον το σύστημα δευτερευόντων χρηστών δεν γνωρίζει άμεσα, αλλά μαθαίνει προοδευτικά, πόσες και ποιές πηγές άλλων ζωνών συχνοτήτων είναι διαθέσιμες, πρέπει να βρεί ένα καλό trade-off μεταξύ εκμετάλλευσης και εξερεύνησης. Για να γίνουν περισσότερο σαφείς οι έννοιες: το σύστημα δευτερευόντων χρηστών έχει να επιλέξει ανάμεσα στην παραμονή του στη ζώνη συχνοτήτων που βρίσκεται τη δεδομένη χρονική στιγμή όπου γνωρίζει πόσα δεδομένα μπορεί να στείλει ή να μεταφερθεί σε μία άλλη ζώνη συχνοτήτων με την ευκαιρία να βρεί περισσότερες διαθέσιμες πηγές φάσματος (φασματικές ευκαιρίες) αλλά και με το ρίσκο να υπάρχουν λιγότερες τέτοιες ευκαιρίες στη ζώνη που θα μεταβεί. Το ζητούμενο για το σύστημα δευτερευόντων χρηστών είναι να μεταβεί σε ζώνη συχνοτήτων η οποία του προσφέρει, εκείνη τη χρονική στιγμή, τη δυνατότητα να μεταδόσει τα περισσότερα δεδομένα. Οι εκτιμώμενες τιμές Q σε αυτή την περίπτωση ανανεώνονται σύμφωνα με την ακόλουθη σχέση:

$$Q_{k+1} = Q_k + a[r_{k+1} - Q_k] \qquad (2.16)$$

όπου α είναι παράμετρος μεγέθους βήματος και r η ληφθείσα ανταμοιβή από την ενέργεια του συστήματος δευτερευόντων χρηστών. Η πιθανότητα να επιλεγεί η επόμενη δράση με βάση τις τιμές Q επιτυγχάνοντας ένα καλό trade-off μεταξύ εξερεύνησης και εκμετάλλευσης, προσδιορίζεται με χρήση της μεθόδου soft-max η οποία δίνει την πιθανότητα

$$P_r(a) = \frac{e^{Q_t(a)/\tau}}{\sum_{b=1}^{n} e^{Q_t(b)/\tau}}, \qquad (2.17)$$

να επιλεγεί η δράση α. Προκύπτει ότι για τ→0 οι δράσεις με την υψηλότερη τιμή Q προτιμώνται, ενώ για υψηλές τιμές του τ όλες οι δράσεις επιλέγονται σχεδόν με την ίδια πιθανότητα. Λόγω της απλότητάς του το σενάριο που περιγράφηκε εξυπηρετεί ως απλή αναφορά για το ακόλουθο σενάριο της προσομοίωσης.

2) *Διατύπωση του προβλήματος ενισχυτικής μάθησης ως διαδικασία απόφασης Markov*: Σε ένα ρεαλιστικό σενάριο η μετάβαση των δευτερευόντων χρηστών σε άλλες ζώνες συχνοτήτων δεν επιτυγχάνεται με μηδενικά κόστη, αφού ολόκληρο το σύστημα δευτερευόντων χρηστών πρέπει να ενημερωθεί για τις προγραμματισμένες αλλαγές. Αυτή η διαδικασία απαιτεί πρόσθετη σηματοδοσία και ανταμοιβές και εξαιτίας αυτού μειώνεται ο αριθμός των πηγών που είναι διαθέσιμες για μετάδοση δεδομένων. Σε αντίθεση με τη διατύπωση του απλού προβλήματος με την ύπαρξη ενός μόνο σταδίου και πολλών δράσεων, σε αυτό το σενάριο χρησιμοποιείται ένα μοντέλο με πολλά στάδια. Υποθέτοντας ότι η πλέον πρόσφατη κατάσταση όπου βρέθηκε το σύστημα περιέχει όλες τις απαραίτητες πληροφορίες για το ιστορικό των δευτερευόντων χρηστών καθώς και για τις επόμενες δράσεις τους· έτσι γίνεται επιλογή με βάση αυτή την κατάσταση. Αυτός ο τύπος προβλήματος μπορεί να διατυπωθεί ως διαδικασία απόφασης Markov όπως αυτή περιγράφηκε στα προηγούμενα.

Έπειτα από την παρατήρηση της τελευταίας κατάστασης όπου βρέθηκε το σύστημα, ο πράκτορας (σύστημα δευτερευόντων χρηστών) πρέπει να επιλέξει μια δράση για το επόμενο στάδιο. Αυτό γίνεται σύμφωνα με την πολιτική π:S×A→[0,1] όπου π(α,s) είναι η πιθανότητα η δράση α να εκτελείται όταν ο πράκτορας βρίσκεται στην κατάσταση s. Η βέλτιστη πολιτική μεγιστοποιεί τις σωρευτικές αναμενόμενες ανταμοιβές οι οποίες συνήθως προεξοφλούνται από τον παράγοντα προεξόφλησης γ∈[0,1] στην περίπτωση άπειρου ορίζοντα. Έτσι ο απώτερος στόχος (goal) είναι να προσδιορισθεί η βέλτιστη πολιτική που μεγιστοποιεί την αναμενόμενη ανταμοιβή.

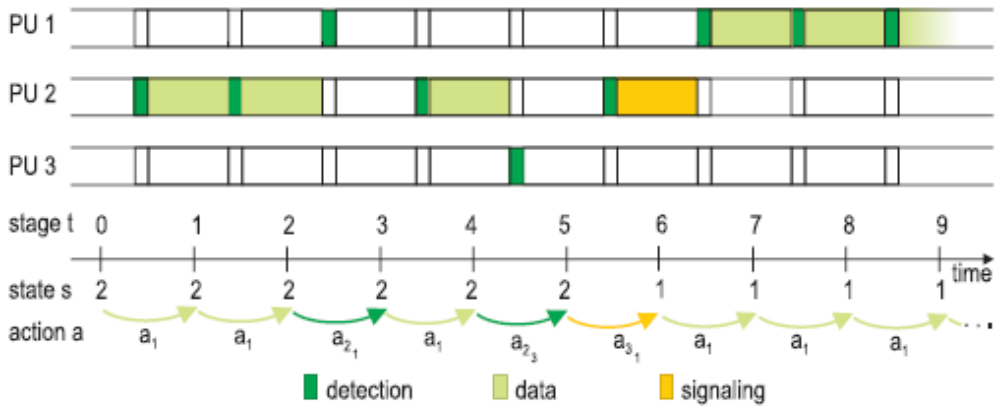
$$R = E[\sum_{t=0}^{\infty} \gamma^t r_t(a_t, s_t)] \tag{2.18}$$

Πριν αναφερθεί η προτεινόμενη στρατηγική επίλυσης του προβλήματος διατυπώνεται σαφώς το πρόβλημα της ανίχνευσης πηγών φάσματος ως MDP. Ο διαθέσιμος αριθμός των ζωνών φάσματος N_f στο σενάριο μπορεί να ερμηνευθεί ως ο μέγιστος αριθμός από καταστάσεις της MDP από τότε που το σύστημα δευτερευόντων χρηστών μπορεί να ανιχνεύει μόνο μία ζώνη συχνοτήτων κάθε φορά. Επομένως, ορίζεται ένα σύνολο από καταστάσεις όπου S=(1,2,3,…, N_f) και η πλέον πρόσφατη κατάσταση αντιστοιχεί στην πλέον πρόσφατη ζώνη συχνοτήτων όπου βρέθηκε το σύστημα των δευτερευόντων χρηστών και μπορεί να μεταδώσει δεδομένα ή να επιλέξει τη φάση ανίχνευσης σε μία άλλη ζώνη συχνοτήτων που συνδέεται με διαφορετικά κόστη μετάβασης. Το σύνολο των πιθανών δράσεων

στην κατάσταση s είναι $A_s=(\alpha_1,\alpha_2,\alpha_3\,\bar{s})$ όπου $\bar{s} \in S$ και οι δράσεις περιγράφονται ως ακολούθως

- α_1: εκτελεί τη φάση ανίχνευσης στη ζώνη συχνοτήτων s και μεταδίδει δεδομένα
- α_2: εκτελεί τη φάση ανίχνευσης στη ζώνη συχνοτήτων $\bar{s}$ (έξω από τη ζώνη που βρίσκεται αυτή τη χρονική στιγμή το σύστημα δευτερευόντων χρηστών)
- α_3: αλλάζει το σύστημα δευτερευόντων χρηστών ζώνη συχνοτήτων s

Είναι φανερό ότι αλλαγή κατάστασης πραγματοποιείται όταν εκτελείται η δράση α_3. Το Σχ.2.8 που ακολουθεί δείχνει ένα παράδειγμα για μια ακολουθία καταστάσεων και δράσεων του συστήματος.



Σχήμα 2.8: Παράδειγμα καταστάσεων και δράσεων ενός συστήματος δευτερευόντων χρηστών σε ένα περιβάλλον τριών ζωνών συχνοτήτων

Για την άμεση συνάρτηση ανταμοιβής r χρησιμοποιείται ο ακόλουθος ορισμός:

$$r(a,s)=\begin{Bmatrix} u_1(s), a=a_1 \\ u_2(s), a=a_2 \\ u_3(s), a=a_3 \end{Bmatrix} \tag{2.19}$$

όπου:

$u_1(s)$ είναι ο αριθμός των πηγών του φάσματος όπου έχει πραγματοποιηθεί μετάδοση στο πλέον πρόσφατο στάδιο της προσομοίωσης ενώ το σύστημα δευτερευόντων χρηστών παρέμενε στην πλέον πρόσφατη ζώνη συχνοτήτων. Σύμφωνα με το μοντέλο του συστήματος πρωτευόντων χρηστών η κατανομή σε κάθε ζώνη συχνοτήτων είναι διαδικασία που υποδεικνύει τον αριθμό των διαθέσιμων υποφερουσών όπου $u_1(s)$ είναι ένα δείγμα αυτών.

Η $u_2(s)$ είναι μια διαφορετική ζώνη συχνοτήτων που είναι ανεξάρτητη από την τελευταία κατάσταση όπου έχει βρεθεί το σύστημα των δευτερευόντων χρηστών. $u_2=0$ είναι συνήθως μια καλή επιλογή. Τη φάση μετάδοσης δεδομένων ακολουθεί η φάση ανίχνευσης η οποία εκτελείται σε διαφορετικές ζώνες συχνοτήτων. Κατά τη διάρκεια αυτής της φάσης δεν μεταδίδονται δεδομένα επειδή το σύστημα δευτερευόντων χρηστών δεν έχει διαθέσιμη πληροφορία για την κατανομή στην πλέον πρόσφατη ζώνη συχνοτήτων στην οποία βρέθηκε ως αποτέλεσμα της μη άμεσης ανταμοιβής. Από την άλλη πλευρά, δεν υπάρχουν σε αυτή την περίπτωση πρόσθετα κόστη σηματοδοσίας.

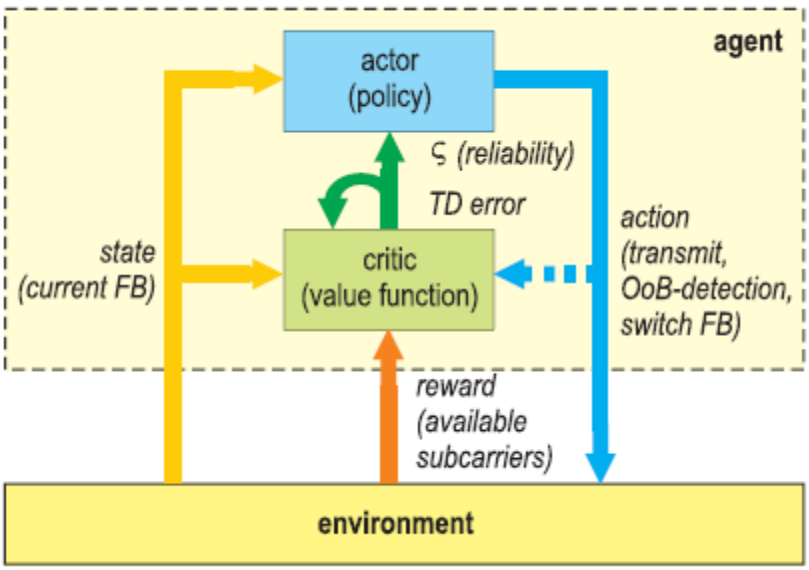
Η $u_3(s)$ περιγράφει τα κόστη της μετάβασης της μετάδοσης από μια ζώνη συχνοτήτων σε άλλη, περιλαμβάνοντας την απαραίτητη προσπάθεια σηματοδοσίας. Η $u_3(s)$ ανήκει στο ειδικό σχέδιο του πρωτοκόλλου εφαρμογής. Στην περίπτωση όπου όλες οι διαδικασίες σηματοδοσίας μπορούν να γίνουν σε ένα διάστημα ανανέωσης, $u_3=0$. Όμως εφόσον η διαδικασία αυτή απαιτήσει περισσότερες από μία προσπάθειες ανανέωσης αποδίδεται μια αρνητική τιμή που αντιστοιχεί σε πρόσθετη απώλεια πηγών. Να σημειωθεί ότι η μετάβαση σε άλλη ζώνη συχνοτήτων είναι απαραίτητη όταν οι πλέον πρόσφατες πηγές φάσματος που χρησιμοποιήθηκαν είχαν χαμηλή απόδοση και, συγκεκριμένα, όταν μόνο ένας μικρός αριθμός υποφερουσών είναι διαθέσιμος. Αυτό σημαίνει ότι η μετάδοση των δεδομένων σηματοδοσίας πραγματοποιείται έπειτα από αρκετές προσπάθειες ανανέωσης.

2.5.2.2 Στρατηγική επίλυσης

Έχοντας ως βάση τη διατύπωση του προβλήματος που έγινε ανωτέρω, εξετάζεται μια στρατηγική επίλυσης χρησιμοποιώντας μεθόδους ενισχυτικής μάθησης. Για την εύρεση της βέλτιστης πολιτικής πολλοί αλγόριθμοι ενίσχυσης χρησιμοποιούν το σχήμα κατάστασης-ανταμοιβής (state-value) συναρτήσεων V^{Π} :$S \rightarrow R$, το οποίο καθορίζει για κάθε κατάσταση του συστήματος μια πραγματική ανταμοιβή που περιγράφει πόσο καλή είναι η επιλογή αυτής της κατάστασης και, κατά συνέπεια, η πολιτική που ακολουθείται. Για την εύρεση της βέλτιστης πολιτικής π* αρκετοί αλγόριμοι ενισχυτικής μάθησης ακολουθούν τη μέθοδο της επανάληψης πολιτικών, η οποία, όπως έχει ήδη αναφερθεί, εκτελείται μέχρι να υπάρξει σύγκλιση στη βέλτιστη κατάσταση για το σύστημα. Τα βασικά σημεία του προβλήματος ανίχνευσης πηγών στο διαθέσιμο φάσμα είναι τα εξής :

Αν και οι πιθανότητες μετάβασης κατάστασης είναι απλές και γνωστές για το πρόβλημα, δεν είναι γνωστές και ακριβείς οι ανταμοιβές μιας δράσης πριν την εκτέλεσή της. Αυτό σημαίνει ότι δεν υπάρχει πλήρης γνώση για το μοντέλο του συστήματος. Επιπλέον, στην περίπτωση μιας μη στατικής διαδικασίας κατανομής, ο χρόνος έχει αρνητική επίδραση στην ακρίβεια της εκτίμησης της κατάστασης-ανταμοιβής. Η ακρίβεια γίνεται όλο και μικρότερη όταν η φάση ανίχνευσης που

εκτελείται σε μια άλλη ζώνη συχνοτήτων από την τρέχουσα διαρκεί όλο και περισσότερο. Οι αλγόριθμοι μάθησης χρονικής διαφοράς TD αποτελούν μια προσέγγιση που προσιδιάζει στον τύπο προβλήματος που μελετάται. Ειδικά οι μέθοδοι προσαρμοστικής ευριστικής κριτικής (actor-critic) έχουν το μεγαλύτερο ενδιαφέρον επειδή χρησιμοποιούν μια δομή μνήμης η οποία περιγράφει σαφώς την ανεξάρτητη πολιτική συνάρτησης ανταμοιβής. Η αρχιτεκτονική ευριστικής κριτικής που έχει αναλυθεί εκτενώς στα προηγούμενα παρουσιάζεται στο Σχ.2.9 που ακολουθεί.



Σχήμα 2.9: Αλληλεπίδραση του δευτερεύοντος χρήστη με το περιβάλλον και η εφαρμοσμένη δομή προσαρμοστικής ευριστικής κριτικής

Ο κριτής (critic) χρησιμοποιεί τη ληφθείσα ανταμοιβή για να ανανεώσει τις ανταμοιβές της κατάστασης που βρίσκεται και να παράξει μερικές πληροφορίες, το σφάλμα TD (TD error) το οποίο στη συνέχεια χρησιμοποιείται από τον actor για να ανανεώσει την πολιτική. Στην περίπτωση που εξετάζεται ο κριτής υπολογίζει πρόσθετες πληροφορίες για την αξιοπιστία των καταστάσεων-ανταμοιβών.

1) Κριτής: Ο κριτής ανανεώνει την κατάσταση-ανταμοιβή σύμφωνα με τον ακόλουθο κανόνα ανανέωσης

$$V(s_t) \leftarrow V(s_t) + \beta[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \tag{2.20}$$

όπου β είναι μια θετική παράμετρος μεγέθους-βήματος και γ ο παράγοντας προεξόφλησης. Το δεύτερο μέρος της 2.20 περιγράφει το TD σφάλμα:

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \tag{2.21}$$

Η αξιοπιστία των καταστάσεων-ανταμοιβών V(s) εξηγείται από την αντιστοίχιση της αξιοπιστίας ανταμοιβής ς(s) με $0 \le s \le 1$. Μικρά s υποδηλώνουν μικρή αξιοπιστία των καταστάσεων-ανταμοιβών. Για να ανανεωθεί η αξιοπιστία των ανταμοιβών ο κριτής χρειάζεται να γνωρίζει ποιά δράση εκτελέστηκε. Σε κάθε στάδιο αυξάνεται μόνο η αξιοπιστία ανταμοιβής της ζώνης συχνοτήτων στην οποία έγινε ανίχνευση (αυτό συμβαίνει όταν εκτελείται η α_1 ή η $\alpha_2 \bar{s}$). Σε όλες τις άλλες περιπτώσεις η αξιοπιστία ανταμοιβών μειώνεται. Να σημειωθεί, επίσης, ότι η ανταμοιβή που παραλαμβάνεται δεν είναι σχετική με την αξιοπιστία. Ένας πιθανός κανόνας ανανέωσης για κάθε s είναι

$$\varsigma_t \leftarrow \varsigma_t + k[d - \varsigma_t] \tag{2.22}$$

όπου d ένας δυαδικός δείκτης που περιγράφει αν μια φάση ανίχνευσης επιλέχθηκε στη ζώνη συχνοτήτων που αντιστοιχεί στην τρέχουσα εκτέλεση (d=1) ή όχι (d=0). Επίσης k∈(0,1) είναι μία άλλη θετική παράμετρος μεγέθους βήματος. Να σημειωθεί ότι κάθε ς ανανεώνεται σε κάθε στάδιο. Στην περίπτωση της δράσης $\alpha_2 \bar{s}$ ανανεώνεται επιπλέον το V(s) σύμφωνα με τον ακόλουθο κανόνα ανανέωσης

$$V(\bar{s}_t) \leftarrow V(\bar{s}) + a[\bar{r}_{t+1} - V(\bar{s}_t)] \tag{2.23}$$

όπου $\bar{r}_{t+1}$ είναι το αποτέλεσμα της ανίχνευσης στο πεδίο της συχνότητας και α παράμετρος μεγέθους-βήματος. Η $\bar{r}_{t+1}$ δεν αποτελεί μία άμεση ανταμοιβή από τη στιγμή που δεν μεταδίδονται δεδομένα αλλά χρησιμοποιείται για να εκτιμηθεί το $V(\bar{s}_t)$.

