



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Εφαρμογή Ενισχυτικής Μάθησης σε Συστήματα Γνωστικών
Ραδιοεπικοινωνιών**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Δ. Χαλικόπουλος

Επιβλέπων : Παναγιώτης Κωττής
Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2010



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Εφαρμογή Ενισχυτικής Μάθησης σε Συστήματα Γνωστικών Ραδιοεπικοινωνιών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Δ. Χαλικόπουλος

Επιβλέπων : Παναγιώτης Κωττής
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16-07-2010

.....
Π. Κωττής
Καθηγητής Ε.Μ.Π

.....
Χ. Καμάλης
Καθηγητής Ε.Μ.Π

.....
Γ. Φικιώρης
Επ. Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2010

.....
Χαλικόπουλος Δημήτριος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright © Χαλικόπουλος Δημήτριος, 2010

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Στα πλαίσια αυτής της διπλωματικής εργασίας θα ήθελα να εκφράσω τις πιο θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή κύριο Π. Κωττή για την πολύτιμη συμβολή του και την άριστη συνεργασία μας, καθώς και τη διδάκτορα Αγγελική Κορδαλή για τις σημαντικές υποδείξεις και την ενδελεχή βοήθειά της κατά την εκπόνηση της. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και το φιλικό μου περιβάλλον για την αμέριστη συμπαράστασή τους.

ΠΕΡΙΛΗΨΗ

Η ραγδαία εξέλιξη των υπηρεσιών ασύρματης επικοινωνίας έχει οδηγήσει στην ελαχιστοποίηση του διαθέσιμου φάσματος συχνοτήτων, καθιστώντας το σπάνιο φυσικό πόρο. Ωστόσο μελέτες έχουν δείξει ότι η έλλειψη αυτή οφείλεται κυρίως στον στατικό τρόπο εκχώρησης του φάσματος, που οδηγεί στην υποχρησιμοποίηση του. Για την επίλυση του προβλήματος προτάθηκε η δυναμική φασματική πρόσβαση, που επιτρέπει σε μη αδειοδοτημένους χρήστες να χρησιμοποιήσουν το φάσμα ευκαιριακά. Η δυναμική πρόσβαση επιτυγχάνεται με τη χρήση γνωστικών ραδιοεπικοινωνιών, που δίνουν την δυνατότητα σε ένα χρήστη να αποκτά γνώση του περιβάλλοντος του και να προσαρμόζει τις παραμέτρους λειτουργίας του σε αυτό. Οι δευτερογενείς χρήστες ανιχνεύουν το αδειοδοτημένο φάσμα και αποκτούν πρόσβαση σε αυτό αν διαπιστωθεί κάποιο φασματικό κενό.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η μελέτη της εφαρμογής τεχνικών ενισχυτικής μάθησης στη διαδικασία ανίχνευσης φάσματος, από δευτερογενείς χρήστες γνωστικών ραδιοεπικοινωνιών. Αρχικά, γίνεται μια εκτενής εισαγωγή στις αρχές των γνωστικών ραδιοεπικοινωνιών, ενώ ακολουθεί η παρουσίαση των κυριότερων τεχνικών ανίχνευσης φάσματος που χρησιμοποιούνται σε συστήματα γνωστικών ραδιοεπικοινωνιών. Στη συνέχεια αναλύεται το πρόβλημα ενισχυτικής μάθησης δίνοντας έμφαση στις τεχνικές που στηρίζονται σε διαδικασίες απόφασης Markov. Τέλος, πραγματοποιείται προσομοίωση της εφαρμογής των τεχνικών ενισχυτικής μάθησης, επανάληψη πολιτικών και εκμάθηση Q, στην διαδικασία ανίχνευσης φάσματος και παρουσιάζονται τα αντίστοιχα αποτελέσματα.

Λέξεις κλειδιά

Δυναμική φασματική πρόσβαση, γνωστικές ραδιοεπικοινωνίες, ανίχνευση φάσματος, ενισχυτική μάθηση, διαδικασίες απόφασης Markov, επανάληψη πολιτικών, εκμάθηση Q.

ABSTRACT

The rapid growth of wireless communication services has led to the minimization of the available frequency spectrum, rendering it a scarce natural resource. However studies have shown that this scarcity is mainly due to the static spectrum allocation that leads to spectrum underutilization. As a solution to the problem, dynamic spectrum access was proposed. Dynamic spectrum access allows non-licensed users to access the spectrum opportunistically and is achieved through the use of cognitive radios, which enable users to acquire knowledge of their radio environment and dynamically adjust their transmission parameters. The secondary users sense the licensed spectrum and gain access to it when a spectrum white space is identified.

This thesis aims at examining the application of reinforcement learning techniques in the process of spectrum sensing from secondary cognitive radio users. At first, the cognitive radio principles are presented extensively, while an introduction to the basic spectrum sensing techniques follows. Then, the problem of reinforcement learning is analyzed, focusing on Markov Decision Processes. Finally, the application of the reinforcement learning techniques, policy iteration and Q learning, in the spectrum sensing process is simulated and the corresponding results are presented.

Keywords

Dynamic spectrum access, cognitive radios, spectrum sensing, reinforcement learning, Markov decision processes, policy iteration algorithm, Q learning.

Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή στις γνωστικές ραδιοεπικοινωνίες.....	13
1.1 Χρησιμοποίηση του φάσματος ραδιοσυχνοτήτων	14
1.2 Δυναμική πρόσβαση φάσματος	20
1.2.1 Δυναμικό μοντέλο αποκλειστικής χρήσης (Dynamic Exclusive Use model) ..	21
1.2.2 Μοντέλο ανοιχτού επιμερισμού (Open Sharing Model)	22
1.2.3 Μοντέλο ιεραρχικής πρόσβασης.....	22
1.2.3.1 Τεχνική φασματικής υπόστρωσης (spectrum underlay)	23
1.2.3.2 Τεχνική φασματικής επίστρωσης (spectrum overlay)	24
1.3 Ραδιοεπικοινωνίες λογισμικού (Software-defined radio – SDR)	25
1.4 Γνωστικά συστήματα ραδιοεπικοινωνιών - Cognitive radios	26
1.4.1 Ορισμός	26
1.4.2 Λειτουργίες συστημάτων γνωστικών ραδιοεπικοινωνιών	28
1.4.3 Αρχιτεκτονική πρωτοκόλλων για συστήματα γνωστικών ραδιοεπικοινωνιών	31
1.4.5 Βασικά στοιχεία συστημάτων γνωστικών ραδιοεπικοινωνιών	33
1.5 Θερμοκρασία παρεμβολής.....	35
1.5.1 Ιδανικό μοντέλο θερμοκρασίας παρεμβολής.....	36
1.5.2 Γενικευμένο μοντέλο θερμοκρασίας παρεμβολής.....	37
1.5.3 Χωρητικότητα διαύλου κάτω από το όριο παρεμβολής	37
1.6 Αρχιτεκτονικές γνωστικών ραδιοεπικοινωνιών συστημάτων επόμενης γενιάς...38	
1.7 Πρότυπα γνωστικών επικοινωνιών	41
1.7.1 IEEE SCC 41	41
1.7.2 IEEE 802.22	44
1.7.2.1 Φυσικό στρώμα	46
1.7.2.2 Στρώμα MAC	47
1.8 Πιθανές εφαρμογές γνωστικών ραδιοεπικοινωνιών	48
Κεφάλαιο 2: Ανίχνευση φάσματος	51
2.1 Μη συνεργατική ανίχνευση πομπού.....	52

2.1.1 Ανίχνευση προσαρμοσμένου φίλτρου.....	52
2.1.2 Ανιχνευτής ενέργειας πομπού	53
2.1.3 Ανίχνευση στοιχείου κυκλοστατικότητας	55
2.2 Συνεργατική ανίχνευση	57
2.2.1 Κεντρικά ελεγχόμενη συνεργατική ανίχνευση	59
2.2.2 Κατανεμημένη συνεργατική ανίχνευση.....	59
2.3 Ανίχνευση παρεμβολής	60
2.4 Περίοδος ανίχνευσης.....	60
2.5 Ανίχνευση φάσματος σε σύγχρονα πρότυπα επικοινωνιών.....	62
2.5.1 IEEE 802.11k	62
2.5.2 Bluetooth	63
2.5.3 IEEE 802.22	63
2.6 Πεδία έρευνας στην ανίχνευση φάσματος	65
Κεφάλαιο 3: Ενισχυτική μάθηση	67
3.1 Κλασσικό μοντέλο ενισχυτικής μάθησης	67
3.2 Στοιχεία ενισχυτικής μάθησης.....	69
3.3 Μέτρηση επίδοσης μάθησης.....	72
3.3.1 Τελική σύγκλιση στη βέλτιστη λύση	72
3.3.2 Ταχύτητα σύγκλισης προς τη βέλτιστη λύση	73
3.3.3 Κόστος.....	73
3.4 Αντιστάθμιση εκμετάλλευσης με έρευνα	73
3.5 Στρατηγικές ενισχυτικής μάθησης.....	74
3.6 Καθυστερημένη ανταμοιβή.....	75
3.7 Διαδικασία απόφασης Markov (MDP)	76
3.8 Εύρεση πολιτικής για δεδομένο μοντέλο.....	76
3.8.1 Επανάληψη αξιών (value iteration)	77
3.8.2 Επανάληψη πολιτικών (policy iteration).....	78
3.9 Μερικώς παρατηρήσιμες διαδικασίες απόφασης Markov (POMDP)	80
3.10 Εύρεση πολιτικής χωρίς τη χρήση μοντέλου.....	83
3.10.1 Εκμάθηση Q.....	83
3.10.2 Εφαρμογή εκμάθησης Q σε συστήματα γνωστικών ραδιοεπικοινωνιών	88

Κεφάλαιο 4: Εφαρμογή ενισχυτικής μάθησης στην ανίχνευση φάσματος σε συστήματα γνωστικών ραδιοεπικοινωνιών	91
4.1 Διατύπωση μοντέλου για τη χρήση του φάσματος	92
4.2 Ανίχνευση και πρόσβαση στο φάσμα	93
4.3 Συνάρτηση ανταμοιβής	93
4.4 Μοντέλο ενισχυτικής μάθησης	95
4.4.1 Επανάληψη πολιτικών	95
4.4.2 Εκμάθηση Q.....	96
4.5 Προσομοίωση	97
4.6 Αποτελέσματα προσομοίωσης.....	98
4.6.1 Πιθανότητα μη εκπομπής λόγω ανίχνευσης κατειλημμένου καναλιού.....	98
4.6.2 Σύγκλιση της εκμάθησης Q προς τη βέλτιστη πολιτική.....	101
4.6.3 Εφαρμογή ενισχυτικής μάθησης όταν η φασματική χρήση δεν παρουσιάζει κανονικότητα	107
4.7 Συμπεράσματα.....	109

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΓΝΩΣΤΙΚΕΣ ΡΑΔΙΟΕΠΙΚΟΙΝΩΝΙΕΣ

Οι γνωστικές ραδιοεπικοινωνίες (cognitive radios) αποτελούν ένα νέο τρόπο σχεδίασης ασύρματων τηλεπικοινωνιακών συστημάτων, που στοχεύει στην αποτελεσματικότερη εκμετάλλευση του φάσματος των ραδιοσυχνοτήτων.