2) Actor : Βασίζεται στο TD σφάλμα δ και οι τιμές αξιοπιστίας ς(s) του actor πρέπει να ανανεώνουν την πρόσφατη πολιτική. Αυτό γίνεται υπολογίζοντας τις τιμές προτίμησης P για κάθε δράση με βάση την πολιτική $\pi_t(s,\alpha)$ με εφαρμογή της μεθόδου soft-max για τον υπολογισμό της, δηλαδή

$$\pi_t(s,a) = P_r(a_t = a \mid s_t = s) = \frac{e^{p(s,a)}}{\sum_{b=1}^{n} e^{p(s,b)}}, \tag{2.24}$$

Οι προτιμήσεις υπολογίζονται με βάση τον τύπο της δράσης. Η δράση α_1 (μετάδοση δεδομένων) ανανεώνεται χρησιμοποιώντας ένα κοινό κανόνα ανανέωσης

$$p(s,a_1) \leftarrow p(s,a_1) + \beta_1 \delta_t \tag{2.25}$$

Η συγκέντρωση των δράσεων $\alpha_2 \bar{s}$ (εκτέλεση ανίχνευσης σε μια ζώνη συχνοτήτων $\bar{s}$ έχει το χαρακτήρα της εξερεύνησης για το σύστημα δευτερευόντων χρηστών. Επομένως, είναι προτιμότερο να εκτελούνται ανιχνεύσεις σε ζώνες συχνοτήτων όπου η αξιοπιστία των καταστάσεων-ανταμοιβών είναι χαμηλές. Αυτή η επιδίωξη

γίνεται εφικτή από την ακόλουθη αντιστοίχιση καταστάσεων-ανταμοιβών στις προτιμήσεις

$$p(s, a_{2_s}) = (1-\varsigma) \cdot V(s) \tag{2.26}$$

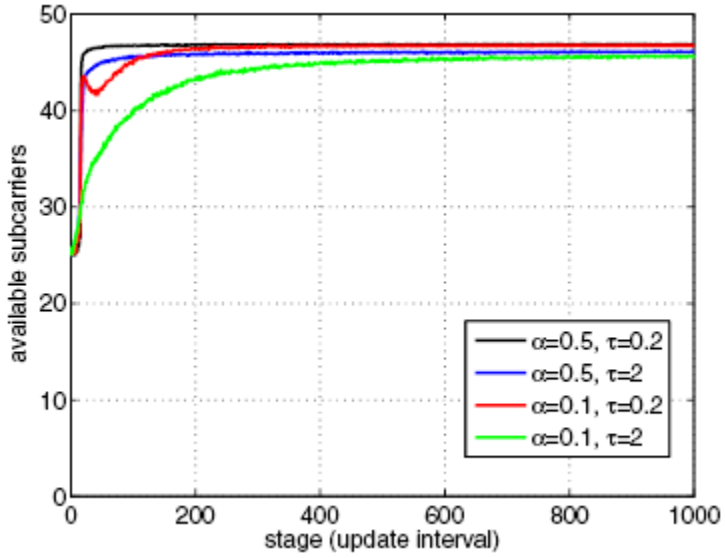
Η δράση α_3 (μετάβαση του συστήματος δευτερευόντων χρηστών σε άλλη ζώνη συχνοτήτων) είναι επιθυμητό να εκτελείται όταν υπάρχει κάποια ζώνη συχνοτήτων με πολλές διαθέσιμες πηγές φάσματος και η πληροφορία περί αυτού να είναι αξιόπιστη. Επίσης, πιθανές αναξιόπιστες πηροφορίες είναι προτιμότερες από αξιόπιστες πληροφορίες για χαμηλό αριθμό πηγών. Η αντιστοίχιση δίδεται από την ακόλουθη σχέση

$$p(s, a_{3_s}) = \varsigma(V(S) - \frac{N_{fb}}{2}) + \frac{N_{fb}}{2} \tag{2.27}$$

2.5.3 Αποτελέσματα προσομοίωσης

Για όλες τις προσομοιώσεις που εκτελέστηκαν έγινε η υπόθεση ότι N_{fb} =15 διαφορετικές ζώνες συχνοτήτων, κάθε μία από τις οποίες διαθέτει 50 κανάλια. Για χάρη απλότητας το σύστημα δευτερευόντων χρηστών χρησιμοποιεί μια υποφέρουσα από κάθε κανάλι πρωτευόντων χρηστών από τις 50 διαθέσιμες υποφέρουσες. Γνωρίζοντας τον αριθμό των OFDM συμβόλων, ο διαθέσιμος αριθμός υποφερουσών μπορεί να μεταφραστεί με ακρίβεια σε ρυθμό μετάδοσης. Κάθε κανάλι του συστήματος πρωτευόντων χρηστών είναι διαθέσιμο με πιθανότητα $p_{avail}(n)$ (όπου n αντιστοιχεί στη ζώνη συχνοτήτων), ως αποτέλεσμα διαφορετικών διωνυμικών κατανομών με τις παραμέτρους N_{fb} και $p_{avail}(n)$ για τον αριθμό των διαθέσιμων υποφερουσών σε κάθε ζώνη συχνοτήτων. Επίσης, πρέπει να σημειωθεί ότι η $p_{avail}(n)$ είναι η ίδια για κάθε κανάλι σε κάποια ζώνη συχνοτήτων. Όλα τα σχήματα μάθησης που χρησιμοποιήθηκαν στην προσομοίωση δείχνουν το μέσο όρο των 2000 επεισοδίων και για κάθε επεισόδιο και κάθε ζώνη συχνοτήτων η $p_{avail}(n)$ υπέστη δειγματοληψία από μια ξεχωριστή κατανομή που έμεινε σταθερή για όλοκληρο το επεισόδιο και επομένως προσομοιώνοντας ένα στατικό σενάριο.

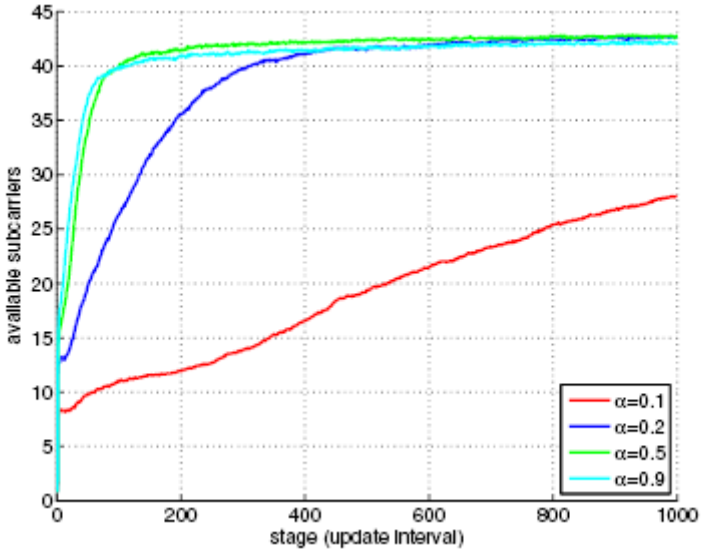
Οι γραφικές παραστάσεις του Σχ.2.10 δείχνουν τα αποτελέσματα της προσομοίωσης για την απλή εκδοχή του προβλήματος. Η βέλτιστη εκτέλεση παρατηρείται για α=0.1 και τ=0.2 τα οποία είναι αποτελέσματα από τις αρχικές ανταμοιβές Q για το μέγιστο αριθμό των πιθανώς διαθέσιμων καναλιών.



Σχήμα 2.10: Διατύπωση απλού προβλήματος: μέσος όρος μάθησης καμπυλών για διαφορετικά α και τ

Αυτό το γεγονός ενθαρρύνει τη διερεύνηση αν η ζώνη συχνοτήτων με τα περισσότερα διαθέσιμα κανάλια έχει ήδη εξεταστεί από το σύστημα δευτερευόντων χρηστών.

Τα αποτελέσματα της προσομοίωσης για το MDP σενάριο και την προτεινόμενη στρατηγική επίλυσης που περιγράφηκε προηγουμένως φαίνεται στις γραφικές παραστάσεις του Σχ.2.11.



Σχήμα 2.11: μέθοδος actor/critic: μέσος όρος διαθέσιμων υποφερουσών για διαφορετικά α με βάση διαφορετικά στάδια του επεισοδίου

Οι προσομοιώσεις δείχνουν το κέρδος εκτέλεσης για αύξηση του α με (β=0.1 , β_1=0.1 και $k = \frac{1}{2N_{fb}}$). Για μεγάλα α η εκτέλεση είναι (μέσος όρος διαθέσιμων υποφερουσών=43, έπειτα από 1000 στάδια), αλλά δεν είναι τόσο καλή η εκτέλεση όσο στο απλό σενάριο (μέσος όρος διαθέσιμων υποφερουών=46-47 για 1000 στάδια). Τα αποτελέσματα του αλγορίθμου που εξετάστηκε είναι καλά όταν επιλέγεται ένα κατάλληλο διάνυσμα μάθησης α. Η επιλογή του πρέπει να γίνεται προοδευτικά καθώς το σύστημα δευτερευόντων χρηστών αναπτύσσεται σε ένα γνωστό περιβάλλον ή με άλλα λόγια πρέπει να ρυθμίζεται δυναμικά.

2.5.4 Επίλογος

Στην εφαρμογή που μελετήθηκε παρουσιάζεται ένα σχήμα βασισμένο στην ενισχυτική μάθηση για την ανίχνευση πηγών φάσματος σε ένα σενάριο γνωστικών ραδιοεπικοινωνιών πολλών ζωνών φάσματος. Με βάση τις υποθέσεις του σεναρίου, αναπτύχθηκε ένα απλό και ένα περισσότερο σύνθετο μοντέλο για την περιγραφή της ανίχνευσης των πηγών φάσματος από το σύστημα δευτερευόντων χρηστών και προτάθηκε μια στρατηγική επίλυσης του προβλήματος βασισμένη στη μέθοδο actor/critic. Τα αποτελέσματα της προσομοίωσης δείχνουν ότι ο αλγόριθμος που παρουσιάστηκε έχει την ικανότητα να αναγνωρίσει γρήγορα τις ζώνες συχνοτήτων με τις περισσότερες διαθέσιμες πηγές. Η δομή του αλγορίθμου αναπτύχθηκε κατά τρόπο που καθιστά ικανή μια εύκολη ολοκλήρωση μέσα από το πλαίσιο βελτιστοποίησης του cross layer και επίσης είναι κατάλληλο για λειτουργικότητα σε ένα περιβάλλον με δυναμική αλλάγή διαθεσιμότητας των πηγών φάσματος.

ΚΕΦΑΛΑΙΟ 3^{ο}: ΕΦΑΡΜΟΓΕΣ ΕΝΙΣΧΥΤΙΚΗΣ ΜΑΘΗΣΗΣ

Στο παρόν κεφάλαιο παρουσιάζονται τα αποτελέσματα και αναλύονται τα συμπεράσματα από τις εφαρμογές διαφόρων τεχνικών ενισχυτικής μάθησης σε συστήματα γνωστικών ραδιοεπικοινωνιών. Γίνεται παρουσίαση τριών χαρακτηριστικών σχεδίων ενισχυτικής μάθησης, των οποίων τα αποτελέσματα αποτελούν αντικείμενο μελέτης και συμβάλλουν στην εξέλιξη της τεχνολογίας των τηλεπικοινωνιών για την επίτευξη του στόχου της αποδοτικής δυναμικής πρόσβασης στο φάσμα. Στόχος είναι η συγκέντρωση βασικών αποτελεσματικών χαρακτηριστικών κάθε σχεδίου ώστε να χρησιμοποιηθούν στην προσομοίωση του επόμενου κεφαλαίου.

3.1 Σχήμα κατανεμημένης εκχώρησης φάσματος (Distributed spectrum sharing scheme)

Παρουσιάζεται ένα νέο σχέδιο εφαρμογής ενισχυτικής μάθησης σε συστήματα γνωστικών ραδιοεπικοινωνιών, γνωστό ως σχήμα κατανεμημένης εκχώρησης φάσματος (distributed spectrum sharing scheme). Αυτό το σχήμα παρέχει τη δυνατότητα μείωσης της ανάγκης για ανίχνευση φάσματος (spectrum sensing) στο χρήστη γνωστικών ραδιοεπικοινωνιών, η οποία επιτυγχάνεται χρησιμοποιώντας την εμπειρία της ενισχυτικής μάθησης. Έτσι, αντί για πραγματοποίηση ανίχνευσης σε ολόκληρο το διαθέσιμο φάσμα αυθαιρέτως ώστε στη συνέχεια να επιχειρηθεί πρόσβαση σε αυτό, το υπό μελέτη σχήμα είναι βασισμένο σε μια στρατηγική βέλτιστης ανίχνευσης και πρόσβασης στο φάσμα που αναλύεται στις επόμενες παραγράφους.

3.1.1 Συνάρτηση ανταμοιβής

Ένα βασικό στοιχείο της ενισχυτικής μάθησης είναι η συνάρτηση ανταμοιβής. Ένας χρήστης γνωστικών ραδιοεπικοινωνιών ανανεώνει τη στρατηγική δράσης του η οποία στηρίζεται στην ανατροφοδότηση της συνάρτησης ανταμοιβής του. Η ακόλουθη γραμμική συνάρτηση χρησιμοποείται στο υπό μελέτη σχήμα

$$W_t = f_1 * W_{t-1} + f_2 \tag{3.1}$$

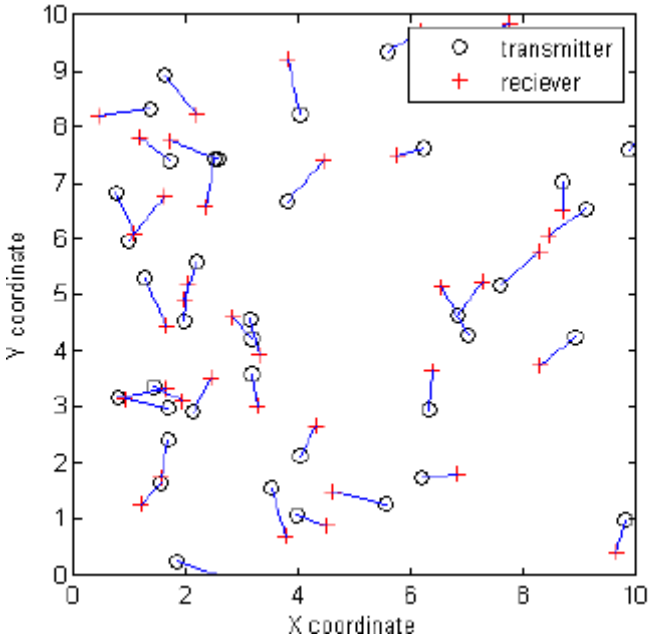
όπου W_{t-1} και W_t είναι τα βάρη προτεραιότητας του καναλιού (φασματικού πόρου) τη στιγμή t-1 και t αντίστοιχα. Μεγαλύτερο βάρος συνεπάγεται μεγαλύτερη προτεραιότητα πρόσβασης στο κανάλι. Επίσης, f_1, f_2 είναι οι παράγοντες βάρους οι οποίοι έχουν διαφορετικές τιμές σύμφωνα με την τοπική κρίση των διαφόρων καταστάσεων του συστήματος και του περιβάλλοντος όπως φαίνεται στον Πίνακα 3.1.

f_1		f_2	
Reward	Punishment	Reward	Punishment
1	1	1	-1

Πίνακας 3.1 Τιμές παραγόντων βάρους

3.1.2 Βήματα αλγορίθμου

Παρουσιάζονται τα βήματα του αλγορίθμου ανίχνευσης του φάσματος για την εξεύρεση του βέλτιστου συνόλου καναλιών για κάθε χρήστη γνωστικών ραδιοεπικοινωνιών μέσω της γνώσης που προσφέρει η ενισχυτική μάθηση. Θεωρούνται ως χρήστες γνωστικών ραδιοεπικοινωνιών (cognitive radio users) ή CR χρήστες ένα σύνολο ζευγών κόμβων πομπών-δεκτών οι οποίοι καλούνται U_i.



Σχήμα 3.1 Δείγμα των ζευγαριών πομών-δεκτών των CR χρηστών

Προκαταρκτικό στάδιο (pre-play) : Στο στάδιο αυτό, οι CR χρήστες βρίσκονται σε αναζήτηση βέλτιστων πηγών από το διαθέσιμο φάσμα και μαθαίνουν από την εμπειρία αυτής της αναζήτησης. Η εξερεύνηση του διαθέσιμου φάσματος γίνεται προσεγγίζοντας όλα τα κανάλια με την ίδια πιθανότητα πρόσβασης σε αυτά. Τα βάρη των καναλιών τα οποία θα χρησιμοποιηθούν τροποποιούνται έπειτα από κάθε δράση ανάλογα με τη συνάρτηση ανταμοιβής.

1) *Έλεγχος κατωφλίου βάρους*. Στο βήμα αυτό οι χρήστες U_i εκτιμούν το τοπικό σύστημα για την εύρεση του βέλτιστου συνόλου πόρων. Ορίζεται ένα προτιμώμενο κατώφλι βάρους (W_{thres}) και σε κάθε προσπάθεια επικοινωνίας ο CR χρήστης συγκρίνει το βάρος το οποίο τον αντιπροσωπεύει με το βάρος κατωφλίου. Εάν το βάρος του χρήστη είναι μεγαλύτερο από το W_{thres}, ο χρήστης θεωρεί το εξεταζόμενο

κανάλι ως προτιμώμενο κανάλι, δηλαδή κατάλληλο κανάλι για μετάδοση. Σε αντίθετη περίπτωση, συνεχίζεται η ανίχνευση.

2) *Ανίχνευση Παρεμβολής.* Αυτό το βήμα αποτελεί συνέχεια του προηγούμενου βήματος. Ο χρήστης ανιχνεύει το επίπεδο παρεμβολής I ορίζοντας και σε αυτή την περίπτωση ένα κατώφλι παρεμβολής I_{thr}**.** Αν ικανοποιείται η συνθήκη $I<I_{thr,}$ ο χρήστης U_i χρησιμοποιεί το κανάλι, αλλιώς το βάρος του καναλιού που εξετάζεται μειώνεται και ο χρήστης U_i ξεκινά την ανίχνευση σε επόμενο κανάλι.

3) *Μέτρηση **SINR** (λόγου ισχύος προς παρεμβολή)*. Ο σκοπός της μέτρησης του SINR είναι να διατηρηθεί η ποιότητα υπηρεσίας των καναλιών. Και σε αυτό το βήμα ορίζεται κατώφλιο $SINR_{thr}$**.** Εάν ο SINR του ενεργοποιημένου καναλιού είναι μεγαλύτερος από τον $SINR_{thr,}$ ο χρήστης U_i χρησιμοποιεί επιτυχώς το κανάλι (φάσμα) με συνέπεια να μεγαλώσει το βάρος του καναλιού κατά συγκεκριμένο παράγοντα f. Στην αντίθετη περίπτωση, ο χρήστης αποκλείεται και το βάρος ανανεώνεται με μία ποινή λάθους ώστε να μειωθεί η προτεραιότητα πρόσβασης του καναλιού.

Εξετάζονται τρία σχήματα :

1) σχήμα πλήρους ανίχνευσης (full sensing scheme) σύμφωνα με το οποίο οι χρήστες γνωστικών ραδιοεπικοινωνιών εξετάζουν το φάσμα στο ξεκίνημα της δράσης τους.

2) σχήμα συγκρατημένης ανίχνευσης (restricted sensing scheme) σύμφωνα με το οποίο οι χρήστες γνωστικών ραδιοεπικοινωνιών πραγματοποιούν ανίχνευση φάσματος στην επιλεγμένη ως βέλτιστη πηγή.

3) σχήμα ελάχιστης ανίχνευσης (minimum sensing scheme) σύμφωνα με το οποίο οι χρήστες γνωστικών ραδιοεπικοινωνιών χρησιμοποιούν ακριβώς τις προηγούμενες φασματικές ζώνες που είχαν χρησιμοποιήσει σε προηγούμενο στάδιο ανίχνευσης για να επικοινωνούν χωρίς νέες απαιτήσεις ανίχνευσης.