Κίνητρο για την ανάπτυξη των γνωστικών ραδιοεπικοινωνιών αποτέλεσε η αλματώδης αύξηση των υπηρεσιών ασύρματης επικοινωνίας, που έχει οδηγήσει στην ελαχιστοποίηση του διαθέσιμου εύρους συχνοτήτων και καθιστά το φάσμα ραδιοσυχνοτήτων σπάνιο φυσικό πόρο και σημαντικό περιοριστικό παράγοντα για την εισαγωγή νέων ασύρματων εφαρμογών. Το μεγαλύτερο μέρος του διαθέσιμου φάσματος έχει ήδη παραχωρηθεί στις υπάρχουσες ασύρματες εφαρμογές με αποτέλεσμα μόνο μικρά τμήματα φάσματος να μπορούν να παραχωρηθούν σε νέες εφαρμογές. Ωστόσο, διάφορες μετρήσεις και μελέτες έχουν δείξει ότι ενώ μερικές ζώνες συχνοτήτων είναι πλήρως κατειλημμένες σε συγκεκριμένες τοποθεσίες και ώρες, υπάρχουν πολλές ζώνες που είτε είναι μερικώς κατειλημμένες είτε επί το πλείστον αχρησιμοποίητες[1][2]. Η έλλειψη του φάσματος δηλαδή, οφείλεται κυρίως στην πολιτική διαχείρισης παρά στη σπανιότητα των διαθέσιμων συχνοτήτων. Συγκεκριμένα, εξαιτίας της στατικής εκχώρησης φάσματος που επιτρέπει την χρησιμοποίησή του μόνο από αδειοδοτημένους φορείς (primary users – licensed users), το φάσμα χρησιμοποιείται μη αποδοτικά και μεγάλο μέρος του παραμένει ανεκμετάλλευτο.

Για την επίλυση του προβλήματος της υποχρησιμοποίησης του φάσματος, η ομοσπονδιακή επιτροπή επικοινωνιών των Ηνωμένων Πολιτειών (FCC – Federal Communication Commission) πρότεινε πρόσφατα μια νέα πολιτική διαχείρισης φάσματος που προβλέπει τη δυναμική πρόσβαση σε αυτό. Οι γνωστικές ραδιοεπικοινωνίες αποτελούν την τεχνολογία φυσικού στρώματος η οποία καθιστά δυνατή τη δυναμική εκχώρηση φάσματος.

Στόχος των γνωστικών ραδιοεπικοινωνιών είναι η αποτελεσματικότερη και αποδοτικότερη πρόσβαση στο φάσμα ραδιοσυχνοτήτων, προσφέροντας νέες ευκαιρίες χρήσης του, που έως σήμερα δεν διατίθενται. Οι γνωστικές

ραδιοεπικοινωνίες στηρίζονται στην τεχνολογία των ραδιοεπικοινωνιών λογισμικού (software defined radio – SDR) όπου οι παράμετροι εκπομπής και λήψης του τηλεπικοινωνιακού συστήματος ρυθμίζονται δυναμικά με χρήση λογισμικού. Στα δίκτυα γνωστικών ραδιοεπικοινωνιών οι μη αδειοδοτημένοι χρήστες μπορούν να ανιχνεύουν περιοδικά το φάσμα με στόχο τον εντοπισμό αχρησιμοποίητων καναλιών και ανάλογα με το αποτέλεσμα, να συγχρονίζουν τους πομποδέκτες τους στο αναγνωρισμένο ελεύθερο κανάλι προκειμένου επικοινωνήσουν μεταξύ τους χωρίς να προκαλούν παρεμβολή στους αδειοδοτημένους χρήστες.

1.1 Χρησιμοποίηση του φάσματος ραδιοσυχνοτήτων

Το ηλεκτρομαγνητικό φάσμα ραδιοσυχνοτήτων αποτελεί φυσικό πόρο η χρήση του οποίου ρυθμίζεται από τις κυβερνήσεις των κρατών μέσω μιας διαδικασίας διαχείρισης που ονομάζεται κατανομή φάσματος ή κατανομή συχνοτήτων. Για λόγους ομοιομορφίας οι κυβερνήσεις συνεργάζονται μεταξύ τους σε θέματα κατανομής φάσματος και προτυποποίησης. Μια σειρά οργάνων και ομάδων προτυποποίησης συμπεριλαμβανομένων των ακόλουθων:

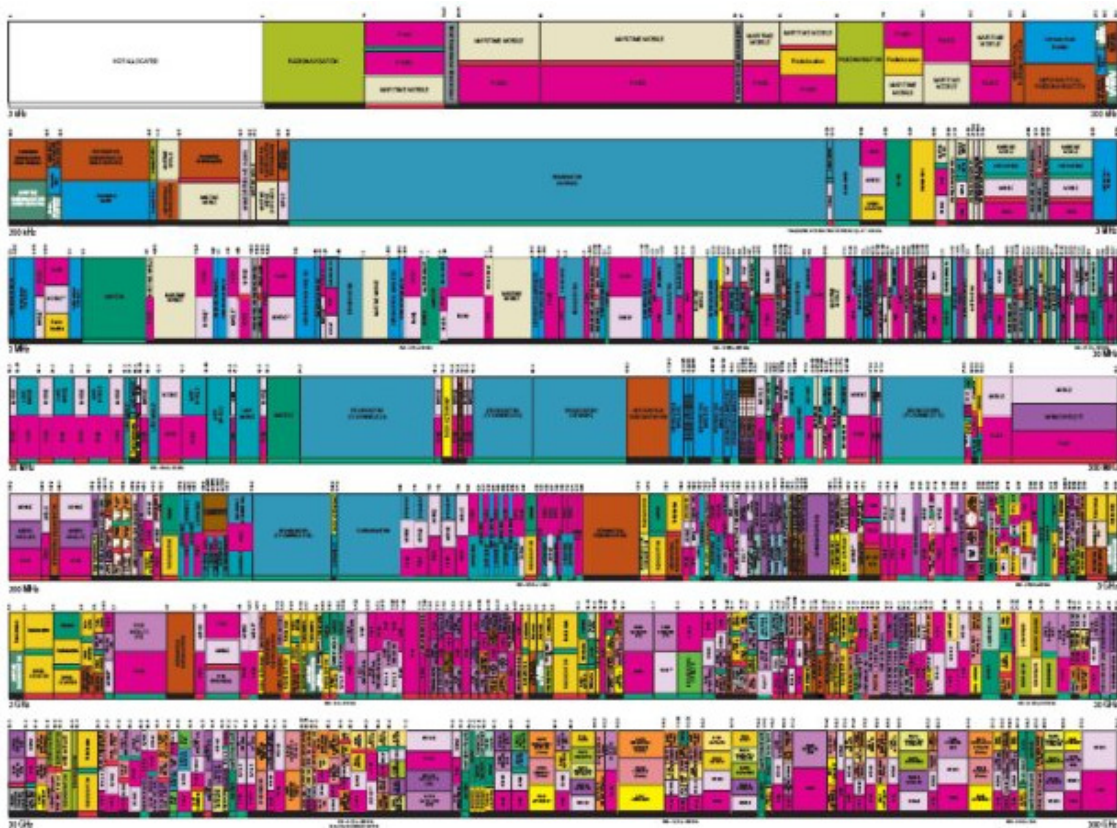
- Διεθνής Ένωση Τηλεπικοινωνιών (International Telecommunication Union - ITU)
- Ευρωπαϊκή Διάσκεψη των Διοικήσεων Ταχυδρομείων και Τηλεπικοινωνιών (European Conference of Postal and Telecommunications Administrations - CEPT)
- Ευρωπαϊκό Ινστιτούτο Τηλεπικοινωνιακών Προτύπων (European Telecommunications Standards Institute - ETSI)
- Διεθνή Ειδική Επιτροπή Ηλεκτρομαγνητικών Παρεμβολών (International Special Committee on Radio Interference - CISPR)

είναι υπεύθυνα για τους κανονισμούς και τις ρυθμιστικές συμφωνίες που αφορούν τη διαχείριση του φάσματος ραδιοσυχνοτήτων. Οι ομάδες αυτές έχουν κατανειμίσει στατικά το φάσμα συχνοτήτων στους τρεις παρακάτω 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 που έχουν παραχωρηθεί σε ραδιοερασιτέχνες, αρκεί να τηρούνται ορισμένοι κανόνες εκπομπής.
- 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 σχήμα 1.1 όπου εικονίζεται ο χάρτης φασματικής κατανομής των Ηνωμένων Πολιτειών Αμερικής.

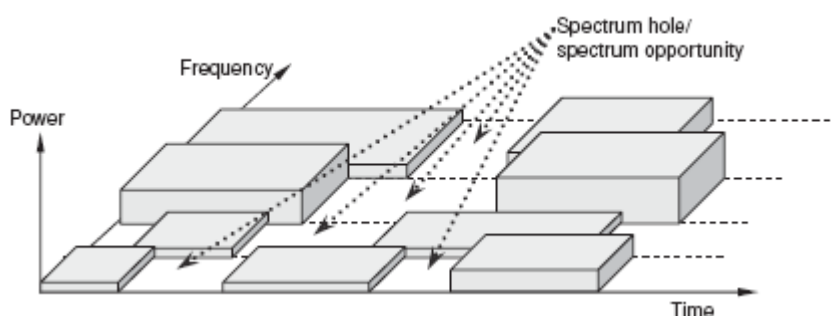


Σχήμα 1.1: Χάρτης φασματικής κατανομής των Ηνωμένων Πολιτειών Αμερικής

Ωστόσο, πρόσφατες μελέτες της ομοσπονδιακής επιτροπής επικοινωνιών των Ηνωμένων Πολιτειών (Federal Communications Commission- FCC) έχουν δείξει ότι η παραδοσιακή πολιτική στατικής κατανομής είναι ανεπαρκής για την