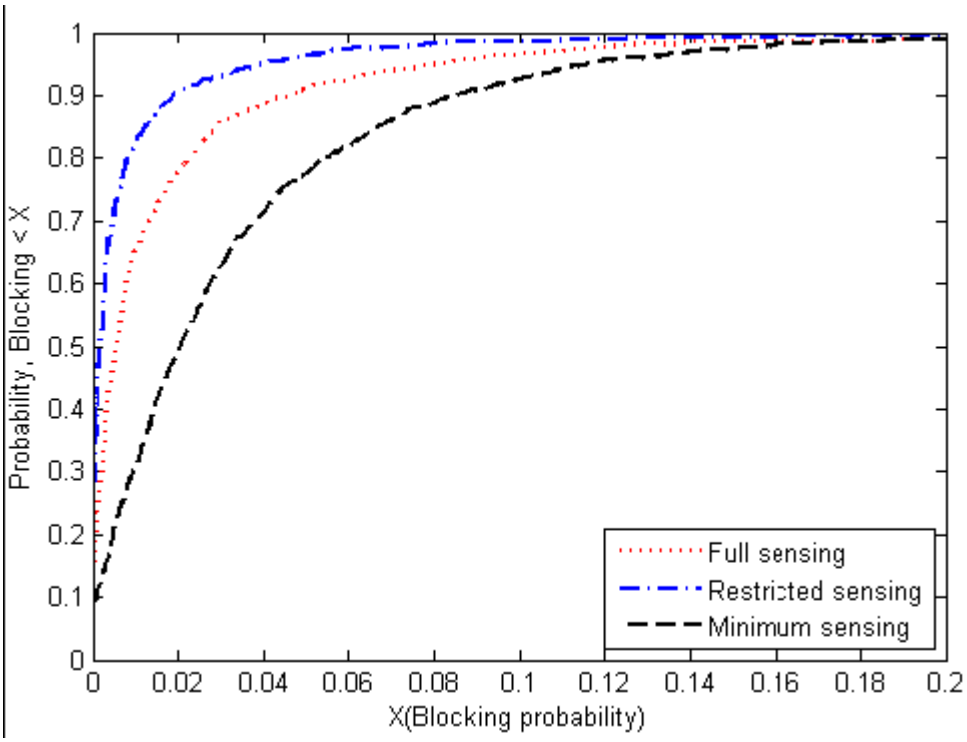
Για την κατανόηση των αποτελεσμάτων του υπό μελέτη συστήματος επικοινωνιών, παρουσιάζεται μια συνάρτηση αθροιστικής κατανομής (cumulative distribution function CDF) τόσο για την πιθανότητα φραγής (blocking) του συστήματος όσο και για την πιθανότητα διακοπής (dropping) επικοινωνίας. Τονίζεται ότι στην προσομοίωση που αναλύεται όλες οι παράμετροι είναι ακριβώς ίδιες για την αξιολόγηση καθενός από τα τρία σχήματα. Το μοντέλο διάδοσης Okumura-Hata χρησιμοποιείται με τυπική απόκλιση 8dB.

- 1000 ζεύγη χρηστών CR κατανέμονται ομοιόμορφα σε μία έκταση $1000km^2$
- Σε κάθε γεγονός μια τυχαία ακολουθία ζευγών ενεργοποιείται, ορίζεται ο αριθμός των 400 ως ο μέγιστος αριθμός της τυχαίας ακολουθίας.
- 100 κανάλια είναι διαθέσιμα για επικοινωνία και από αυτά το 5% είναι ελεύθερα για επικοινωνία.
- Η φέρουσα συχνότητα είναι 300MHz

- Ύψος κεραίας των ζευγών επικοινωνίας στα 30 m.
- Ισχύς πομπού 1 Watt
- Κέρδος κεραιών πομπού και δέκτη ζευγών επικοινωνίας στα 0 dbi
- SNR_{thres}=10db.
- Επίπεδο θορύβου στα -124dbm
- Θερμοκρασία δέκτη 300K

3.1.3 Αποτελέσματα προσομοίωσης

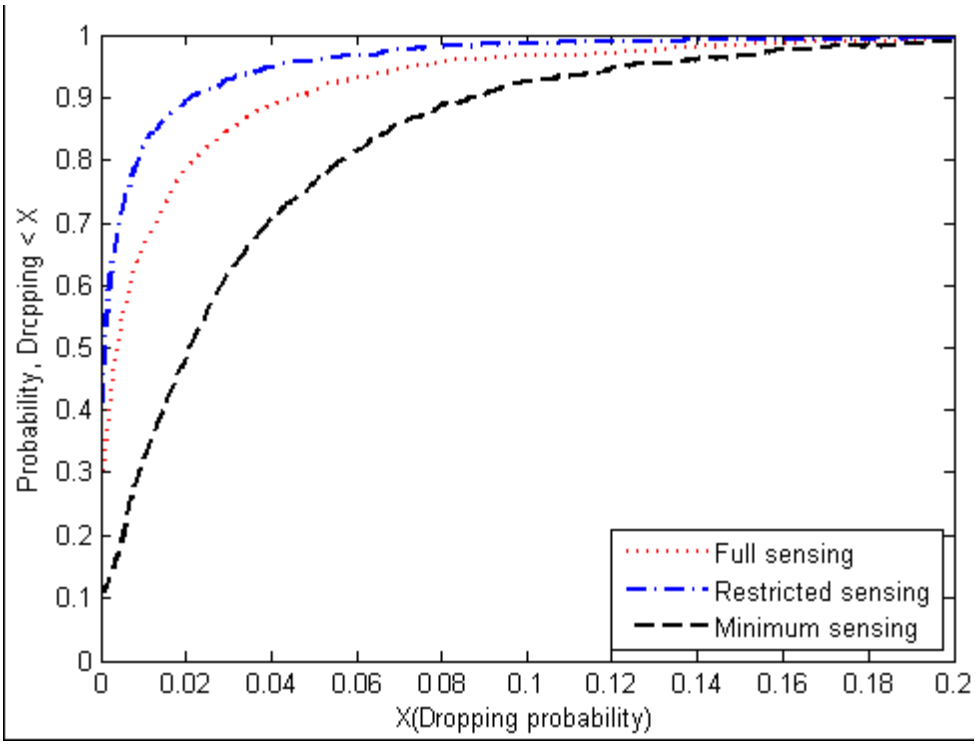
Η γραφική παράσταση του Σχ.3.2 επεξηγεί την πιθανότητα φραγής (blocking probability).



Σχήμα 3.2: Συνάρτηση CDF για φραγή του συστήματος

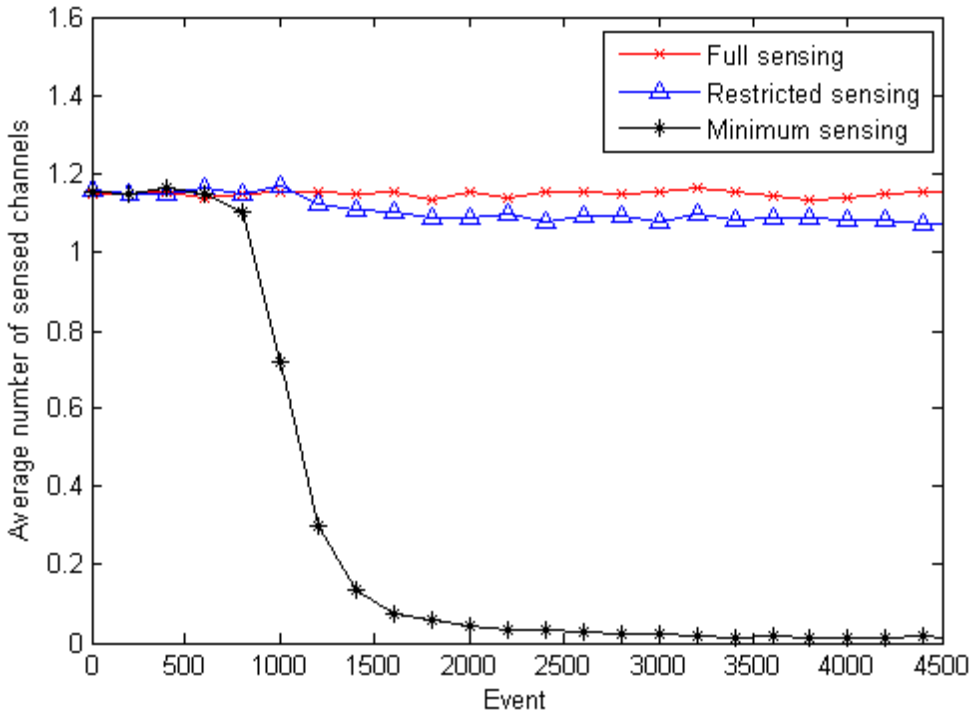
Φαίνεται ότι η πιθανότητα φραγής περίπου 70% των χρηστών στο σχήμα ελάχιστης ανίχνευσης είναι κάτω από την τιμή πιθανότητας p=0.04. Αντίθετα στο σχήμα πλήρους ανίχνευσης και στο σχήμα συγκρατημένης ανίχνευσης είναι γύρω στο 87% και 95% αντίστοιχα η πιθανότητα φραγής για την ίδια p=0.04. Συκρίνοντας το σχήμα ελάχιστης ανίχνευσης με το σχήμα πλήρους ανίχνευσης η πιθανότητα φραγής του του πρώτου είναι υψηλότερη, κάτι που είναι λογικό διότι ένα σχήμα το οποίο πάντα επιλέγει ένα ελεύθερο κανάλι για να λειτουργήσει έχει καλύτερα αποτελέσματα από ένα σχήμα το οποίο περιστασιακά διαλέγει ένα κανάλι χωρίς ανίχνευση. Επίσης φαίνεται ότι το σχήμα συγκρατημένης ανίχνευσης έχει και την καλύτερη συμπεριφορά ως προς την πιθανότητα φραγής, διότι ο χρήστης είναι ικανός να ανιχνεύσει τα κανάλια τα οποία έχουν υψηλότερη πιθανότητα να επιτύχουν (επικοινωνία) σύμφωνα με την προηγούμενη εμπειρία την οποία έχει αποκτήσει, το οποίο είναι και το ζητούμενο στο υπό μελέτη σύστημα επικοινωνιών. Είναι φανερό ότι σε κάθε σχήμα υπάρχουν περίπου 2% των χρηστών οι οποίοι έχουν πιθανότητα φραγής πάνω από 0.2. Αυτό συμβαίνει επειδή υπάρχουν χρήστες που βρίσκονται είτε σε περιοχές υψηλής πυκνότητας ή σε χώρους που υποφέρουν από σημαντική σκίαση.

Η γραφική παράσταση που ακολουθεί επεξηγεί την πιθανότητα διακοπής (dropping probability) που ακολουθούν τα τρία σχήματα ανίχνευσης του συστήματος επικοινωνιών που μελετάται.



Σχήμα 3.3: Συνάρτηση CDF για διακοπή του συστήματος

Τα συμπεράσματα που προκύπτουν από τη μελέτη της γραφικής παράστασης της συνάρτησης για τη διακοπή του συστήματος είναι αντίστοιχα με αυτά της προηγούμενης γραφικής παράστασης του Σχ.3.2.



Σχήμα 3.4: Μέσο πλήθος καναλιών ανίχνευσης

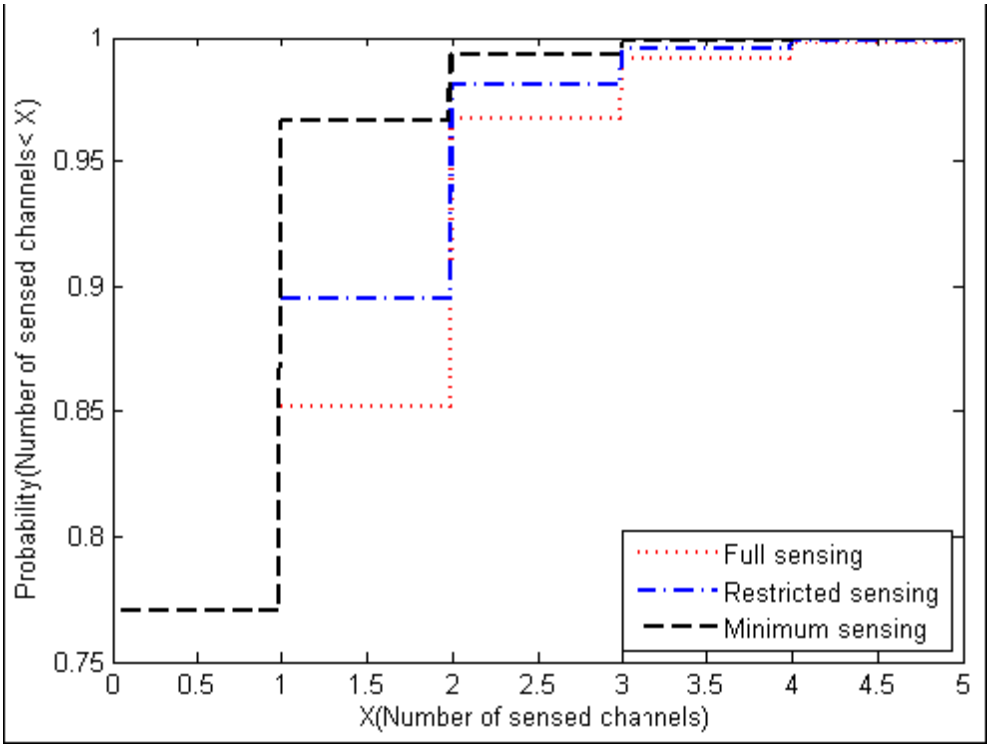
Η γραφική παράσταση του Σχ.3.4 απεικονίζει το μέσο πλήθος των καναλιών που οι χρήστες γνωστικών ραδιοεπικοινωνιών πρέπει να ανιχνεύσουν ως συνάρτηση των γεγονότων-προσπαθειών της προσομοίωσης. Παρατηρείται ότι στο σχέδιο πλήρους ανίχνευσης οι κόμβοι δεν σταματούν ποτέ την ανίχνευση καναλιών και επιλέγουν το φάσμα σε τυχαία βάση. Η κόκκινη γραμμή της γραφικής παράστασης που αντιπροσωπεύει τη συμπεριφορά αυτή διατηρεί τη θέση της, γύρω στο 1.15 σε όλη τη διάρκεια της προσομοίωσης.

Όσο αφορά το σχήμα συγκρατημένης ανίχνευσης πού αντιστοιχεί στην μπλε γραμμή της γραφικής παράστασης συγκλίνει στο ένα κανάλι το οποίο αποτελεί την ιδανική κατάσταση για αυτό το σχέδιο. Αυτό ισχύει διότι ο μέσος όρος των καναλιών στα οποία θα πραγματοποιηθεί ανίχνευση στο σχέδιο συγκρατημένης ανίχνευσης σε κάθε κατάσταση-γεγονός δεν μπορεί να είναι μικρότερο από ένα. Αυτό συμβαίνει διότι δεν υπάρχει ούτε στην περίπτωση αυτή διακοπή της ανίχνευσης για ελεύθερο κανάλι. Τόσο στο σχέδιο πλήρους όσο και στο σχέδιο συγκρατημένης ανίχνευσης, οι χρήστες CR εφαρμόζουν το προκαταρκτικό στάδιο. Συνεπώς, όλες οι προσπάθειες για πρόσβαση στο φάσμα απαιτούν την ανίχνευση ενός τουλάχιστον καναλιού.

Τέλος, η συμπεριφορά του σχήματος ελάχιστης ανίχνευσης μπορεί να χωριστεί σε τρείς περιόδους. Η πρώτη περίοδος είναι από το γεγονός 1 έως το γεγονός 600. Οι χρήστες σε αυτή την περίοδο είναι όλοι στο προκαταρκτικό στάδιο. Η δεύτερη περίοδος είναι από το γεγονός 600 μέχρι το γεγονός 2000. Οι ανάγκες για ανίχνευση φάσματος είναι δραστικά μειωμένες σε αυτή την περίοδο. Έπειτα από ένα βέβαιο χρόνο προσομοίωσης μια ισορροπημένη κατανομή φάσματος

αποκαθίσταται από την απαίτηση του αλγόριθμου της ενισχυτικής μάθησης. Οι χρήστες γνωστικών ραδιοεπικοινωνιών ξεκινούν να εισέρχονται στο φάσμα στις προτιμώμενες ομάδες καναλιών χωρίς ανίχνευση. Στην τρίτη περίοδο παρατηρώντας τη γραμμή που αντιπροσωπεύει το σχήμα ελάχιστης ανίχνευσης, πλησιάζει την τιμή 0,03 που σημαίνει ότι η κατάσταση του συστήματος είναι σταθερή. Έπειτα από την ισορροπία της κατανομής φάσματος, οι χρήστες γνωστικών ραδιοεπικοινωνιών έχουν την ικανότητα να αποφύγουν συγκρούσεις, χρησιμοποιώντας την εμπειρία της μάθησης. Αυτό έχει ως συνέπεια τη δραστική μείωση των αναγκών για εξεύρεση νέων πόρων φάσματος.

Συγκρίνοντας το σχήμα πλήρους ανίχνευσης με το σχήμα συγκρατημένης ανίχνευσης παρατηρείται ότι η σπατάλη χρόνου και η κατανάλωση ενέργειας είναι 5% χαμηλότερη στο δεύτερο. Στο σχέδιο ελάχιστης ανίχνευσης το μέσο πλήθος καναλιών τα οποία ανιχνεύονται είναι στο 23% του σχήματος πλήρους ανίχνευσης. Βεβαίως, μετά το γεγονός 2000 είναι μόλις στο 1.72% εξαιτίας της μείωσης των αναγκών για ανίχνευση με την επίδραση της ενισχυτικής μάθησης. Έτσι, καθίσταται εμφανές το πλεονέκτημα του σχήματος ενισχυτικής μάθησης που αναλύεται. Το πλεονέκτημα αυτό είναι ότι μειώνεται το πλήθος των καναλιών στα οποία πραγματοποιείται ανίχνευση από τους χρήστες γνωστικών ραδιοεπικοινωνιών όσο προχωρά η εξέλιξη της μάθησης. Το πλήθος καναλιών στα οποία έχει πραγματοποιηθεί ανίχνευση περιγράφει αποτελεσματικά την δαπάνη χρόνου (time consuming) και την κατανάλωση ενέργειας (power intensive) της διαδικασίας ανίχνευσης φάσματος. Επίσης, η γραφική παράσταση του Σχ.3.4 δείχνει τη σύγκλιση της συμπεριφοράς του σχήματος ενισχυτικής μάθησης. Από το πρώτο γεγονός της προσομοίωσης έως το γεγονός 2000 το σχήμα μάθησης συγκλίνει στην ιδανική στρατηγική διανομής φάσματος (spectrum sharing).



Σχήμα 3.5: Συνάρτηση CDF για τον αριθμό των καναλιών που ανιχνεύονται

Η γραφική παράσταση του Σχ.3.5 είναι η συνάρτηση αθροιστικής κατανομής του πλήθους των καναλιών τα οποία ένας χρήστης γνωστικών ραδιοεπικοινωνιών ανιχνεύει. Προκύπτει ότι το 99% των προσπαθειών για ανίχνευση φάσματος επιτυγχάνεται με την ανίχνευση τεσσάρων καναλιών. Δηλαδή ο χρήστης γνωστικών ραδιοεπικοινωνιών πρέπει να δοκιμάσει το πολύ τέσσερα διαφορετικά κανάλια για να επιτύχει πρόσβαση. Σε κάποιο από αυτά τα τέσσερα κανάλια θα εντοπίσει το προτιμητέο κανάλι που του υποδεικνύει η ενισχυτική μάθηση.

3.1.4 Συμπεράσματα

Σε αυτό το μοντέλο που παρουσιάστηκε προτείνεται ένας τρόπος ενισχυτικής μάθησης βασιζόμενος στην κατανεμημένη εκχώρηση φάσματος για συστήματα γνωστικών ραδιοεπικοινωνιών στα οποία υπάρχει η δυνατότητα μείωσης των αναγκών για ανίχνευση φάσματος. Χρησιμοποιώντας την ικανότητα της μάθησης οι πράκτορες γνωστικών ραδιοεπικοινωνιών μπορούν να κρατούν ιστορικό των βέλτιστων πόρων πρόσβασης στο φάσμα. Αυτή η ικανότητα καθιστά δυνατή την αποτελεσματική προσέγγιση της βέλτιστης ανίχνευσης και πρόσβασης στο φάσμα. Αυτή η βελτίωση της ανίχνευσης του διαθέσιμου φάσματος οδηγεί σε εξοικονόμηση χρόνου και ενέργειας.

3.2 Ενισχυτική μάθηση συνεργασίας πολλών πρακτόρων (Multi-agent reinforcement learning)

3.2.1 Ενισχυτική μάθηση συνεργασίας πολλών πρακτόρων για ισορροπημένη κατανομή υπηρεσιών (Adaptive load balancing- Multi-agent learning)

Σε αυτό το σχήμα μελετάται η ενισχυτική μάθηση με συνεργασία πολλών πρακτόρων (multi-agent reinforcement learning), με σκοπό την επίτευξη ισορροπημένης κατανομής υπηρεσιών (load balancing) σε ένα τηλεπικοινωνιακό σύστημα. Ορίζεται ένα στοχαστικό σύστημα το οποίο εμπλέκει ένα σύνολο από πράκτορες (agents) και ένα σύνολο από πηγές παροχής υπηρεσιών επικοινωνίας (resources). Ο στόχος είναι η επίτευξη της πλέον αποδοτικής εκτέλεσης των υπηρεσιών που αναμένεται να εξυπηρετήσει το σύστημα μέσω της συνεργασίας των πρακτόρων που αναλαμβάνουν την περάτωση των υπηρεσιών. Μεταβάλλοντας με πιθανοτικό τρόπο (είναι πιθανότερο να περατωθεί μια υπηρεσία στην πηγή με τη μεγαλύτερη χωρητικότητα) τις χωρητικότητες των πηγών πραγματοποιείται αντιστοίχως πιθανοτική μεταβίβαση νέων υπηρεσιών (jobs) στους πράκτορες. Ένας πράκτορας πρέπει να επιλέγει μια πηγή για την εξυπηρέτηση κάθε νέας υπηρεσίας. Η έννοια της ισορροπημένης κατανομής υπηρεσιών αντιστοιχεί στην επίτευξη της καλύτερης δυνατής χρησιμοποίησης των πόρων του συστήματος επικοινωνιών. Το σύστημα επικοινωνιών που εξετάζεται έχει βέλτιστη επίδοση όταν οι υπηρεσίες που αναλαμβάνουν οι πράκτορες περατώνονται (εκτελούνται) το συντομότερο δυνατό στις πηγές του συστήματος.

3.2.2 Η γενική θέση του συστήματος

Αρκετές ερευνητικές προσπάθειες στην επιστήμη διαχείρισης (management science) και στα κατανεμημένα συστήματα τεχνητής νοημοσύνης (distributed artificial intelligence) υιοθετούν μια διαφορετική σκοπιά στα κατανεμημένα συστήματα πρακτόρων–πηγών. Οι πράκτορες είναι αυτόνομες οντότητες που διαπραγματεύονται μεταξύ τους. Αντίθετα, στο υπό μελέτη σύστημα κατανομής υπηρεσιών υιοθετείται αυστηρός διαχωρισμός μεταξύ πρακτόρων και πηγών. Οι πηγές είναι παθητικές και δεν λαμβάνουν αποφάσεις. Οι πράκτορες ενεργούν σε ένα καθαρά τοπικό περιβάλλον. Η ισορροπημένη κατανομή υπηρεσιών επιτυγχάνεται μέσω της επικοινωνίας των ενεργών πρακτόρων ενισχυτικής μάθησης.

Το πλαίσιο του συστήματος ορίζεται ως ακολούθως:

- $A=(\alpha_1,\alpha_2,...,\alpha_N)$ *είναι το σύνολο των πρακτόρων (agents)*
- $R=(r_1,r_2,...,r_N)$ *είναι το σύνολο των πηγών (resources)*
- $P: A \times N \rightarrow [0,1]$ *συνάρτηση απόδοσης υπηρεσίας (job submission function)*
- $D: A \times N \rightarrow R$ *είναι μια πιθανοτική συνάρτηση μεγέθους υπηρεσίας*

- *C: $R \times N \rightarrow R$ είναι μια πιθανοτική συνάρτηση χωρητικότητας*

Κάθε μία από τις πηγές έχει μια συγκεκριμένη χωρητικότητα, η οποία είναι ένας πραγματικός αριθμός που μεταβάλλεται με το χρόνο σύφωνα με τη συνάρτηση C. Κάθε χρονική στιγμή ένας πράκτορας χαρακτηρίζεται ως αδρανής (idle) ή ως απασχολημένος (engaged). Όταν είναι αδρανής περατώνει μια νέα υπηρεσία με πιθανότητα που καθορίζεται από τη συνάρτηση P. Κάθε υπηρεσία έχει ένα συγκεκριμένο μέγεθος το οποίο είναι επίσης πραγματικός αριθμός. Το μέγεθος αυτό καθορίζεται από τη συνάρτηση D. Θα χρησιμοποιηθεί ο όρος ένδειξη (token) που χαρακτηρίζει το συνδυασμό του μεγέθους υπηρεσίας και των χωρητικοτήτων των πηγών. Για κάθε νέα υπηρεσία ο πράκτορας επιλέγει μια από τις πηγές ώστε να την περατώσει (εκτελέσει). Η επιλογή αυτή πραγματοποιείται σύμφωνα με τον κανόνα επιλογής (selection rule) ο οποίος αναλύεται στην επόμενη παράγραφο.

Οποιαδήποτε υπηρεσία έχει τη δυνατότητα να περατωθεί σε οποιαδήποτε πηγή. Επιπλέον, δεν υπάρχουν όρια στο πλήθος των υπηρεσιών που εξυπηρετούνται συγχρόνως από μια πηγή. Βεβαίως η ποιότητα της παρεχόμενης υπηρεσίας από μια πηγή σε δεδομένη χρονική στιγμή χειροτερεύει όσο αυξάνεται ο αριθμός των πρακτόρων που τη χρησιμοποιούν την ίδια χρονική στιγμή. Ο χρόνος που απαιτείται για να εκτελεστεί μια υπηρεσία εξαρτάται από το μέγεθός της που καθορίζεται από τη συνάρτηση D, τη χωρητικότητα της πηγής και από τον αριθμό των άλλων πρακτόρων που χρησιμοποιούν τη συγκεκριμένη πηγή. Σκοπός της διαδικασίας απόδοσης των υπηρεσιών σε πηγές του συστήματος είναι να ελαττωθεί κατά το δυνατό, ο μέσος χρόνος περάτωσης όλων των υπηρεσιών. Αυτό, συνήθως, προϋποθέτει δικαιοσύνη στο διαμοιρασμό υπηρεσιών στο σύστημα.

3.2.3 Προσαρμοσμένοι κανόνες επιλογής πηγών (Selections rules)

Ο κανόνας, σύμφωνα με τον οποίο οι πράκτορες επιλέγουν μια πηγή για την περάτωση μιας νέας υπηρεσίας, είναι η βάση του υπο μελέτη προσαρμοσμένου συστήματος. Αρχικά γίνεται η υπόθεση ομοιογένειας, δηλαδή ότι όλοι οι πράκτορες χρησιμοποιούν τον ίδιο κανόνα επιλογής. Βεβαίως, κάθε πράκτορας βασίζεται στη δική του τοπική πληροφόρηση καθώς και τη δική του εμπειρία για το σύστημα. Η εμπειρία του πράκτορα εξάγεται από τις προηγούμενες υπηρεσίες τις οποίες έχει περατώσει. Τα στοιχεία που ο πράκτορας έχει στη διάθεσή του είναι το όνομα της χρησιμοποιούμενης πηγής (r) καθώς και οι χρονικές στιγμές έναρξης (t_{start}) και περάτωσης (t_{stop}) μιας υπηρεσίας . Η είσοδος του κανόνα επιλογής είναι ένα σύνολο κόμβων με την εξής μορφή πληροφορίας (r, t_{start} t_{stop}). Οποτεδήποτε ο πράκτορας επιλέξει μια πηγή για την εκτέλεση μιας υπηρεσίας, αυτό έχει ως αποτέλεσμα ο πράκτορας να λάβει ανατροφοδότηση από την πηγή μετά την περάτωση της υπηρεσίας. Αυτό αποτελεί τμήμα του ιστορικού του. Κάθε πράκτορας A συμπυκνώνει το συνολικό ιστορικό του σε ένα διάνυσμα που καλείται υπολογιστής αποδοτικότητας (efficiency estimator) και συμβολίζεται ως ee_A. Το μέγεθος του

διανύσματος αυτού είναι ο αριθμός των πηγών που εξυπηρετεί ο εκάστοτε πράκτορας, δηλαδή σε πόσες πηγές έχει αποδώσει υπηρεσίες για περάτωση. Επίσης, κάθε πράκτορας διατηρεί ένα διάνυσμα jd_A στο οποίο αποθηκεύει το πλήθος των υπηρεσιών που έχει περατώσει σε μια συγκεκριμένη πηγή. Για να επιλέξει ο πράκτορας μια πηγή για να υλοποιήσει μια νέα υπηρεσία πρέπει το διάνυσμα jd_A να είναι όσο το δυνατό μικρότερο, ώστε να είναι μικρός ο αριθμός των ήδη υλοποιημένων υπηρεσιών από τη συγκεκριμένη πηγή.

Σε αυτό το σημείο μελετώνται δύο είδη συστημάτων. Αυτά στα οποία η χωρητικότητα των πηγών παραμένει σταθερή και αυτά στα οποία υπάρχει αλλαγή των χωρητικοτήτων των πηγών και, συνεπώς, νέα δεδομένα στο σύστημα.

Στα συστήματα με σταθερή τη χωρητικότητα των πηγών επιλέγεται ως κανόνας επιλογής αυτός που δίνει προτεραιότητα στις πηγές οι οποίες είχαν καλή επίδοση στο παρελθόν, δηλαδή μεγαλύτερη πιθανότητα να επιτύχουν την εκτέλεση μιας υπηρεσίας. Αυτή η επιλογή είναι γνωστή ως BSCR (the best choice selection rule). Στα συστήματα, όμως, όπου η χωρητικότητα δεν είναι σταθερή και υπεισέρχεται και ο παράγοντας της εκμετάλλευσης πληροφοριών (exploitation) μεταξύ των πρακτόρων, ο κανόνας BSCR δεν συνεπάγεται την καλύτερη δυνατή αξιοποίηση του συστήματος. Πριν ληφθεί η απόφαση για το ποια πηγή θα εκτελέσει την υπηρεσία, το σύστημα πρέπει να λαμβάνει υπόψη του και τα νέα δεδομένα που προκύπτουν από τις αλλαγές των χωρητικοτήτων των πηγών.

3.2.4 Ετερογενείς πληθυσμοί Πρακτόρων (Heterogeneous populations)

Μέχρι το σημείο αυτό έχει γίνει η υπόθεση ότι όλοι οι πράκτορες χρησιμοποιούν τον ίδιο κανόνα επιλογής πηγών για την περάτωση μιας υπηρεσίας. Αυτού του είδους η ομοιογένεια περιγράφει μια κατάσταση στη οποία υπάρχουν κεντρικοί ελεγκτές (offline controllers) οι οποίοι αρχικά ορίζουν τη συμπεριφορά των πρακτόρων και, στη συνέχεια, τους επιτρέπουν να λαμβάνουν αποφάσεις με βάση αυτό το μοντέλο.

Στη συνέχεια, υποτίθεται ότι κάθε πράκτορας έχει τη δυνατότητα να ορίσει τη δική του στρατηγική για τη λήψη αποφάσεων. Γίνεται η θεώρηση ότι ένα ένα τμήμα του πληθυσμού των πρακτόρων χρησιμοποιεί κάποιο κανόνα επιλογής πηγών ενώ ένα άλλο τμήμα του πληθυσμού χρησιμοποιεί άλλο κανόνα επιλογής. Πλέον, η επίδοση του συστήματος βασίζεται στα στοιχεία της εκμετάλλευσης και της εξερεύνησης που συσχετίζονται μέσω της αλληλεπίδρασης της συμπεριφοράς των διαφορετικών πληθυσμών πρακτόρων. Στην περίπτωση όπου δύο πληθυσμοί πρακτόρων αλληλεπιδρούν, η επίδοση του συστήματος βασίζεται στη συνεργασία που θα αναπτύξουν. Ο αποτελεσματικότερος τρόπος περάτωσης υπηρεσιών στις πηγές για το υπό μελέτη προσαρμοσμένο σύστημα, προκύπτει όταν

πραγματοποιείται αμοιβαία εξερεύνηση και εκμετάλλευση από τους διαφορετικούς πληθυσμους, και ο ένας επωφελείται από τον άλλο.

Όταν ένας πράκτορας φροντίζει να αποσπά τα καλύτερα αποτελέσματα και, επομένως, την καλύτερη δυνατή επίδοση για τον ίδιο, χαρακτηρίζεται ως μη συνεργατικός (non cooperative). Όμως, στην περίπτωση όπου όλοι οι πράκτορες είναι μη συνεργατικοί, υπάρχει ο κίνδυνος να μην επωφεληθεί κανένας από την εκτέλεση υπηρεσιών στις πηγές του συστήματος. Δηλαδή το ατομικό συμφέρον του πράκτορα δεν συμβαδίζει με το συμφέρον του συνολικού πληθυσμού σαν μία οντότητα. Για την ισορροπία του συστήματος, φροντίζουν πράκτορες οι οποίοι στέλνουν προς περάτωση μια υπηρεσία στην πηγή η οποία έχει το μικρότερο φορτίο από υπηρεσίες εκείνη τη χρονική στιγμή. Οι (load querying agents) δεν δρουν ως παρασιτικά στοιχεία στο σύστημα και είναι υπεύθυνοι για την εξισορρόπηση του φορτίου των πηγών στο σύστημα.

3.2.5 Επικοινωνία μεταξύ των πρακτόρων (Communication among agents)

Μέχρι τώρα έχει γίνει η θεώρηση ότι δεν υπάρχει επικοινωνία μεταξύ των πρακτόρων. Όμως, η ιδέα της ενισχυτικής μάθησης πολλών πρακτόρων αποκτά ακόμα μεγαλύτερο ερευνητικό ενδιαφέρον και γίνεται περισσότερο αποδοτική όταν υπάρχει επικοινωνία μεταξύ των πρακτόρων. Στο σημείο αυτό παρουσιάζεται ένα απλό σχήμα επικοινωνίας μεταξυ πρακτόρων ώστε να γίνουν αντιληπτά τα οφέλη που αυτό προσφέρει.

Γίνεται η υπόθεση ότι κάθε πράκτορας έχει τη δυνατότητα να επικοινωνήσει μόνο με ορισμένους πράκτορες, οι οποίοι καλούνται γείτονες (neighbours). Συνεπώς, ο πληθυσμός των πρακτόρων διαιρείται σε ισοδύναμες αυτοτελείς οντότητες ομάδων πρακτόρων που καλούνται γειτονιές (neighborhoods). Το σχήμα επικοινωνίας το οποίο υιοθετείται βασίζεται στην ιδέα ότι τα διανύσματα υπολογιστικών αποδοτικοτήτων ee_A των πρακτόρων σε μια γειτονιά (όταν γίνεται εκτέλεση μιας υπηρεσίας από κάποιον πράκτορα), διαμοιράζονται μεταξύ των πρακτόρων της γειτονιάς. Επίσης, γίνεται η υπόθεση ότι οι πράκτορες που συγκροτούν μια γειτονιά έχουν τον ίδιο κανόνα επιλογής ενώ πράκτορες που ανήκουν σε διαφορετικές γειτονιές μπορούν να έχουν διαφορετικό κανόνα επιλογής. Το διάνυσμα υπολογιστικής αποδοτικότητας κάθε πράκτορα ανανεώνεται με βάση το μέσο όρο των υπολογοστικών αποδοτικοτήτων όλων των πρακτόρων της γειτονιάς που καλείται υπολογιστική αποδοτικότητα γειτονιάς, (the neighborhood efficiency estimator). Η τιμή του υπολογίζεται κάθε στιγμή από το μέλος-πράκτορα της γειτονιάς που εκτελεί μια νέα υπηρεσία. Για να συγκριθεί η συμπεριφορά των πρακτόρων που επικοινωνούν με αυτή των πρακτόρων που δεν επικοινωνούν, γίνεται η θεώρηση ενός μοναδιαίου (single) πληθυσμού, παραλληλα με τις γειτονιές πρακτόρων που ορίστηκαν. Επίσης, γίνεται η παραδοχή ότι ορισμένες γειτονιές πρακτόρων δεν επιτρέπουν την κατανομή των διανυσμάτων υπολογιστικών

αποδοτικοτήτων ανάμεσα στα μέλη τους. Οι γειτονιές αυτές ονομάζονται μη επικοινωνούσες γειτονιές (non communicating neighborhood, NCN). Ορισμένες γειτονιές επιτρέπουν την κατανομή των διανυσμάτων υπολογιστικών αποδοτικοτήτων και ονομάζονται (communicating neighborhood CN). Η επίδοση του συστήματος και η εξαγωγή συμπερασμάτων εξαρτώνται από τους εξής παράγοντες:

1) Εκμετάλλευση: ένας πράκτορας εκμεταλλεύεται προς όφελός του τις πληροφορίες που παρέχουν οι υπόλοιποι πράκτορες της γειτονιάς που ανήκει περί των πηγών του συστήματος

2) Εξερεύνηση: ορισμένοι πράκτορες που ανήκουν σε μία γειτονιά δεν δρουν με γνώμονα το ατομικό τους συμφέρον αλλά οι επιλογές πηγών που κάνουν για την περάτωση υπηρεσιών αποσκοπούν στην αποκόμιση καλύτερης γνώσης του συστήματος για ολόκληρη τη γειτονιά πρακτόρων

3) Κανόνας επιλογής: όταν μια NCN χρησιμοποιεί ένα συντηρητικό που δεν ευνοεί την ύπαρξη επικοινωνίας κανόνα επιλογής, οι CNs έχουν καλύτερη επίδοση στο σύστημα· άρα, η επικοινωνία μεταξύ τους λειτουργεί προς όφελος των πρακτόρων

Όταν οι πράκτορες-μέλη μιας γειτονιάς δρουν εγωιστικά και φροντίζουν να εξασφαλίζουν την καλύτερη δυνατή εκτέλεση για τους ίδιους, χαρακτηρίζονται ως παρασιτικά στοιχεία και χειροτερεύουν την εκτέλεση των CNs έναντι των NCNs. Σε αυτή την περίπτωση η εκμετάλλευση δρα ενάντια στην εξερεύνηση των πρακτόρων. Οι πράκτορες επιδιώκουν ο καθένας να εξυπηρετήσει το ατομικό του συμφέρον (εκτελώντας μια υπηρεσία στην πηγή με τη μικρότερη χωρητικότητα). Συνεπώς, η CN έχει αρνητικά αποτελέσματα και δεν βελτιώνει τον τρόπο λειτουργίας του συστήματος επικοινωνιών. Μια NCN εξασφαλίζει καλύτερη και αποδοτικότερη λειτουργία ως προς τον τρόπο περάτωσης των υπηρεσιών, διότι σε αυτή την περίπτωση οι πράκτορες δεν μοιράζονται τα διανύσματα υπολογιστικών αποδοτικοτήτων και δρουν προς όφελός τους χωρίς να επηρεάζουν τον τρόπο δράσης της γειτονιάς όπου ανήκουν.

Ένα άλλο πιθανό σενάριο είναι όταν κάποια NCN χρησιμοποιεί ένα κανόνα επιλογής πηγών που είναι πολύ συντηρητικός ενώ μια CN χρησιμοποιεί ένα κανόνα επιλογής που ευνοεί την επικοινωνία μεταξύ των πρακτόρων. Σε αυτή την περίπτωση, οι πράκτορες της CN συνεργάζονται αρμονικά και δρουν προς όφελος της γειτονιάς. Αντίθετα, οι πράκτορες της NCN λόγω του συντηρητικού κανόνα επιλογής δεν εξασφαλίζουν την καλύτερη δυνατή ατομική επίδοση (ως προς την περάτωση υπηρεσιών στις πηγές με τη μικρότερη χωρητικότητα) και κατά συνέπεια και της γειτονιάς ως σύνολο. Συνεπώς, η γειτονιά στην οποία υπάρχει επικοινωνία και ανταλλαγή πληροφορίας μεταξύ των πρακτόρων έχει καλύτερη επίδοση από τη γειτονιά όπου οι πράκτορες δεν επικοινωνούν μεταξύ τους.

Παρατηρείται ότι υπάρχει σε κάθε εκτέλεση του συστήματος ένα trade-off μεταξύ των τριών θεμελιωδών παραγόντων που αναφέρθηκαν. Δηλαδή για να χαρακτηριστεί αποδοτική ή όχι η επικοινωνία μεταξύ των πρακτόρων λαμβάνεται υπόψη το πώς δρουν ξεχωριστά οι πράκτορες εντός της γειτονιάς. Απαιτείται ο κατάλληλος συνδυασμός των τριών παραγόντων (εκμετάλλευση, εξερεύνηση, κανόνας επιλογής) που αναφέρθηκαν ώστε να υπάρξει η αποδοτικότερη δυνατή συνεργασία μεταξύ των πρακτόρων.