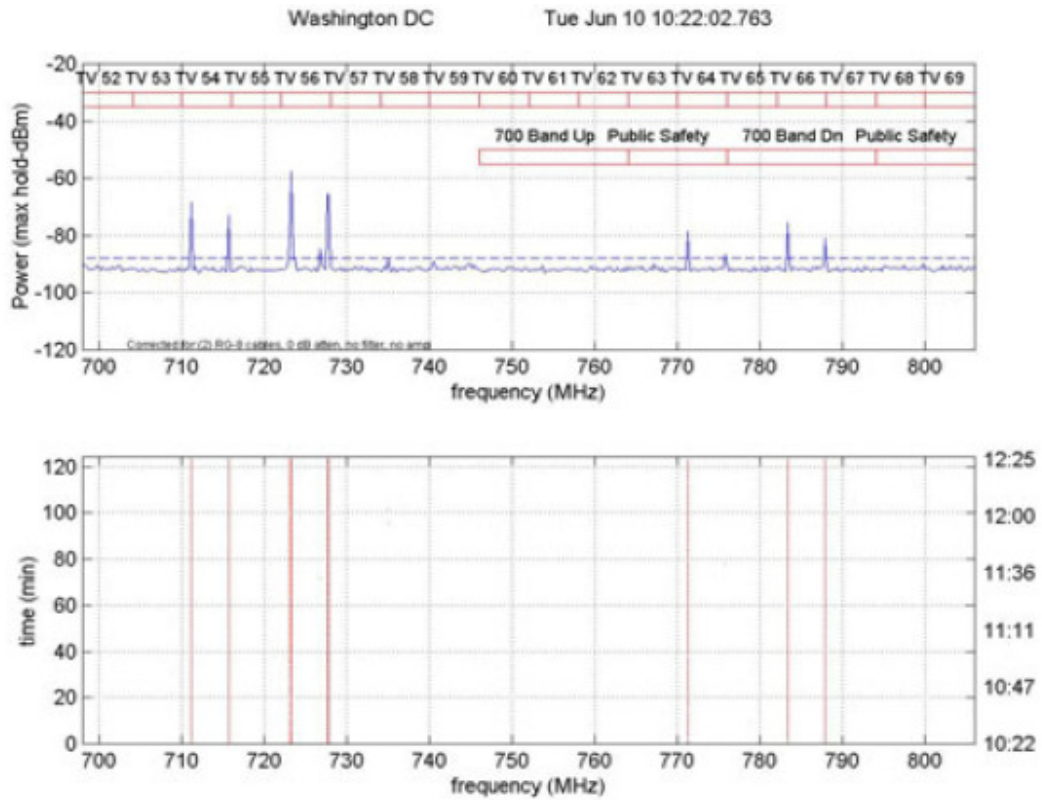
αντιμετώπιση της ραγδαίας εξέλιξης των ασύρματων επικοινωνιών. Η σύγχρονη τάση για ευρυζωνική πρόσβαση στο διαδίκτυο, οι δυνατότητες των νέων κινητών τερματικών και οι συνεχώς αυξανόμενες απαιτήσεις των καταναλωτών έχουν ως αποτέλεσμα τη σταθερή αύξηση της ζήτησης πρόσθετου εύρους ζώνης, σε αντιδιαστολή με το διαθέσιμο φάσμα συχνοτήτων που είναι πεπερασμένο.

Το Νοέμβριο του 2002 η FCC δημοσίευσε μια αναφορά σύμφωνα με την οποία ορισμένες αδειοδοτημένες ζώνες συχνοτήτων σε διάφορες γεωγραφικές περιοχές δεν χρησιμοποιούνται καθόλου ή παραμένουν ανενεργές για μεγάλα χρονικά διαστήματα[1]. Επιπλέον, ακόμη και όταν κάποια φασματική ζώνη είναι ενεργή, η εκρηκτική φύση της κίνησης των δεδομένων έχει ως αποτέλεσμα την ύπαρξη φασματικών κενών για δεδομένες χρονικές περιόδους όπως φαίνεται από το σχήμα 1.2. Τα φασματικά αυτά κενά είναι γνωστά και ως κενός φασματικός χώρος (spectrum white space) ή φασματικές τρύπες (spectrum holes).

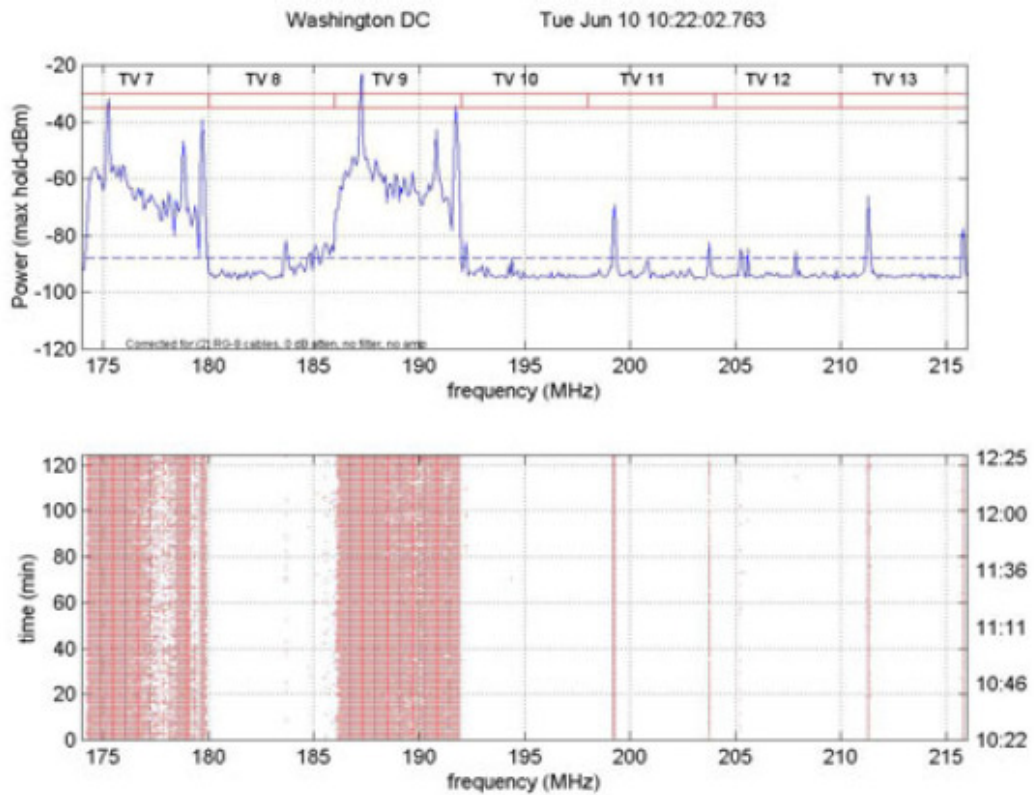


Σχήμα 1.2: Αναπαράσταση φασματικών κενών

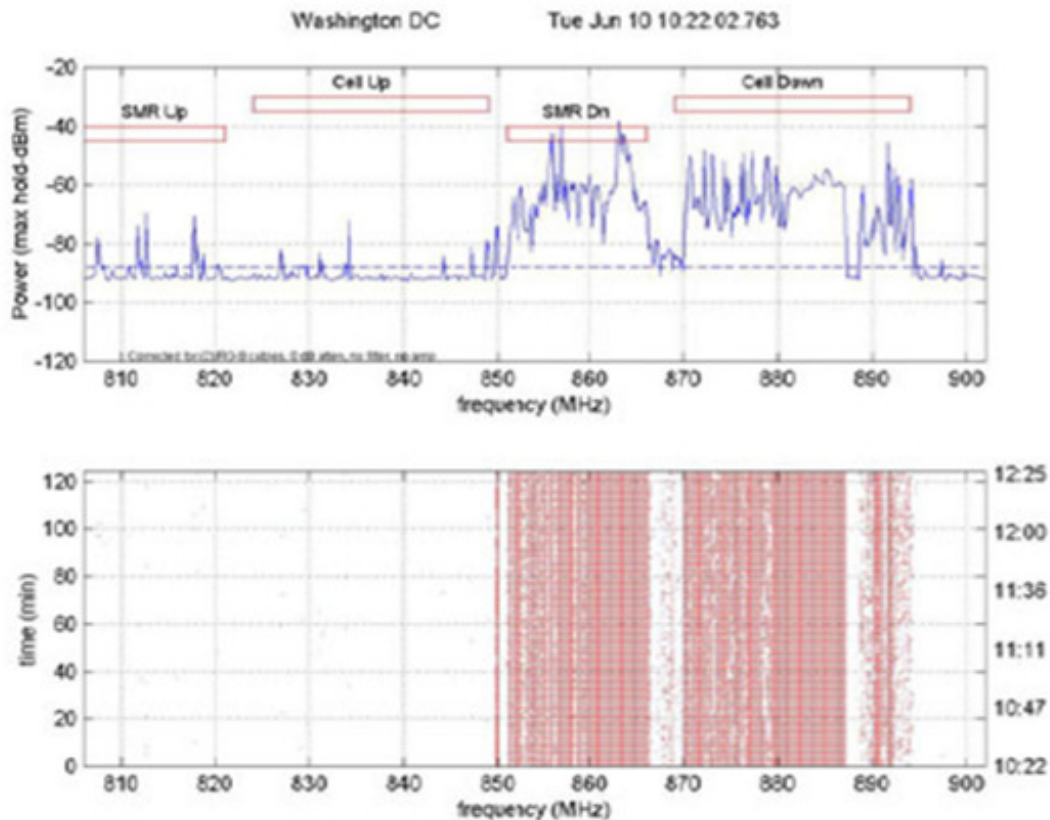
Μετρήσεις που πραγματοποιήθηκαν από την εταιρία Shared Spectrum Company στην Ουάσινγκτον δείχνουν ότι ακόμα και στην πλέον συνωστισμένη περιοχή κοντά στο κέντρο της πόλης, όπου υπάρχει έντονη χρήση τόσο εμπορικού όσο και κρατικού ραδιοφάσματος, το 62% του φάσματος παραμένει αχρησιμοποίητο. Στα παρακάτω σχήματα ακολουθούν παραδείγματα των μετρήσεων αυτών σε διάφορες ζώνες συχνοτήτων που παρουσιάζουν χαμηλή, μέση και υψηλή χρησιμοποίηση φάσματος αντίστοιχα.