3.2.6 Συμπεράσματα

Παρουσιάστηκε η ιδέα της ενισχυτικής μάθησης πολλαπλών πρακτόρων και πώς αυτή επιδρά στην ισορροπημένη κατανομή υπηρεσιών ενός τηλεπικοινωνιακού συστήματος. Μελετήθηκαν οι παράμετροι της προσαρμοστικής συμπεριφοράς πρακτόρων (adaptive behavior) η οποία προκύπτει από τη μορφή του κανόνα επιλογής, της εξερεύνησης και της εκμετάλλευσης μεταξύ των πρακτόρων καθώς και η αλληλεπίδραση μεταξύ τους. Η επικοινωνία των πρακτόρων προϋποθέτει βασικούς κανόνες αλληλεπίδρασης ώστε να είναι αποδοτική και να μην έχει επιζήμια αποτελέσμτα για τους πράκτορες που επικοινωνούν. Στόχος είναι να γίνει ένα περαιτέρω βήμα σε σχέση με τη λειτουργία όπου ο κάθε πράκτορας εμπιστεύεται μόνο τη δική του τοπική πληροφορία. Δηλαδή κάθε πράκτορας πλέον να δρα για να επιτευχθεί ένα συνολικά καλό αποτέλεσμα για ολόκληρο τον πληθυσμό πρακτόρων, και να τροποποιεί αντίστοιχα τη δράση που αποτελεί την καλύτερη επιλογή για το δικό του μόνο συμφέρον. Ένα σύστημα που προϋποθέτει συνεργασία πρακτόρων αποτελεί καινοτομία στον τομέα της ενισχυτικής μάθησης. Τα αποτελέσματα που προκύπτουν δεν είναι πάντα περισσότερο αποδοτικά από λειτουργίες όπου κάθε πράκτορας δρα μόνος του ως μεμονωμένη οντότητα. Όμως, η διαρκής μελέτη και έρευνα του τομέα της ενισχυτικής μάθησης με συνεργασία πολλών πρακτόρων αναμένεται να επιφέρει επαναστατικά αποτελέσματα στις εφαρμογές της ενισχυτικής μάθησης.

3.3 Ετερογενή τηλεπικοινωνιακά επίγεια συστήματα πολυεκπομής (Heterogeneous multicast terrestrial communication systems)

3.3.1 Ετερογενή τηλεπικοινωνιακά επίγεια συστήματα πολυεκπομής (Heterogeneous multicast terrestrial communication systems)

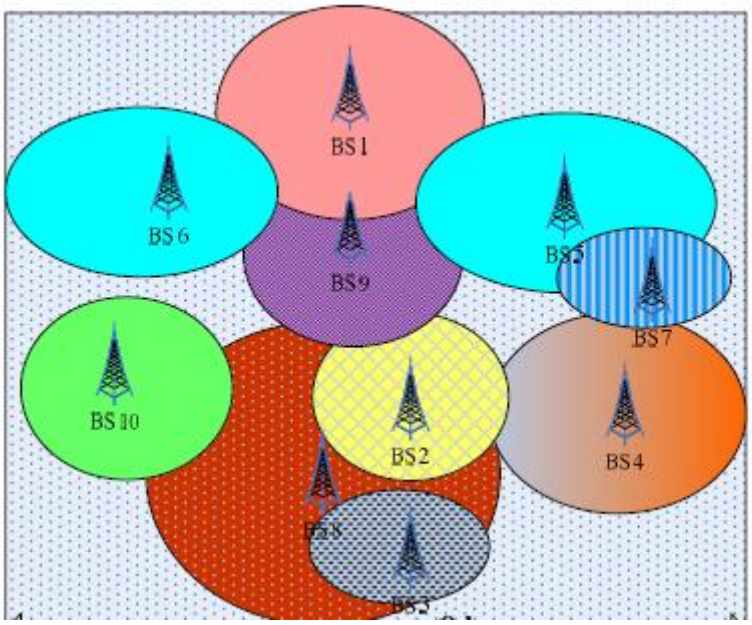
Το παρόν τηλεπικοινωνιακό σύστημα δείχνει ότι η εκχώρηση καναλιών (ανάθεση ζωνών συχνοτήτων σε σταθμούς βάσης) σε ετερογενή επίγεια τηλεπικοινωνιακά συστήματα πολυεκπομπής μπορεί να βελτιωθεί με τη χρήση της νοημοσύνης που προσφέρει η ενισχυτική μάθηση. Παρουσιάζονται δύο σχέδια ανάθεσης καναλιών (channel assignment), ένα με χρήση προτεραιότητας μεταξύ των καναλιών και ένα τυχαίας προσπέλασής τους, στα οποία η χρήση της ενισχυτικής μάθησης βελτιώνει την ταχύτητα και την ποιότητα εκχώρησης των καναλιών στους επίγειους σταθμούς βάσης. Αυτό επιτυγχάνεται με μείωση των επανεκχωρήσεων (αλλαγή επιλογής καναλιού για το σταθμό βάσης) και των ποσοστών φραγής (blocking) και διακοπής (dropping) της επικοινωνίας του τηλεπικοινωνιακού συστήματος. Ένας συντελεστής βαρύτητας (weighting factor) χρησιμοποιείται για να προσδιορίσει την υψηλότερη προτεραιότητα μεταξύ των καναλιών και να βοηθήσει στη βελτιστοποίηση της επίδοσης του συστήματος εκχωρήσεων των καναλιών στους επίγειους σταθμούς βάσης. Τα σχήματα με κανάλια ανάθεσης είναι από τα σημαντικότερα έργα για έλεγχο της αποδοτικότητας της multicasting χρησιμοποίησης του φάσματος. Γενικά, οι εκχωρήσεις καναλιών διακρίνονται στις τρείς ακόλουθες κατηγορίες, 1) καθορισμένη εκχώρηση καναλιού (Fixed Channel Assignment, FCA), 2) δυναμική εκχώρηση καναλιού (Dynamic Channel Assignment, DCA) και 3) μικτή εκχώρηση καναλιού (Hybrid Channel Assignment, HCA).

Στη συγκεκριμένη περίπτωση μελετάται μια περιοχή καναλιών φάσματος βασισμένη σε δυναμική DCA μέθοδο βελτιστοποίησης ώστε να προσδιοριστούν τα καταλληλότερα κανάλια με βάση τα στατιστικά στοιχεία των SINR (λόγος σήματος προς παρεμβολή) που περιέχονται στους χρήστες της περιοχή κάλυψης. Σκοπός είναι να γίνει η επιλογή του καλύτερου δυνατού καναλιού από πλευράς παροχής ποιότητας υπηρεσίας για τους σταθμούς βάσης που βρίσκονται στην περιοχή κάλυψης που μελετάται. Τα σχήματα που χρησιμοποιούνται επιλέγουν ένα κανάλι βασισμένα σε ένα κατώφλιο του λόγου SINR εντός της περιοχής κάλυψης.

3.3.2 Το μοντέλο του σχήματος επιλογής καναλιών από τους επίγειους σταθμούς βάσης

Αντίθετα στη λογική άλλων σχετικών επίγειων μοντέλων ανάθεσης καναλιών τα οποία ασχολούνται με μεμονωμένους χρήστες, το σχήμα ενισχυτικής μάθησης που εξετάζεται δίνει έμφαση στην ταυτόχρονη διανομή καναλιών επικοινωνίας σε πολλούς χρήστες εντός της περιοχής κάλυψης. Ολόκληρη η περιοχή θα θεωρηθεί μια ενιαία οντότητα. Για να απλοποιηθεί το σενάριο πολυεκπομπής αρχικά

θεωρούνται ένα σύνολο καναλιών και ένα σύνολο σταθμών βάσης που κατανέμονται σε τυχαίες περιοχές όπως φαίνεται και στο Σχ. 3.6. Για την προσομοίωση γίνεται η θεώρηση ότι υπάρχουν πέντε υποψήφια κανάλια για εκχώρηση και τριάντα σταθμοί βάσης που εξυπηρετούν τις ανάγκες τηλεπικοινωνιακής κάλυψης των χρηστών της περιοχής υπηρεσιών.



Σχήμα 3.6 Κατανομή σταθμών βάσης και καναλιών

Για να προσδιοριστεί η επίδοση του συστήματος είναι απαραίτητος ο υπολογισμός του λόγου ισχύος λήψης προς το άθροισμα των ισχύων παρεμβολής και θορύβου αντίστοιχα μέσω της σχέσης:

$$SINR = \frac{P_s}{P_n + \sum P_i} \tag{3.2}$$

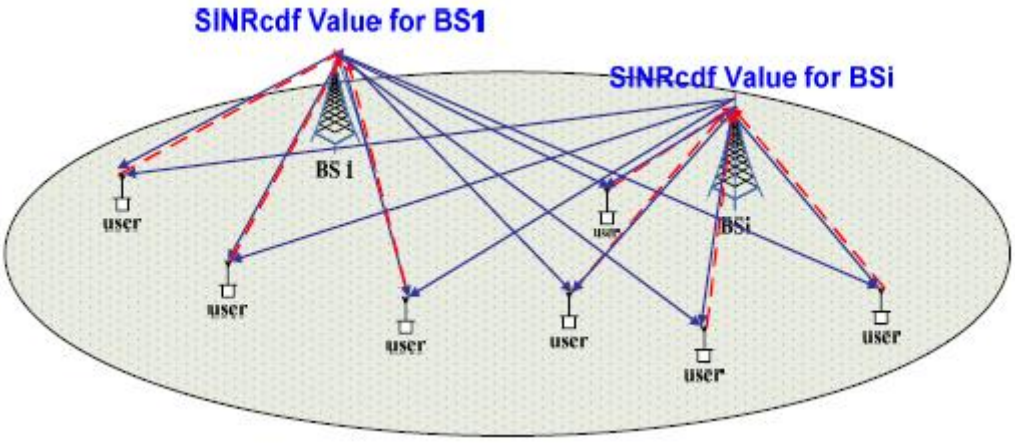
όπου P_s είναι η επιθυμητή λαμβανόμενη ισχύς σήματος ενός χρήστη, P_n είναι η ισχύς θορύβου και $\sum P_i$ η ισχύς όλων των παρεμβαλλόντων σταθμών στο ίδιο κανάλι. Το μοντέλο διάδοσης που χρησιμοποιείται είναι αυτό των Okumura-Hata.

Ο σκοπός του σχεδίου που μελετάται είναι η καλύτερη δυνατή χρησιμοποίηση των καναλιών από τους σταθμούς βάσης του συστήματος ώστε να επιτευχθεί ταυτόχρονη κάλυψη της περιοχής υπηρεσιών των σταθμών βάσης παρέχοντας στους χρήστες την καλύτερη δυνατή ποιότητα υπηρεσίας. Στόχος είναι κάθε σταθμός βάσης να καλύπτει το μικρότερο δυνατό ποσοστό της συνολικής περιοχής κάλυψης με το υψηλότερο δυνατό SINR ώστε να εξασφαλίζεται η καλύτερη δυνατή ποιότητα της multicast κίνησης. Βεβαίως πρέπει να τονιστεί ότι οι σταθμοί βάσης οι

οποίοι χρησιμοποιούν διαφορετικά κανάλια, δεν επηρεάζουν την τιμή του SINR σε άλλους σταθμούς βάσης. Για να επιτευχθεί μια υψηλή τιμή SINR χρησιμοποιείται ένα κατώφλιο ποσόστοσης ως μέθοδος για τον υπολογισμό του ανεκτού επιπέδου παρεμβολής μεταξύ χρηστών. Το κατώφλιο υπολογίζεται από τη σχέση:

$$\kappa\alpha\tau\omega\varphi\lambda\iota o\_\pi o\sigma\acute{o}\sigma\tau o\sigma\eta\varsigma = 1 - \frac{1}{N} \tag{3.3}$$

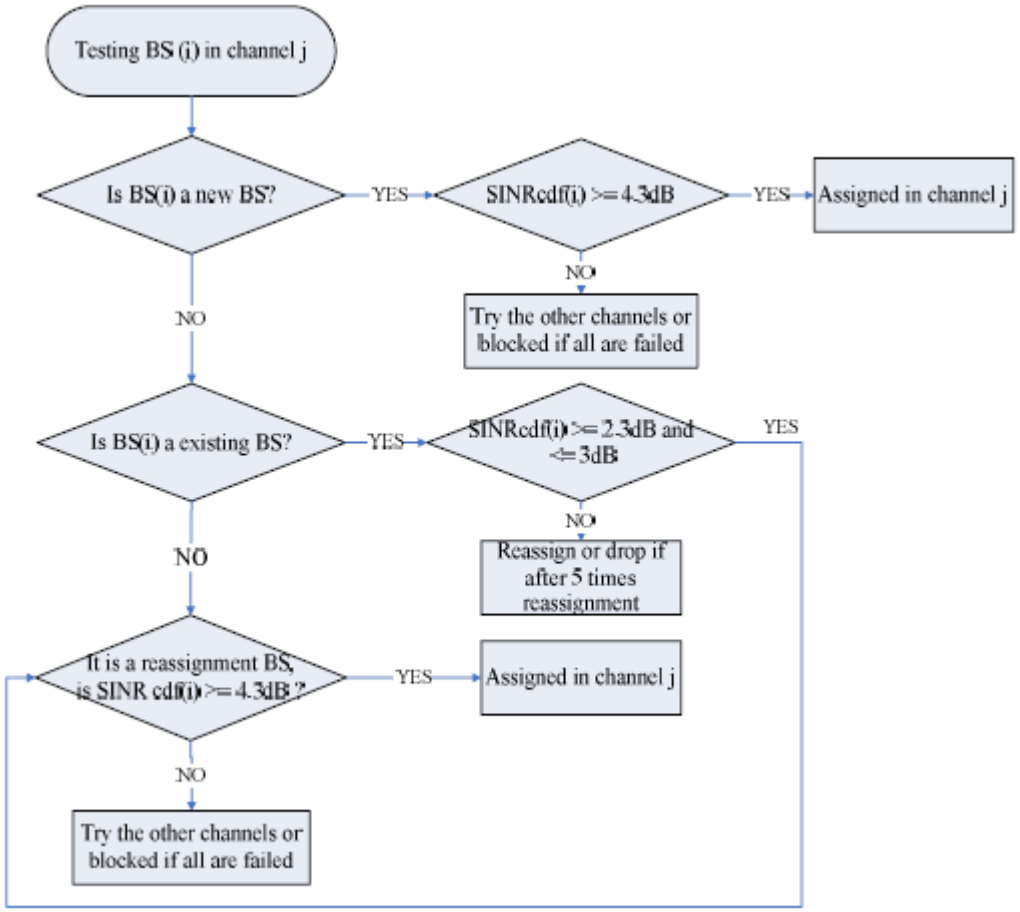
όπου N είναι ο συνολικός αριθμός των σταθμών βάσης του επίγειου τηλεπικοινωνιακού συστήματος. Για να γίνει κατανοητός ο τρόπος υπολογισμού του κατωφλίου ποσόστοσης αναφέρεται το εξής παράδειγμα: εάν έπρεπε να καλυφθεί μια περιοχή υπηρεσιών από δέκα σταθμούς βάσης θα ήταν επιθυμητό ο καθένας να καλύπτει το 10% της συνολικής περιοχής. Όπως φαίνεται και στο Σχ.3.6 εάν η τιμή του $SINR_{cdf}$ είναι μεγαλύτερη από το κατώφλιο, εντός της περιοχής ποσόστοσης του σταθμού βάσης, τότε το κανάλι είναι διαθέσιμο για εκχώρηση στο σταθμό βάσης που ανήκει το κατώφλιο αυτό. Ένα παράδειγμα του σεναρίου πολυεκπομπής παρουσιάζεται στο Σχ.3.7 που ακολουθεί.



Σχήμα 3.7: Το multicast σενάριο

Εξετάζονται δύο σχήματα κατανομής των καναλιών στους σταθμούς βάσης: το πρώτο είναι το σχήμα κατανομής καναλιών με προτεραιότητα (channel priority distributed scheme) και το δεύτερο είναι το σχήμα τυχαίας κατανομής καναλιών (random picking distributed scheme). Η διαφορά των δύο σχημάτων έγκειται στον τρόπο επιλογής των καναλιών από τους σταθμούς βάσης. Στο σχήμα κατανομής καναλιών με προτεραιότητα τα κανάλια επιλέγονται με βάση τις υψηλότερες τιμές στα βάρη των καναλιών (υψηλότερη τιμή βάρους συνεπάγεται μεγαλύτερη προτεραιότητα για το κανάλι). Στο σχήμα τυχαίας κατανομής ο τρόπος επιλογής των καναλιών από τους σταθμούς βάσης είναι τυχαίος, αφού όμως πρώτα έχει γίνει η επιλογή των τριών καναλιών με το μεγαλύτερο παράγοντα βάρους. Δηλαδή από τα τρία κανάλια με τη μεγαλύτερη προτεραιότητα γίνεται τυχαία επιλογή εκχώρησης

στο σταθμό βάσης που μελετάται. Η διαδικασία ανάθεσης ενός καναλιού σε ένα σταθμό βάσης περιγράφεται από το ακόλουθο λειτουργικό διάγραμμα.



Σχήμα 3.8: Διάγραμμα ροής για την εκώρηση ενός καναλιού

Είναι φανερό ότι η εκχώρηση ενός καναλιού σε ένα σταθμό βάσης εντός της περιοχής κάλυψης γίνεται με βάση τα εξής κριτήρια:

- Αν είναι νέος σταθμός βάσης ή ένας από τους ήδη υπάρχοντες
- Με βάση τον έλεγχο των κατωφλίων (Threscolds, T) του SINR και την αντιστοίχιση τους σε βάρη προτεραιότητας των καναλιών που προκύπτουν από τον ακόλουθο πίνακα

Type of BSs		Thresholds	Weights (W_f)
New	Acceptance	T > 4.3dB	+2
	No acceptance	T <= 4.3dB	0
Existing	Reassignment	2.3dB=<T<= 3dB	-1
	Dropping	T < 2.3dB	-2
Reassignment	New acceptance	T > 4.3dB	+1

Σχήμα 3.9: Διάγραμμα κατωφλίων

Οι τιμές των κατωφλίων για το SINR προκύπτουν ως εξής. Όταν απαιτείται εκχώρηση πέντε καναλιών σε ένα σύστημα τριάντα σταθμών βάσης (όπως έχει οριστεί για τις ανάγκες της προσομοίωσης) χρησιμοποιώντας το σχήμα κατανομής καναλιών με προτεραιότητα, ένα τυπικό $SINR_{cdf}$ αποδεκτό κατώφλιο ώστε να έχει καλή επίδοση το τηλεπικοινωνιακό σύστημα είναι στα 4.3dB [53]. Επίσης το κατώφλιο των 2.3 dB, χρησιμοποιείται ως κατώφλιο της διαμόρφωσης GMSK ώστε κατά την αποδιαμόρφωση να επιτευχθεί ένα αποδεκτό ποσοστό εσφαλμένων ψηφίων (bit error rate).

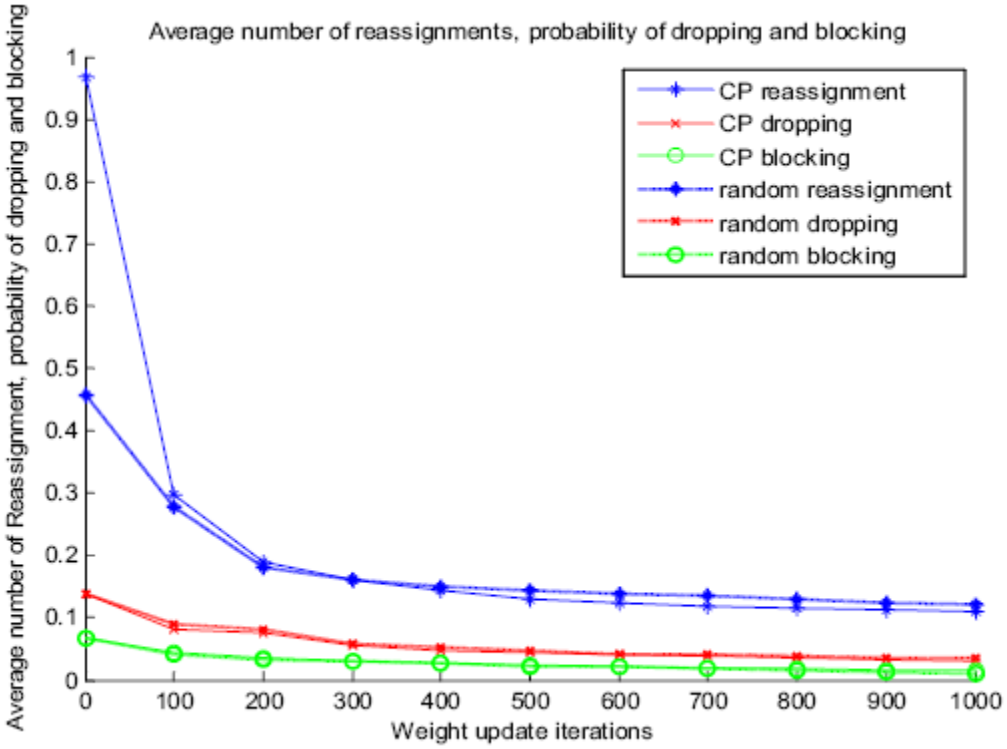
Ο αλγόριθμος υπολογισμού των συντελεστών βαρύτητας για κάθε νέα προσπάθεια εκχώρησης ενός καναλιού σε ένα σταθμό βάσης βασίζεται στη σχέση

$$W_i = F_1 * W_{i-1} + F_2 * W_e + W_f \tag{3.4}$$

όπου W_i είναι το βάρος που έχει ένα κανάλι στην τρέχουσα επανάληψη, έπειτα από την ανανέωση της πληροφορίας από τον προηγούμενο παράγοντα βάρους W_{i-1}. Ο παράγοντας W_f προκύπτει από τον πίνακα του Σχ.3.9 από τις εκχωρήσεις των καναλιών. Δηλαδή είναι ο παράγοντας βάρους που προσδιορίζει ουσιαστικά πώς μεταβάλλεται η προτεραιότητα των καναλιών. Τέλος, ο παράγοντας W_e αντιπροσωπεύει το περιβάλλον, αλλά στο συγκεκριμένο σχήμα έχει την τιμή 0. Έτσι, μετά το πέρας κάθε επανάληψης, κάθε κανάλι χαρακτηρίζεται από ένα ανανεωμένο παράγοντα βάρους ο οποίος αντιστοιχεί στη νέα προτεραιότητα του καναλιού. Σε κάθε δράση οι σταθμοί βάσης επιλέγουν το κανάλι με τη μεγαλύτερη προτεραιότητα (μεγαλύτερο παράγοντα βάρους) για να επιχειρήσουν εκχώρηση. Έτσι, η ενισχυτική μάθηση επιτρέπει στα κανάλια να βρούν τον ταχύτερο τρόπο εκχώρησης σε κάποιο σταθμό βάσης του τηλεπικοινωνιακού συστήματος ώστε να μειωθούν η παρεμβολή και οι συγκρούσεις μεταξύ σταθμών βάσης.

Για να γίνει σαφής η πρόοδος που επιφέρει η ενισχυτική μάθηση στην επίδοση του συστήματος των σταθμών βάσης, παρουσιάζεται στο Σχ.3.10 το διάγραμμα του

μέσου πλήθους επανεκχωρήσεων καθώς και των πιθανοτήτων απόρριψης και φραγής ως συνάρτηση των επαναλήψεων εκτέλεσης.



Σχήμα 3.10: Πρόοδος ενισχυτικής μάθησης

Είναι φανερό ότι με την αύξηση του αριθμού των επαναλήψεων το σύστημα επιτυγχάνει βελτιωμένα αποτελέσματα, δηλαδή καλύτερη εξυπηρέτηση των καναλιών από τους σταθμούς βάσης. Η συμπεριφορά των επανεκχωρήσεων και για τα δύο σχήματα της προσομοίωσης μπορεί να χωριστεί σε τρείς περιόδους. Η πρώτη περίοδος, η οποία περιλαμβάνει τις 100 πρώτες επαναλήψεις, είναι η περίοδος έρευνας (investigation period). Σε αυτή την περίοδο ολόκληρο το σύστημα σε συνδυασμό με την ενισχυτική μάθηση είναι σε μια δυναμική διεργασία. Το σύστημα ξεκινά να μαθαίνει ποιά είναι η αποδοτικότερη συμπεριφορά που μπορεί να έχει, αλλά δεν διαθέτει ακόμα προηγούμενη εμπειρία η οποία θα επηρεάσει το αποτέλεσμα της μάθησης. Η δεύτερη περίοδος περιλαμβάνει τις επαναλήψεις 100-300 είναι η περίοδος συσσώρευσης (accumulation period), όπου ο αριθμός των επανεκχωρήσεων μειώνεται από 10-40%. Πρόκειται για την περίοδο κατα την οποία τα αποτελέσματα της ενισχυτικής μάθησης είναι φανερά όσον αφορά τη βελτίωση της εκτέλεσης του συστήματος. Μπορεί να χαρακτηριστεί ως περίοδος συνύπαρξης και αλληλεπίδρασης καθώς η προηγούμενη εμπειρία βοηθά αποτελεσματικά τη συμπεριφορά των σταθμών βάσης ως προς την εκχώρηση των καναλιών. Ως φυσικό επακόλουθο παρατηρείται μια διαρκής μείωση των επανεκχωρήσεων. Τέλος, η τρίτη περίοδος η οποία ξεκινά από την επανάληψη 300 και διαρκεί μέχρι το τέλος της προσομοίωσης είναι η περίοδος ωρίμανσης (mature period). Σε αυτή την

περίοδο οι σταθμοί βάσης έχουν σταθεροποιήσει τα κανάλια που τους επιφέρουν τα μικρότερα επίπεδα παρεμβολής. Βέβαια υπάρχει ακόμα μια μικρή τάση μείωσης του πλήθους των επανεκχωρήσεων διότι η ενισχυτική μάθηση εξακολουθεί να βελτιώνει την εκτέλεση του συστήματος.

3.3.3 Συμπεράσματα

Το επίγειο τηλεπικοινωνιακό σύστημα που μελετάται παρουσιάζει δύο σχέδια εκχώρησης (κατανομής) καναλιών σε ένα σύστημα γνωστικών ραδιοεπικοινωνιών χρησιμοποιώντας ενισχυτική μάθηση και έναν παράγοντα βάρους (που χαρακτηρίζει την προτεραιότητα κάθε καναλιού). Προκύπτει το αποτέλεσμα ότι τηλεπικοινωνιακά συστήματα που χρησιμοποιούν ενισχυτική μάθηση μπορούν να βελτιώσουν με αποδοτικό τρόπο την ταχύτητα εκχώρησης των καναλιών στους σταθμούς βάσης ενός συστήματος, περιορίζοντας σε σημαντικό βαθμό τα επίπεδα παρεμβολής και συγκρούσεων κατα τη διάρκεια ανάθεσης ενός καναλιού σε κάποιο σταθμό βάσης. Οι βελτιώσεις της επίδοσης του συστήματος είναι αποτέλεσμα της προτεραιότητας την οποία αποκτούν τα κανάλια ως προς την εκχώρησή τους σε ένα σταθμό βάσης η οποία επιτυγχάνεται μέσα από τη μάθηση των παρελθοντικών επιτυχημένων και αποτυχημένων εκχωρήσεων, καθώς και με την αύξηση του αποδεκτού κατωφλίου ή του αριθμού των καναλιών, που τα καθιστούν λιγότερο θορυβώδη.

3.4 Σύνοψη

Στο κεφάλαιο αυτό παρουσιάστηκαν τρία χαρακτηριστικά σχέδια εφαρμογών της ενισχυτικής μάθησης. Ο σκοπός που επιλέχθηκαν τα συγκεκριμένα σχέδια ήταν ότι η προσομοίωση του σεναρίου που ακολουθεί στο επόμενο κεφάλαιο έχει υιοθετήσει ορισμένα βασικά χαρακτηριστικά απο τα τρία αυτά σχέδια. Η κατανεμημένη εκχώρηση φάσματος και η χρήση συνάρτησης ανταμοιβής με βάρη είναι τα βασικά συστατικά που χρησιμοποιούνται από το πρώτο σχήμα. Ο πιθανοτικός τρόπος επιλογής καναλιών καθώς και η επέκταση της ενισχυτικής μάθησης με συνεργασία πολλών πρακτόρων είναι τα χαρακτηριστικά που αποκομίζονται από το δεύτερο σχήμα. Τα χαρακτηριστικά της ποιότητας υπηρεσίας σε κάθε σταθμό βάσης κατά την εκχώρηση καναλιών είναι στοιχεία που παρέχει το τρίτο σχήμα. Με βάση αυτά τα χαρακτηριστικά προτείνεται ένα σχήμα ενισχυτικής μάθησης και εξάγονται βασικά συμπεράσματα από τον τρόπο που λειτουργεί και τα αποτελέσματα που προσφέρει. Το μοντέλο που μελετάται ακολουθεί στο επόμενο κεφάλαιο.

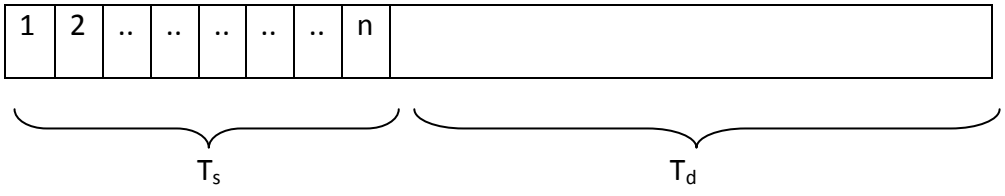
ΚΕΦΑΛΑΙΟ 4^{ο}: ΠΡΟΣΟΜΟΙΩΣΗ

4.1 Μοντέλο χρήσης του υπάρχοντος φάσματος

Θεωρείται ότι το διαθέσιμο αδειοδοτημένο φάσμα αποτελείται από N κανάλια καθένα από τα οποία χρησιμοποιείται διαιρώντας το χρόνο σε χρονοσχισμές (time slots). Οι πρωτεύοντες χρήστες καταλαμβάνουν το φάσμα υπό συγκεκριμένη πιθανότητα κατάληψης για το χρονικό διάστημα κατά το οποίο εξετάζεται η πρόσβαση σε αυτό. Αυτή η πιθανότητα κατάληψης των καναλιών αντιπροσωπεύει τη χρησιμοποίηση του διαθέσιμου φάσματος. Ουσιαστικά, αποτελεί την εικόνα προς το διαχειριστή του φάσματος ώστε να επιτευχθεί η καλύτερη δυνατή αξιοποίησή του. Οι δευτερεύοντες χρήστες προσπαθούν να μάθουν τη συμπεριφορά των πρωτευόντων χρηστών και να επιλέγουν σωστά κάποιο κανάλι ώστε να πραγματοποιούν μετάδοση σε κάθε χρονική στιγμή. Ο τρόπος με τον οποίο πραγματοποιείται αυτή η αλληλεπίδραση των δευτερευόντων χρηστών με τους πρωτεύοντες χρήστες και το διαθέσιμο αδειοδοτημένο φάσμα περιγράφεται στη συνέχεια.

4.2 Ανίχνευση και πρόσβαση στο φάσμα

Θεωρείται ένας μόνο δευτερεύων χρήστης, ο οποίος θα επιχειρήσει μετάδοση σε κάποιο κανάλι του διαθέσιμου φάσματος. Κάθε χρονοσχισμή χωρίζεται σε δύο φάσεις. Η πρώτη φάση είναι η φάση κατα την οποία ο δευτερεύων χρήστης πραγματοποιεί ανίχνευση σε n από τα N κανάλια με σκοπό να εντοπίσει ένα διαθέσιμο κανάλι, δηλαδή ένα κανάλι το οποίο δεν χρησιμοποιεί κάποιος πρωτεύων αδειοδοτημένος χρήστης τη συγκεκριμένη χρονοσχισμή. Αυτή ονομάζεται φάση ανίχνευσης (sensing). Η δεύτερη φάση είναι η φάση όπου πραγματοποιείται η μετάδοση στο κανάλι το οποίο επιλέχθηκε κατα την πρώτη φάση. Η φάση αυτή ονομάζεται φάση μετάδοσης (transmitting).



time slot

T_s: χρόνος φάσης ανίχνευσης

Ο χρόνος ανίχνευσης λαμβάνεται ίσος προς 2.5msec ώστε να επιτευχθεί το επιθυμητό αποτέλεσμα [54].

T_d: χρόνος φάσης μετάδοσης

n: ο αριθμός των καναλιών που ανιχνεύονται

1) Φάση ανίχνευσης

Σε κάθε χρονοσχισμή όλα τα κανάλια χαρακτηρίζονται από συγκεκριμένη πιθανότητα κατάληψής τους από κάποιον πρωτεύοντα χρήστη. Ο δευτερεύων χρήστης προσπαθεί σε κάθε χρονοσχισμή να αποκτήσει πρόσβαση σε κάποιο από τα διαθέσιμα κανάλια του φάσματος, το οποίο δεν χρησιμοποιείται από κάποιον πρωτεύοντα χρήστη. Η επιλογή του καναλιού από την ανίχνευση την οποία πραγματοποιεί ο δευτερεύων χρήστης γίνεται με πιθανοτικό (probabilistic) τρόπο, δηλαδή με βάση τις πιθανότητες κατάληψης που έχουν αποδοθεί στους πρωτεύοντες χρήστες. Με βάση τον πιθανοτικό τρόπο ανίχνευσης όσο μικρότερη είναι η πιθανότητα ένα κανάλι να είναι κατειλημμένο από κάποιο πρωτεύοντα χρήστη, τόσο πιθανότερη είναι η πρόσβαση σε αυτό το κανάλι από το δευτερεύοντα χρήστη. Κατα τη διάρκεια της φάσης αυτής ο δευτερεύων χρήστης ανανεώνει την αντίληψη την οποία έχει για το περιβάλλον. Ως περιβάλλον ορίζεται το διαθέσιμο αδειοδοτημένο φάσμα και οι πρωτεύοντες χρήστες οι οποίοι το καταλαμβάνουν. Η αντίληψή του βασίζεται στα ακόλουθα διανύσματα.

- value(f): η τιμή που έχει ο δευτερεύων χρήστης για το κανάλι f ή άλλως για τη φέρουσα συχνότητα που αντιστοιχεί στο κανάλι.
- test(f): πόσες φορές έχει ανιχνεύσει ή έχει προσπαθήσει να εκπέμψει στο κανάλι f
- success(f): πόσες φορές έχει επιτυχημένη μετάδοση στο κανάλι f
- last_check(f): σε πόσες χρονοσχισμές πριν από την τρέχουσα χρονοσχισμή κάθε φορά πραγματοποιήθηκε ανίχνευση του καναλιού ή προσπάθεια μετάδοσης

Τα ανωτέρω διανύσματα σχετίζονται μεταξύ τους μέσω των ακολούθων σχέσεων:

$$value(f) = W \cdot success(f) + (1 - W) \cdot value(f) \tag{3.1}$$

$$W = w + \frac{1 - w}{test(f)} \tag{3.2}$$

όπου W η βαρύτητα που αποδίδεται στη νέα πληροφορία (weight of the new information)

Η επιλογή του καναλιού όπου θα προσπαθήσει να εκπέμψει ο δευτερεύων χρήστης βασίζεται στην κανονικοποιημένη συνάρτηση

$$\text{choice(f)'} = \frac{choice(f)}{\sum choice(f)} \tag{3.3}$$

όπου

$$choice(f) = \begin{cases} \dfrac{value(f)}{\sum value(f)} \cdot \dfrac{1}{last\_check(f)} & , last\_check(f) > 0 \\ \dfrac{value(f)}{\sum value(f)} & , last\_check(f) = 0 \end{cases} \tag{3.4}$$

Είναι φανερό ότι όσο μεγαλύτερη είναι η τιμή του value(f) που διατηρεί ο δευτερεύων χρήστης για ένα κανάλι, τόσο πιθανότερη είναι η επιλογή του. Επίσης, όσο μικρότερη είναι η τιμή του last_check(f) για κάποιο κανάλι τόσο μεγαλύτερη είναι η πιθανότητα να επιχειρήσει μετάδοση σε αυτό ο δευτερεύων χρήστης. Επιπλέον, όσο μεγαλύτερη είναι η τιμή του διανύσματος test(f) τόσο περισσότερο έγκυρο είναι το αποτέλεσμα της μάθησης, διότι μεγάλη τιμή αυτού έχει ως συνέπεια μικρότερο εναποτιθέμενο βάρος W στη νέα πληροφορία η οποία προστίθεται.

2)Φάση μετάδοσης

Μετά το πέρας της φάσης ανίχνευσης ακολουθεί η φάση μετάδοσης για το δευτερεύοντα χρήστη. Διακρίνονται δύο πιθανές περιπτώσεις κατά την εκτέλεση της φάσης αυτής:

i) Εφόσον το κανάλι στο οποίο αποφασίστηκε να γίνει εκπομπή από το δευτερεύοντα χρήστη δεν είναι κατειλημμένο από πρωτεύοντα χρήστη, πραγματοποιείται μετάδοση. Στην περίπτωση αυτή η διέλευση (throughput) είναι

$$R = \frac{T - n \cdot T_S}{T} \cdot C_0 \tag{3.5}$$

όπου T η χρονική διάρκεια της χρονοσχισμής, n ο αριθμός του καναλιού στο οποίο πραγματοποιείται μετάδοση και

$$C_0 = \log_2(1 + SNR_s) \tag{3.6}$$

αντιπροσωπεύει τη διέλευση, που επιτυγχάνει ο δευτερεύων χρήστης όταν πραγματοποιεί μετάδοση σε κάποιο κανάλι, χωρίς την παρουσία κάποιου

πρωτεύοντα χρήστη. Ο SNR_s είναι ο σηματοθορυβικός λόγος του δευτερεύοντος χρήστη σε μία μετάδοση.

ii) Εφόσον το κανάλι στο οποίο αποφασίστηκε να πραγματοποιήσει μετάδοση ο δευτερεύων χρήστης είναι κατειλημμένο από πρωτεύοντα χρήστη, η διέλευση του δευτερεύοντος χρήστη είναι

$$R = \frac{T - n \cdot T_s}{T} \cdot C_1 \tag{3.7}$$

όπου

$$C_1 = \log_2(1 + \frac{SNR_s}{1 + SNR_p}) \tag{3.8}$$

είναι η διέλευση, που επιτυγχάνει ο δευτερεύων χρήστης όταν πραγματοποιεί μετάδοση σε κανάλι, όπου μεταδίδει πρωτεύων χρήστης. Ο SNR_p είναι ο σηματοθορυβικός λόγος του πρωτεύοντος χρήστη σε μία μετάδοση. Δηλαδή γίνεται η θεώρηση ότι ο δευτερεύων χρήστης μεταδίδει στο κανάλι που επιλέγει μετά τη φάση ανίχνευσης, ακόμα και στην περίπτωση όπου το κανάλι είναι κατειλημμένο από κάποιο πρωτεύοντα χρήστη. Συνεπώς, η εκπομπή του δευτερεύοντος χρήστη γίνεται πάντα υπό ανεκτή (non harmful) παρεμβολή προς τον πρωτεύοντα χρήστη. Βέβαια, σε αυτό το σενάριο μετάδοσης, η διέλευση έχει πολύ χαμηλότερη τιμή σε σχέση με το σενάριο όπου απουσιάζει ο πρωτεύων χρήστης.

Είναι φανερό το ότι η μετάδοση στο επιλεγμένο κανάλι απουσία πρωτεύοντος χρήστη εξασφαλίζει μια σαφώς καλύτερη διέλευση και, συνεπώς, την καλύτερη δυνατή χρησιμοποίηση του καναλιού από το δευτερεύοντα χρήστη. Ο στόχος της διαδικασίας μάθησης είναι να επιτευχθεί η υψηλότερη δυνατή διέλευση για το δευτερεύοντα χρήστη με την χαμηλότερη δυνατή απώλεια χρόνου και κατανάλωση ενέργειας στο στάδιο της ανίχνευσης. Η επιδίωξη αυτή αποτελεί και το βασικό trade-off της προσομοίωσης. Δηλαδή επίτευξη της καλύτερης δυνατής διέλευσης από το δευτερεύοντα χρήστη συνδυασμένη με όσο το δυνατό μικρότερη αλλά με ικανοποιητικά αποτελέσματα ανίχνευση.

4.3 Μελέτη συμπεριφοράς σε διαφορετικού τύπου μηνύματα

Το κύριο ερώτημα που τίθεται αφορά τη συμπεριφορά του δευτερεύοντος χρήστη ώστε να εξυπηρετήσει διαφορετικής φύσης μηνύματα. Διακρίνονται τέσσερις κατηγορίες μηνυμάτων που διαχωρίζονται κατά δύο τρόπους.