Σχήμα 1.3: Παράδειγμα χαμηλής φασματικής χρήσης

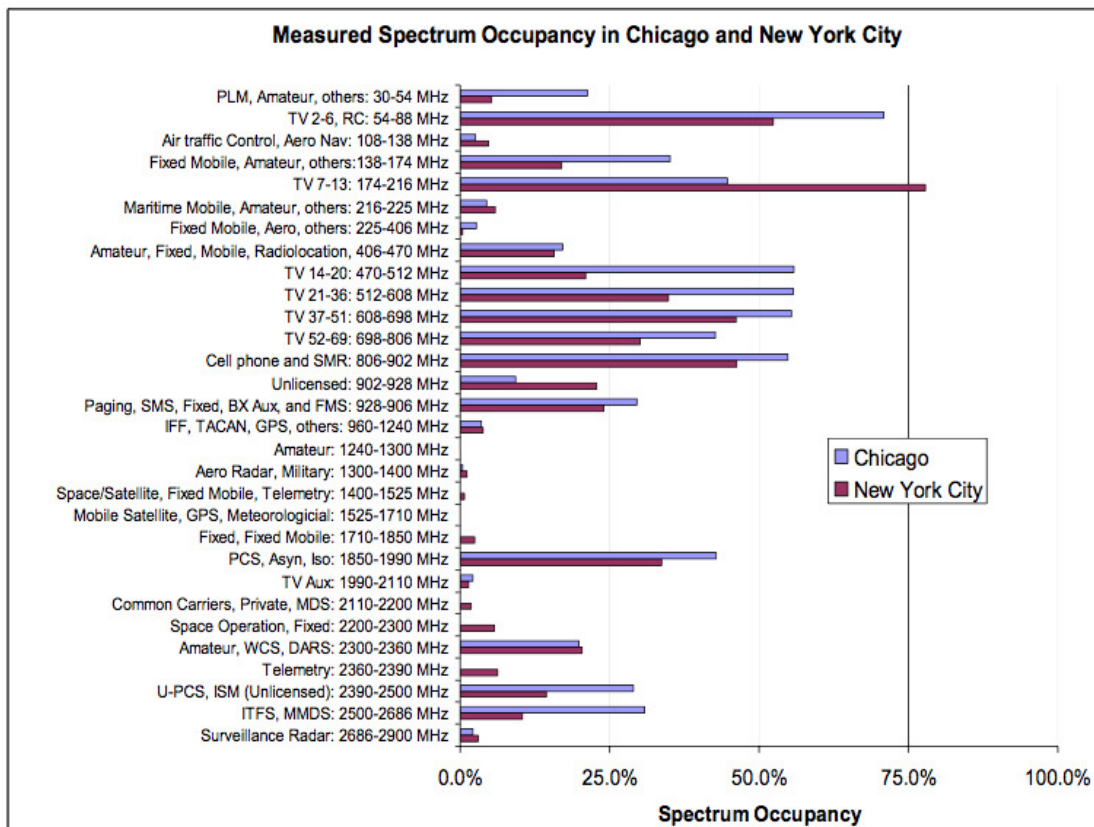


Σχήμα 1.4: Παράδειγμα μέσης φασματικής χρήσης



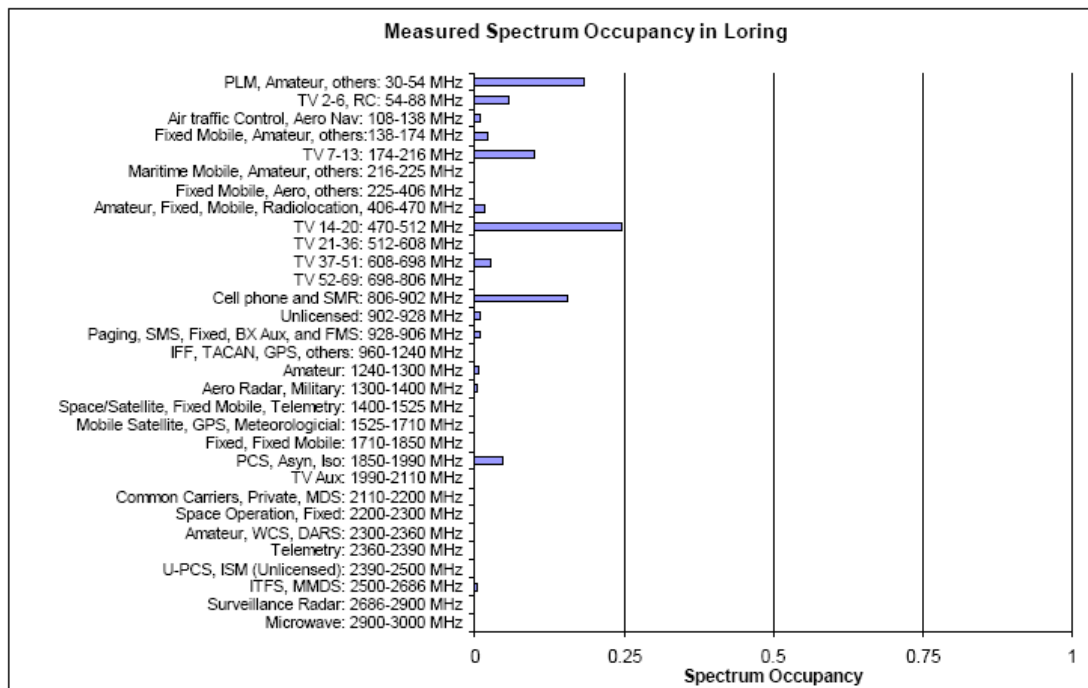
Σχήμα 1.5: Παράδειγμα υψηλής φασματικής χρήσης

Άλλες μετρήσεις που πραγματοποίησε η Shared Spectrum Company το 2004 στη Νέα Υόρκη[3] έδειξαν ότι η χρήση φάσματος κυμαίνεται από 1% στις ζώνες 1240-1300 MHz μέχρι και 77% στις ζώνες 174 MHz – 216 MHz με τη μέση χρήση να είναι 1.1%. Επιπλέον, μετρήσεις της Shared Spectrum Company στο Σικάγο το 2005[2] έδειξαν ότι η φασματική χρήση κυμαίνεται από λιγότερο του 1% στις ζώνες 1240-1300 MHz μέχρι και 70.9% στη ζώνη 54MHz – 88MHz με τη μέση χρήση να μην υπερβαίνει το 17.4%. Στο σχήμα 1.5 φαίνονται αναλυτικά τα ποσοστά φασματικής χρήσης σε κάθε ζώνη συχνοτήτων για τις πόλεις του Σικάγο και της Νέας Υόρκης. Από τα παραπάνω καθίσταται σαφές ότι ακόμα και σε πολυπληθείς αστικές περιοχές το αδειοδοτημένο φάσμα παραμένει σε μεγάλο βαθμό αναξιοποίητο .