1) *Μηνύματα τύπου επείγουσας ανάγκης (emergency)*

Είναι μηνύματα που πρέπει επειγόντως να φθάσουν στον προορισμό τους. Για το λόγο αυτό, πρέπει να είναι άμεση η πρόσβασή τους στο διαθέσιμο αδειοδοτημένο φάσμα. Η εξυπηρέτηση τέτοιων μηνυμάτων γίνεται με βάση την ανάγκη εύρεσης ενός καναλιού το οποίο θα εξασφαλίζει υψηλή πιθανότητα ως προς την επιτυχημένη πρόσβαση στο αδειοδοτημένο φάσμα. Από την άλλη πλευρά, λόγω του επείγοντος χαρακτήρα του μηνύματος, δεν υπάρχει πολύς διαθέσιμος χρόνος για ανίχνευση πολλών από τα N διαθέσιμα κανάλια. Συνεπώς, ο στόχος είναι να προσδιοριστεί ο κατάλληλος αριθμός n καναλιών από το διαθέσιμο προς ανίχνευση φάσμα ο οποίος θα εξασφαλίζει το συνδυασμό υψηλής πιθανότητας επιτυχημένης πρόσβασης στο φάσμα και μικρό χρονικό διάστημα διαδικασίας ανίχνευσης σε αυτό. Η απαίτηση αυτή αποτελεί το trade-off για τα μηνύματα αυτού του τύπου.

2) *Τα μηνύματα τύπου μη επείγουσας ανάγκης (non emergency)*

Είναι μηνύματα στα οποία υπάρχει ανοχή ως προς την καθυστέρηση της πρόσβασης στο διαθέσιμο φάσμα. Δηλαδή ο δευτερεύων χρήστης έχει περιθώριο για καλύτερη ανίχνευση του διαθέσιμου φάσματος ώστε να επιτύχει την υψηλότερη δυνατή πιθανότητα πρόσβασης στο διαθέσιμο φάσμα. Συνεπώς, ο στόχος σε αυτή την περίπτωση είναι να προσδιορισθεί ο κατάλληλος αριθμός n καναλιών ανίχνευσης που θα εξασφαλίζει την υψηλότερη δυνατή πιθανότητα επιτυχημένης πρόσβασης στο διαθέσιμο φάσμα.

3) *Μηνύματα τα οποία έχουν μεγάλη διάρκεια*

Τα μηνύματα αυτού του είδους απαιτούν τη χρήση του καναλιού που θα επιλέξουν για να μεταδώσουν, για μεγάλο χρονικό διάστημα. Δηλαδή τα κανάλια που θα εξυπηρετήσουν τέτοια μηνύματα δεν πρέπει να εμφανίζουν απότομες μεταβολές στις τιμές της πιθανότητας κατάληψης τους από τους πρωτεύοντες χρήστες κατα τη διάρκεια της εκτέλεσης της προσομοίωσης. Συνεπώς, ο στόχος είναι να προσδιοριστεί ο κατάλληλος αριθμός n καναλιών ανίχνευσης που θα εξασφαλίζει την επιλογή καναλιών με σταθερή αλλά και τη μικρότερη δυνατή τιμή πιθανότητας κατάληψης από τους πρωτεύοντες χρήστες για όλη τη διάρκεια της προσομοίωσης.

4) *Μηνύματα που δεν έχουν μεγάλη διάρκεια*

Πρόκειται για μηνύματα που δίνουν το μεγαλύτερο βαθμό ελευθερίας για επιλογή καναλιού στο δευτερεύοντα χρήστη. Δεν υπάρχει απαίτηση για επιλογή καναλιού με μικρές μεταβολές στις τιμές πιθανότητας κατάληψης από τους πρωτεύοντες χρήστες. Ασφαλώς, ο δευτερεύων χρήστης πρέπει να επιλέγει το

κανάλι για μετάδοση τέτοιων μηνυμάτων, χωρίς να σπαταλά χρόνο στην ανίχνευση.

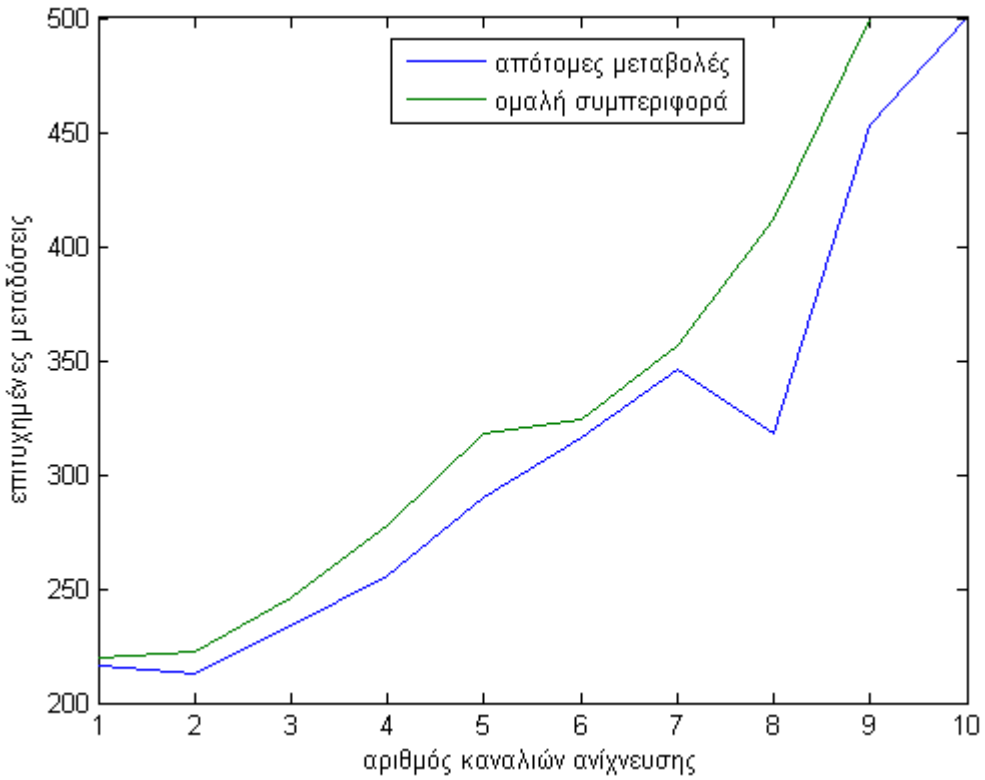
Συνδυάζοντας τους τέσσερις ανωτέρω τύπους μηνυμάτων σχηματίζονται τεσσερα διαφορετικά γεγονότα προς επεξεργασία

4.4 Αποτελέσματα προσομοίωσης

4.4.1 Σύγκριση σταθερής κατανομής-απότομων μεταβολών

Προσομοιώνοντας το μοντέλο ανίχνευσης φασματικών πόρων με τις παραμέτρους που αναφέρθηκαν προηγουμένως εξάγονται τα ακόλουθα συμπεράσματα:

Εκτελώντας το σενάριο για 500 χρονοσχισμές (slots=500) συγκρίνεται το πλήθος των επιτυχημένων μεταδόσεων του δευτερεύοντος χρήστη ως συνάρτηση του αριθμού των καναλιών στα οποία πραγματοποιεί ανίχνευση, για δύο διαφορετικές κατανομές πιθανοτήτων κατάληψης του διαθέσιμου αδειοδοτημένου φάσματος από τους πρωτεύοντες χρήστες. Κατά την πρώτη κατανομή, οι πρωτεύοντες χρήστες έχουν σταθερές τιμές πιθανοτήτων κατάληψης των διαθέσιμων καναλιών που αντιπροσωπεύουν το αδειοδοτημένο φάσμα. Κατά τη δεύτερη κατανομή, οι πιθανότητες κατάληψης των πρωτευόντων χρηστών παρουσιάζουν απότομες μεταβολές κατά τη διάρκεια της προσομοίωσης. Ως επιτυχημένη μετάδοση χαρακτηρίζεται αυτή στην οποία ο δευτερεύων χρήστης έχει εκτελέσει με επιτυχία και τις δύο φάσεις (ανίχνευση και πρόσβαση στο φάσμα).



Σχήμα 4.1: Σύγκριση επιτυχημένων μεταδόσεων των δύο σεναρίων

Όπως προκύπτει από τις γραφικές παραστάσεις και για τα δύο σενάρια, όσο αυξάνεται ο αριθμός των n καναλιών στα οποία ο δευτερεύων χρήστης πραγματοποιεί ανίχνευση τόσο αυξάνεται και ο αριθμός των επιτυχημένων μεταδόσεων. Φαίνεται ότι όταν πραγματοποιείται ανίχνευση και στα 10 κανάλια του φάσματος, οι επιτυχημένες μεταδόσεις είναι 100% των συνολικών μεταδόσεων, κάτι που είναι λογικό αφού ο χρήστης έχει την πλήρη γνώση του διαθέσιμου φάσματος. Βεβαίως κάτι τέτοιο δεν είναι επιθυμητό, αφού η πλήρης ανίχνευση του φάσματος είναι μια διαδικασία που συνεπάγεται σπατάλη χρόνου και κατανάλωση ενέργειας. Στόχος του δευτερεύοντος χρήστη είναι να επιτύχει ένα αρκετά υψηλό ποσοστό επιτυχίας πρόσβασης στο φάσμα έχοντας πραγματοποιήσει τη μικρότερη σε διάρκεια δυνατή ανίχνευση. Αυτό είναι και το βασικό trade-off που προκύπτει από τη λειτουργία του συστήματος. Από τη γραφική παράσταση που αντιπροσωπεύει το σενάριο της ομαλής κατανομής πιθανοτήτων (πράσινη γραμμή), προκύπτει ότι ο δευτερεύων χρήστης αποκτά μια καλή εκτίμηση για να επιχειρήσει πρόσβαση στο φάσμα, όταν ανιχνεύει n=5 κανάλια, όπου το ποσοστό επιτυχίας του (επιτυχημένη πρόσβαση στο φάσμα) είναι 63,6%.

Στο δεύτερο σενάριο των απότομων μεταβολών στις πιθανότητες κατάληψης του φάσματος από τους πρωτεύοντες χρήστες (μπλε γραμμή), φαίνεται ότι ο δευτερεύων χρήστης αποκτά μια επαρκή γνώση για το φάσμα (ώστε να επιχειρήσει μετάδοση με επιτυχία) όταν πραγματοποιεί ανίχνευση σε n=6 κανάλια. Το ποσοστό

της επιτυχίας του είναι 63,2%. Επίσης, στο σενάριο αυτο παρατηρείται μία βύθιση της καμπύλης των επιτυχημένων μεταδόσεων στο πέρασμα από την ανίχνευση n=7 καναλιών στην ανίχνευση n=8 καναλιών. Αυτή η βύθιση είναι αποτέλεσμα της μετάβασης του δευτερεύοντος χρήστη σε κανάλια με πολύ υψηλή πιθανότητα κατάληψης από πρωτεύοντες χρήστες, εξαιτίας του απότομου χαρακτήρα των μεταβολών στις πιθανότητες κατάληψης του φάσματος σε αυτό το σενάριο.

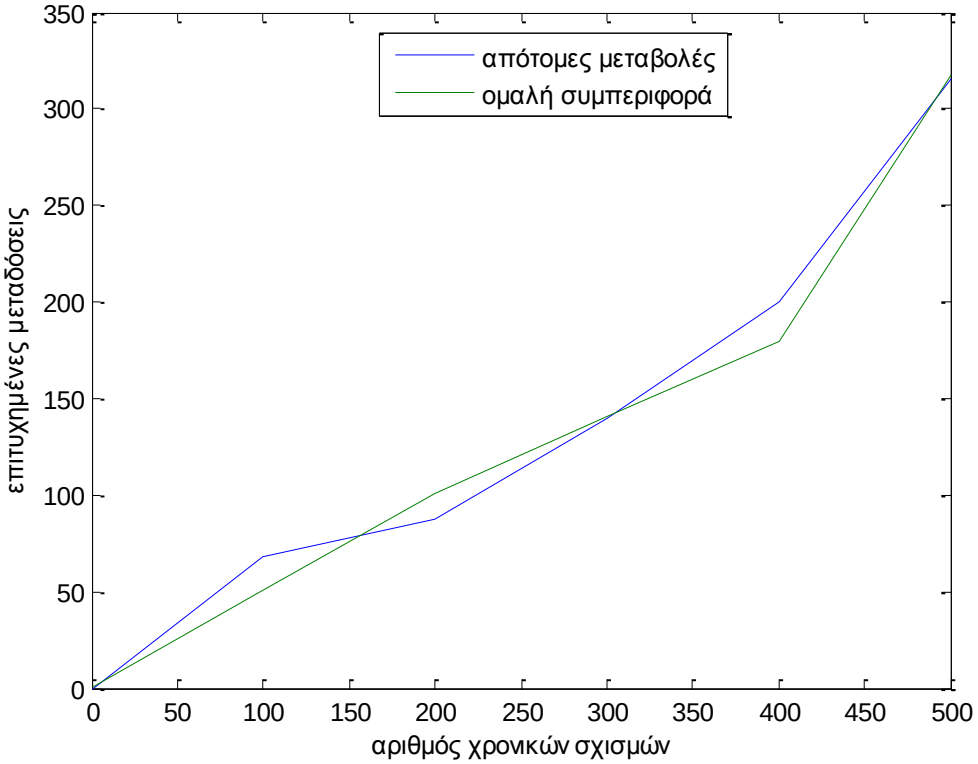
Αν υποτεθεί ότι ο δευτερεύων χρήστης καλείται να εξυπηρετήσει ένα μήνυμα τύπου επείγουσας ανάγκης (emergency), όπου απαιτείται η άμεση πρόσβασή του στο φάσμα, τότε στην περίπτωση ομαλής κατανομής των πιθανοτήτων (χωρίς απότομες μεταβολές) υπάρχει η δυνατότητα από το δευτερεύοντα χρήστη να αποκτήσει μια αρκετά αξιόπιστη εικόνα για την πρόσβασή του στο φάσμα έπειτα από ανίχνευση πέντε καναλιών. Αντίθετα, στην περίπτωση της κατανομής με απότομες μεταβολές, ο δευτερεύων χρήστης αποκτά μια αξιόπιστη εικόνα για το διαθέσιμο αδειοδοτημένο φάσμα έπειτα από ανίχνευση έξι καναλιών. Συνεπώς παρατηρείται ότι στην ομαλή κατανομή μηνύματα τύπου επείγουσας ανάγκης (emergency) εξυπηρετούνται καλύτερα.

Επίσης παρατηρείται συγκριτικό πλεονέκτημα της ομαλής κατανομής έναντι της κατανομής με απότομες μεταβολές σε μηνύματα τα οποία απαιτούν μεγάλη διάρκεια, δηλαδή επιθυμούν τη χρηση του καναλιού στο οποίο θα επιχειρήσουν πρόσβαση επί μεγάλο χρονικό διάστημα. Το πλεονέκτημα εντοπίζεται στο γεγονος ότι ένα μήνυμα τέτοιου τύπου σε ενδεχόμενη απότομη μεταβολή της χρησιμοποίησης του καναλιού από τον πρωτεύοντα χρήστη που το χρησιμοποιεί, θα προκληθεί απότομος τερματισμός της μετάδοσης του μηνύματος του δευτερεύοντος χρήστη που επιχείρησε εκεί πρόσβαση, γεγονός μη επιθυμητό για αυτού του είδους τα μηνύματα. Συνεπώς, και τα μηνύματα μεγάλης διάρκειας εμφανίζουν καλύτερη εκτέλεση στην περίπτωση ομαλής φασματικής κατανομής των πρωτευόντων χρηστών. Μάλιστα, όσο περισσότερο απότομες είναι οι μεταβολές στη φασματική κατανομή τόσο μεγαλύτερος είναι ο κίνδυνος για διακοπή της πλήρους μετάδοσης μηνυμάτων μεγάλης διάρκειας.

4.4.2 Πρόοδος ενισχυτικής μάθησης

Στο σημείο αυτό παρουσιάζεται η εξέλιξη της προόδου της ενισχυτικής μάθησης, δηλαδή πόσες τελικά επιτυχημένες μεταδόσεις πραγματοποιούνται στο διαθέσιμο αδειοδοτημένο φάσμα, από το δευτερεύοντα χρήστη στην πορεία του χρόνου της προσομοίωσης. Τα διαγράμματα που ακολουθούν απεικονίζουν γραφικά αυτή την εξέλιξη της ενισχυτικής μάθησης για τις δύο κατανομές πρωτευόντων χρηστών καθώς και για συγκεκριμένο αριθμό n καναλιών στα οποία επιχειρεί ανίχνευση ο δευτερεύων χρήστης. Στο Σχ.4.2 απεικονίζεται η εξέλιξη της ενισχυτικής μάθησης και για τα δύο σενάρια κατανομής των πιθανοτήτων των πρωτευόντων χρηστών που μελετούνται.

Η πράσινη γραμμή του Σχ.4.2 απεικονίζει την εξέλιξη της ενισχυτικής μάθησης για το σενάριο ομαλής κατανομής των πιθανοτήτων των πρωτευόντων χρηστών. Όπως αναφέρθηκε προηγουμένως στην περίπτωση της ομαλής κατανομής παρατηρείται το επιθυμητό trade-off πρόσβασης με αυξημένο ποσοστό επιτυχίας και εξοικονόμησης χρόνου και ενέργειας για n=5 κανάλια όπου ο δευτερεύων χρήστης πραγματοποιεί ανίχνευση ώστε να πραγματοποιήσει την επιθυμητή μετάδοση μηνυμάτων στο διαθέσιμο αδειοδοτημένο φάσμα. Στον κατακόρυφο άξονα μεταβάλλεται το πλήθος των επιτυχημένων μεταδόσεων και στον οριζόντιο άξονα το πλήθος των χρονικών σχισμών της προσομοίωσης.



Σχήμα 4.2: Απεικόνιση της εξέλιξης της μάθησης των δύο σεναρίων

Είναι σαφής η επιτυχία της ενισχυτικής μάθησης στην πορεία του χρόνου, δηλαδή ο δευτερεύων χρήστης αποκτά τη δυνατότητα για μετάδοση σε όλο και περισσότερα κανάλια στην πορεία του χρόνου. Επίσης φαίνεται ότι, μετά τις 400 χρονικές στιγμές η κλίση της (πράσινης) καμπύλης αυξάνεται, συνεπώς ο ρυθμός μάθησης γίνεται μεγαλύτερος. Αυτό είναι και το επιθυμητό αποτέλεσμα της λειτουργίας του συστήματος. Δηλαδή ο δευτερεύων χρήστης να αποκτά όλο και καλύτερη εικόνα για το διαθέσιμο αδειοδοτημένο φάσμα από την εμπειρία της μάθησης και ως φυσικό επακόλουθο να αυξάνονται οι επιτυχημένες μεταδόσεις του στο διαθέσιμο φάσμα.

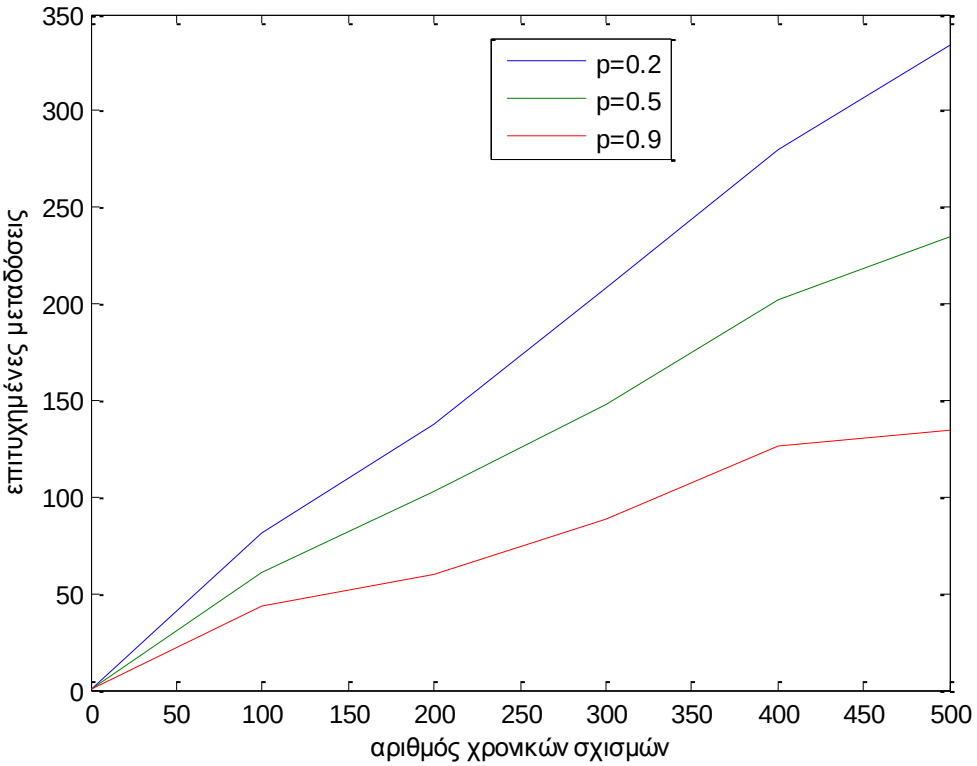
Η μπλέ γραμμή του Σχ.4.2 απεικονίζει την εξέλιξη της ενισχυτικής μάθησης για το σενάριο με τις απότομες μεταβολές των πιθανοτήτων στους πρωτεύοντες χρήστες οι οποίοι κατέχουν το διαθέσιμο αδειοδοτημένο φάσμα. Σε αυτό το σενάριο το επιθυμητό trade-off επιτυγχάνεται για n=6 κανάλια στα οποία ο δευτερεύων χρήστης πραγματοποιεί ανίχνευση ώστε να επιτύχει τελικά πρόσβαση στο διαθέσιμο φάσμα. Παρατηρείται και σε αυτό το σενάριο η επιτυχία της ενισχυτικής μάθησης στο δευτεύοντα χρήστη που επιχειρεί πρόσβαση στο διαθέσιμο φάσμα. Η κλίση της καμπύλης που αντιπροσωπεύει την εξέλιξη της ενισχυτικής μάθησης είναι και σε αυτό το σενάριο αυξανόμενη. Ο δευτερεύων χρήστης αποκτά όλο και καλύτερη εικόνα για τα φασματικά κενά τα οποία έχει τη δυνατότητα να εκμεταλλευτεί. Συγκρίνοντας τις δύο καμπύλες των δύο σεναρίων είναι φανερό ότι η μεταβολή του ρυθμού αύξησης της μάθησης, δηλαδή το πόσο ταχέως αυξάνεται η κλίση των καμπυλών ποικίλει κατά το συνολικό χρόνο εκτέλεσης των σεναρίων. Αυτό εξαρτάται από τις τιμές των πιθανοτήτων κατάληψης των πρωτευόντων χρηστών. Δηλαδή όταν οι τιμές πιθανοτήτων κατάληψης του φάσματος από τους πρωτεύόντες χρήστες είναι μικρότερες για ένα από τα δύο σενάρια για κάποιο χρονικό διάστημα, τότε η κλίση της καμπύλης που δείχνει την πρόοδο της ενισχυτικής μάθησης αυξάνεται με ταχύτερο ρυθμό για αυτό το σενάριο.

4.4.3 Συνδυασμός τυχαίων μεταβολών και σταθερής κατανομής πιθανοτήτων

Στη συνέχεια, μελετάται ένα σενάριο κατανομής πιθανοτήτων των πρωτευόντων χρηστών που αποτελεί ουσιαστικά το συνδυασμό των δύο προηγούμενων σεναρίων σε ένα. Σε αυτή την περίπτωση η πιθανοτική κατάληψη των δέκα καναλιών του διαθέσιμου φάσματος έχει την εξής μορφή:

- για τα πρώτα πέντε κανάλια του φάσματος η κατανομή των πιθανοτήτων ακολουθεί τις απότομες μεταβολές του δεύτερου σεναρίου
- για τα πέντε επόμενα κανάλια του φάσματος η κατανομή των πιθανοτήτων ακολουθεί την ομαλή κατανομή (σταθερές τιμές) του πρώτου σεναρίου. Θα μελετηθούν τρεις υποπεριπτώσεις, 1) χαμηλές τιμές πιθανοτήτων κατάληψης, p=0.2 2) υψηλές τιμές πιθανοτήτων κατάληψης, p=0.9 και 3) p=0.5.

Το Σχ.4.4 που ακολουθεί απεικονίζει την εξέλιξη της μάθησης για τις τρείς υποπεριπτώσεις του σεναρίου που εξετάζεται και για ανίχνευση σε n=6 κανάλια του διαθέσιμου φάσματος.



Σχήμα 4.4 Συνδυασμός τυχαίων μεταβολών και σταθερής κατανομής πιθανοτήτων

Είναι σαφές ότι η καλύτερη δυνατή εκτέλεση του συστήματος ενισχυτικής μάθησης προκύπτει όταν η κατανομή των καναλιών (6-10) έχει πιθανότητα κατάληψης p=0.2 από τους πρωτεύοντες χρήστες του συστήματος. Αυτό συμβαίνει διότι οι πιθανότητες κατάληψης των καναλιών του διαθέσιμου φάσματος από τους πρωτεύοντες χρήστες έχουν χαμηλή τιμή. Συνεπώς, είναι πιθανότερη η εύρεση ενός καναλιού για μετάδοση από το δευτερεύοντα χρήστη. Σε αυτή την περίπτωση όπως προκύπτει και από το διάνυσμα success(f) το οποίο υποδηλώνει πόσες φορές υπήρξε επιτυχής μετάδοση σε κάθε κανάλι του διαθέσιμου φάσματος, οι περισσότερες επιτυχημένες μεταδόσεις για το δευτερεύοντα χρήστη πραγματοποιούνται στα κανάλια (6-10) στα οποία υπάρχει χαμηλή φασματική χρησιμοποίηση. Αντίθετα στα κανάλια (1-5) όπου παρουσιάζονται απότομες μεταβολές των πιθανοτήτων κατάληψης του φάσματος από τους πρωτεύοντες χρήστες πραγματοποιούνται σαφώς λιγότερες μεταδόσεις όπως προκύπτει και από τα στοιχεία του πίνακα 4.1. Κατά την εκτέλεση ενός τέτοιου σεναρίου θα είχαν τη δυνατότητα να εξυπηρετηθούν από το δευτερεύοντα χρήστη μηνύματα τύπου μεγάλης διάρκειας διότι θα περατώνονταν από τα κανάλια (6-10) με τη χαμηλή σταθερή φασματική χρησιμοποίηση από τους πρωτεύοντες χρήστες. Επίσης, ο δευτερεύων χρήστης μπορεί να εξυπηρετήσει και μηνύματα τύπου επείγουσας ανάγκης καθώς παρατηρείται μεγάλος αριθμός επιτυχημένων μεταδόσεων και κατά

συνέπεια ένα καλό trade-off ανάμεσα στο μικρότερο δυνατό χρόνο ανίχνευσης και την επιτυχημένη μετάδοση ενός μηνύματος.

success(f)	success(1)	success(2)	success(3)	success(4)	success(5)	success(6)	success(7)	success(8)	success(9)	success(10)
p=0.2	25	30	20	33	28	51	57	52	68	60
P=0.5	24	42	38	31	30	39	29	36	33	38
P=0.9	28	28	25	39	37	9	4	4	5	4

Πίνακας 4.1

Η εκτέλεση της προσομοίωσης του συστήματος όταν η πιθανότητα κατάληψης των πρωτευόντων χρηστών στα κανάλια (6-10) είναι p=0.9 παρουσιάζει τις λιγότερες επιτυχημένες μεταδόσεις του δευτερεύοντος χρήστη στο διαθέσιμο φάσμα, γεγονός που είναι αναμενόμενο διότι στο σενάριο αυτό τόσο τα πρώτα πέντε κανάλια παρουσιάζουν απότομες μεταβολές στις πιθανότητες κατάληψης τους από τους πρωτεύοντες χρήστες αλλά και στα κανάλια (6-10) η πιθανότητα κατάληψης από τους πρωτεύοντες χρήστες είναι πολύ υψηλή. Συνεπώς, πρόκειται για το δυσμενέστερο σενάριο όπως προκύπτει και από το συνολικό αριθμό των επιτυχημένων μεταδόσεων του δευτερεύοντος χρήστη. Τα μόνα μηνύματα τα οποία θα μπορούσε με επιτυχία να μεταδώσει ο δευτερεύων χρήστης στην περίπτωση αυτή είναι τα μηνύματα που είτε είναι μη επείγουσας ανάγκης (non emergency) είτε μηνύματα που δεν έχουν μεγάλη διάρκεια, διότι αυτού του είδους μηνύματα είναι τα λιγότερο απαιτητικά από πλευράς ζήτησης των πόρων του συστήματος.

Τέλος στο σενάριο όπου τα κανάλια (6-10) έχουν πιθανότητα κατάληψης από τους πρωτεύοντες χρήστες p=0.5 παρατηρείται μια ενδιάμεση από πλευράς επιτυχιών εκτέλεση του συστήματος. Στην περίπτωση αυτή είναι φανερό ότι ο δευτερεύων χρήστης πρέπει να διακινδυνεύσει κατά την αποστολή μηνυμάτων επείγουσας ανάγκης διότι δεν εξασφαλίζεται με υψηλό ποσοστό η ασφαλής μετάδοση του μηνύματος. Επίσης, ένα μήνυμα μεγάλης διάρκειας θα μπορούσε να εξυπηρετηθεί από τα κανάλια (6-10) όπου δεν παρατηρούνται απότομες μεταβολές ως προς την πιθανότητα κατάληψης του φάσματος από τους πρωτεύοντες χρήστες. Ωστόσο, δεν είναι η καλύτερη δυνατή επιλογή για το δευτερεύοντα χρήστη, διότι η πιθανότητα p=0.5 δεν εξασφαλίζει τη βέβαιη μετάδοση του μηνύματος.

4.5 Συμπεράσματα

Με βάση τα αποτελέσματα της προσομοίωσης που παρουσιάστηκαν προηγουμένως, το σχήμα ενισχυτικής μάθησης που εφαρμόζεται στο υπό μελέτη σύστημα γνωστικών ραδιοεπικοινωνιών έχει επιτυχή συμβολή στην αποδοτικότερη φασματική πρόσβαση. Ο δευτερεύων χρήστης έχει τη δυνατότητα πρόσβασης στο διαθέσιμο αδειοδοτημένο φάσμα με όλο και καλύτερες συνθήκες όσο προχωρά η εξέλιξη της μάθησης. Είναι φανερό ότι τα αποτελέσματα της ενισχυτικής μάθησης μπορούν να εξυπηρετήσουν ειδικού τύπου μηνύματα, όπως μηνύματα τύπου επείγουσας ανάγκης και μηνύματα μεγάλης διάρκειας, όταν η κατανομή πιθανοτήτων των πρωτευόντων χρηστών είναι σταθερή και δεν παρουσιάζει απότομες μεταβολές. Επίσης, το βέλτιστο δυνατό trade off μεταξύ εξοικονόμησης χρόνου και ενέργειας και ανίχνευσης φάσματος επιτυγχάνεται σε κανάλια όπου οι σταθερές τιμές πιθανοτήτων κατάληψης των πρωτευόντων χρηστών έχουν και χαμηλή τιμή. Διότι σε αυτή την περίπτωση υπάρχει μεγαλύτερη πιθανότητα για το δευτερεύοντα χρήστη να έχει επιτυχή πρόσβαση στο φάσμα.

Βιβλιογραφία

[1] J. Mitola III, “Cognitive radio: An integrated agent architecture for software defined radio,” PhD Dissertation, Royal Institute of Technology, Stockholm, Sweden, May 2000

[2] FCC Spectrum Policy Task Force, “FCC Report of the Spectrum Efficiency Working Group”, Nov.2002.

[3] Qing Zhao and Brian M. Sadler, A Survey of Dynamic Spectrum Access, IEEE signal processing magazine May 2007

[4] J. Mitola, “Cognitive radio for flexible mobile multimedia communications,” in Proc. IEEE Int. Workshop Mobile Multimedia Communications, 1999, pp. 3-10.

[5] FCC ET Docket No 03-237, November 2003

[6] S. Haykin, “Cognitive Radio: Brain-Empowered Wireless Communications,” IEEE JSAC, vol. 23, no. 2, Feb. 2005, pp. 201-20.

[7] F. K. Jondral, “Software–defined radio: basics and evolution to cognitive radio,” EURASIP Journal on Wireless Communications and Networking, vol. 5, no. 3, pp. 275-283, 2005

[8] A. Sahai, N. Hoven, R. Tandra, Some fundamental limits in cognitive radio, Allerton Conf. on Commun., Control and Computing 2004, October 2004.

[9] D. Cabric, S. M. Mishra, and R.W. Brodersen, “Implementation issues in spectrum sensing for cognitive radios,” in Proceedings of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, vol.1, November 2005, pp.137-143.

[10] J. Zhao, H. Zheng, and G.-H.Yang, “Distributed coordination in dynamic spectrum allocation networks,” in Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), November 2005, pp.259-268

[11] S. Mangold and L. Berlemann, “IEEE 802.11k: improving confidence in radio resource measurements,” in Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), vol. 2, September 2005, pp. 1009-1013.

[12] “IEEE 1900 standards comitee, IEEE SCC 41. “http://www.scc41.org

[13] Kaelbling, Leslie P., Michael L. Littman, Andrew W.Moore, “Reinforcement Learning: A Survey” (1996), Journal of Artificial Intelligence Research 4:237-285.

[14] Sutton, Richard S., Andrew G. Barto, "Reinforcement Learning: An Intriduction (1998). MIT Press. ISBN 0-262-19398-1-.

[15] Kok-Lim Alvin Yau, Peter Komisarczuc, Paul D.Teal, "Achieving Context Awareness and Intelligence in Cognitive Radio Networks using Reinforcement Learning for Stateful Applications", Victoria University of Wellington, New Zealand, Jan 2010.

[16] C.Watkins, *Learning From Delayed Rewards*, PhD Theis, The Univer- sity of Cambridge, UK, 1989.

[17] Watkins, J C. H., Dayan, P. (1992). Technical Note: Q-Learning. Machine Learning 8:279-292.

[18] Singh, S., and Bertsekas,D.P. (1996), "Reinforcement Learning for Dynamic Channe; Allocation in Ceccular Telephone Systems", MIT Press, Cambridge, Massachusetts 1996.

[19] Hang Su and XiZhang,"Cross-Layer Based Opportunistic MAC Protocols for QoS Provisionings Over Cognitive Radio Wireless Networks". IEEE Journal on selected areas in communications, vol.26, no. 1, January 2008.

[20] M. McHenry, "NSF spectrum occupancy measurements ," The Shared Spectrum Company, http://www.sharedspectrum.com/?section=nsf_measurments, Tech. Rep., 2005.

[21] S.Brandles, M Schnell, U Berthold, and F. K. Jondral, "OFDM based overlay systems-design challenges and solutions ," *in Personal Indoor and Mobile Communications, 2007. IEEE 18th International Symposium on, 3-7 Sept. 2007 , pp 1-5.*

[22] T. Weiss and F. K. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency,"*IEEE Commun. Mag.,* vol. 42, pp. 8-14, 2004.

[23] U. Berthold and F. K. Jondral,: Distributed detection in OFDM based ad hoc overlay systems," *in IEEE 67th Vehicular Technology Conference, VTC Spring,* 2008.

[24] P. Papadimitratos, S. Sankaranarayanan, and A. Mishra,"A bandwidth sharing approach to improve licensed spectrum utilization," IEEE Communications Magazine, vol. 43, no. 12, pp. S10-S14, December 2005.

[25] Z. Tian and G. B. Giannakis, "*Compressed sensing for wideband cognitive radios," in IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, vol.4, 2007,pp. IV-1357-IV-1360.

[26] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: APOMDP framework, "*IEEE J. Select. Areas Commun*., vol. 25, no. 3, pp. 589-600, April 2007.

[27] F. Fu and M. van der Schaar, "A new theoretic framework for cross-layer optimization," in Proc. IEEE Int. Conf. on Image Process. 2008 (ICIP 2008), to appear 2008

[28] I. F. Akyilidiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, pp. 201-220, Feb. 2005

[29] B. Fette, *Cognitive Radio Technology* : Newnes, 2006.

[30] L. Dasilva and A. Mackenzie, "Cognitive Networks: Tutorial," in *CrownCom Orlando*, FL, July 2007.

[31] M. Bublin, J. Pan, I. Kambourov, and P. Slanina, "Distributed spectrum sharing by reinforcement and game theory," presented at 5th Karlsruhe workshop on software radio, Karlsruhe, Germany, March.2008.

[32] S. R.Saunders, *Antennas and propagation for wireless communication systems*: Wiley, 1999.

[33] J.Nie and S. Haykin , "A Dynamic Channel Assignment Policy Through Q-Learning,"IEEE Transactions on Neural Networks and Applications, vol.11, pp.779-797, December, 2006.

[34] AAbdel-Fattah, Y .M. (1987). Stochastic automata modelling of gertain problems of collective behaviour. *IEEE Transactions on Systems, Man and Cybernetics*, 13(3), 236-241.

[35] Bertsekas, D., & Tsitsiklis, J. (1989). *Parallel and Distributed Computation: Numercial Methods*. Prentice Hall.

[36] BBillard, E., & Pasquale, J. (1993). Effects of delayed communication in dynamic group formation. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(5), 1265-1275.

[37] Bond, A. H., & Gasser, L. (1988). *Readings in Distributed Artificial Intelligence.* Ablex Publishing Corporation.

[38] Bonomi, F., Doshi, B., Kaufmann, J., Lee, T., & Kumar, A. (1990). A.900). A case study of adaptive load balancing algorithm. *Queuing Systems*, 7, 23-49

[39] Durfee, E. H., Lesser, V . R., Corkill, D. (1987). Coherent cooperation among communicating problem solvers, *IEEE Transactions on Computers*, 36, 1275-1291.

[40] Glockner, A., &Pasquale, J. (1993). Coadaptive behaviour in a simple distributed job scheduling system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 902-907.

[41] Gmytrasiewicz, P., Durfee, E., &Wehe,D. (1991). The utility of communication in coordinating intelligent agents. In *Proc. Of the 9th Nat. Conf. on Artificial Intelligence (AAAAI-91)*, pp. 166-172.

[42] Kindermann, R., & Snell, S. L. (1980). *Markov Random Fields and their Applications. American Mthematical Society*

[43] Kosoresow, A. P. (1993). A fast first-cut protocol for agent coordination. In Proc. *of the 11th Nat. Conf. on Artificial Intelligence* (AAAI-93), pp. 237-242.

[44] Kraus, S., &Wilkenfeld, J. (1991). The function of time in cooperative negotiations. In *Proc. of the 9th Nat. Conf. on Artificial Intelligence (AAAI-91)*, pp. 179-184.

[45] Mehra, P. (1992). Automated Learning of Load-Balancing Strategies For A Distributed Computer System. Ph.D thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign.

[46] Mehra, P., &Wah, B. W.(1993). Population-based learning of load balancing policies for a distributed computer system. In *proceedings of Computing in Aerospace 9 Conference, AIAA*, pp. 1120-1130.

[47] Mirchandaney, R., & Stankovic, J. (1986). Using stochastic learning automata for job scheduling in distributed processing systems. Journal of Parallel and Distributed Computing, 3, 527-552.

[48] Shoham, Y., &Tennenholtz, M. (1994). Co-learning and the evolution of social activity. Tech. rep. STAN-CS-TR-94-1511, Dept. of Computer Science, Standford University.

[49] J. MitolaIII; Maquire , G.Q., Jr., “Cognitive radio: making software radios more personal”, Personal Communications, IEEE [see also IEEE Wireless Communications], Vol. 6, No.4, pp.13-18, Aug 1999

[50] B. A. Fette, “Cognitve Radio Technology”, Elsevier Science & Technology Books, published in july, 2006

[51] I. Katzela and M. Naghshineh, Channel Assignment Schemes for Ceccular Mobile Telecommunication Systems: A Comprehesive Survey, Personal Communications, IEEE, Vol.3, pp10-31, June 1996

[52] L. Berlemann, S. Mangold, G. Hiertz, B. Walke, “Policy Defined Spectrum Sharing and Medium Access for Cognitive Radios”, Journal of communications. Vol. 1, No April 2006

[53] M. Yang and D. Grace, “Interaction and Coexistance of Multicast Terrestial Communication Systems with Area Optimized Channel Assignments”, COGCOM 2008, Hangzhou, China, Aug, 2008

[54] Ying-Chang Liang, Yonghong Zeng, Edward Peh, and Anh Tuan Hoang, “Sensing-Throughpout Tradeoff for Cognitive Radio Networks”, Institute for Infocomm Reasearch 21 Heng Mui Terrace, Singapore 119613