Σχήμα 1.6: Ποσοστό απασχόλησης φάσματος σε Νέα Υόρκη και Σικάγο

Αντίστοιχες μετρήσεις σε αγροτικές περιοχές δείχνουν ακόμα μεγαλύτερη υποχρησιμοποίηση του φάσματος. Ενδεικτικά, μετρήσεις που πραγματοποιήθηκαν από τη Shared Spectrum Company στο Loring Commerce Centre στην περιοχή Λάιμστον (Limestone) της πολιτείας Μέιν [5] έδειξαν ότι η μέση χρήση φάσματος δεν υπερβαίνει το 1.7%. Η χρησιμοποίηση φάσματος φαίνεται αναλυτικά στο παρακάτω σχήμα.



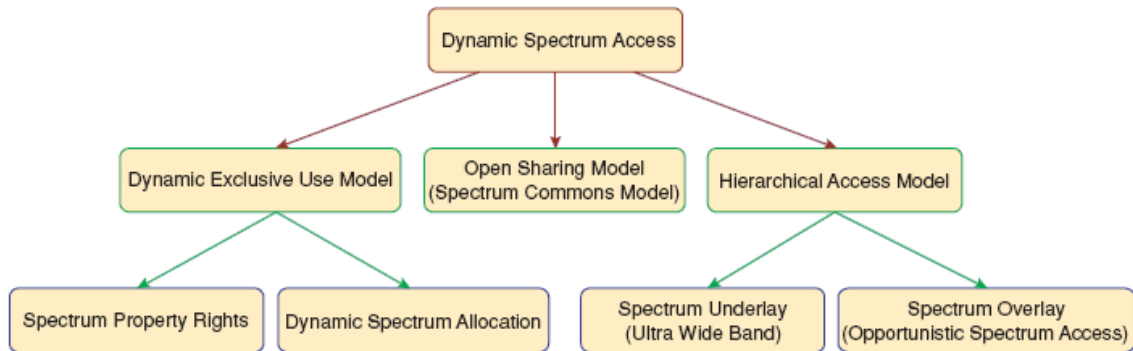
Σχήμα 1.7 Ποσοστό απασχόλησης φάσματος στο Λόρινγκ

Τα ανωτέρω αποτελέσματα έρχονται σε αντίθεση με την αντίληψη έλλειψης φάσματος, δεδομένου ότι παρατηρείται αφθονία ανεκμετάλλευτων συχνοτήτων ενώ η εξάντλησή του είναι φυσικό επακόλουθο των πολιτικών πρόσβασης και της διαδικασίας αδειοδότησης. Τα χαμηλά αυτά επίπεδα χρησιμοποίησης αδειοδοτημένων συχνοτήτων φανερώνουν την ανάγκη αλλαγής της πολιτικής εκχώρησης φάσματος, γεγονός που συνδέεται άμεσα με την αντιμετώπιση της έλλειψης νέων διαθέσιμων ραδιοσυχνοτήτων. Είναι αναγκαία, συνεπώς, η μετάβαση από τη στατική εκχώρηση φάσματος στην πολιτική της δυναμικής φασματικής πρόσβασης (DSA – Dynamic Spectrum Access), με στόχο την αποτελεσματικότερη κατανομή συχνοτήτων αλλά και την αύξηση της χωρητικότητας των ασύρματων υπηρεσιών.

1.2 Δυναμική πρόσβαση φάσματος

Σε αντίθεση με την τωρινή πολιτική διαχείρισης φάσματος, ο όρος δυναμική πρόσβαση φάσματος περιλαμβάνει διάφορες προσεγγίσεις αναδιάρθρωσης

φάσματος. Οι διάφορες στρατηγικές δυναμικής πρόσβασης μπορούν να κατηγοριοποιηθούν σε τρία μοντέλα όπως φαίνεται από το σχήμα 1.8.



Σχήμα 1.8: Κατηγορίες στρατηγικών δυναμικής πρόσβασης

1.2.1 Δυναμικό μοντέλο αποκλειστικής χρήσης (Dynamic Exclusive Use model)

Το μοντέλο αυτό διατηρεί τη βασική δομή της υφιστάμενης πολιτικής ρύθμισης του φάσματος: οι ζώνες συχνοτήτων αδειοδοτούνται σε υπηρεσίες για αποκλειστική χρήση. Ωστόσο, οι χρήστες που έχουν πρόσβαση στο φάσμα και ο τύπος της ασύρματης υπηρεσίας που χρησιμοποιείται ενδέχεται να αλλάζουν ανά χρονική στιγμή. Η βασική ιδέα είναι η εισαγωγή ευελιξίας για την αποτελεσματικότερη αξιοποίηση του φάσματος. Μέχρι τώρα, έχουν προταθεί δύο προσεγγίσεις για το μοντέλο αυτό: η προσέγγιση φασματικών δικαιωμάτων ιδιοκτησίας (spectrum property rights) και η δυναμική κατανομή φάσματος (dynamic spectrum allocation). Η πρώτη προσέγγιση επιτρέπει στους δικαιούχους (licensees – primary users) να εμπορεύονται το φάσμα και να επιλέγουν ελεύθερα την τεχνολογία που υιοθετείται. Η οικονομία και η αγορά, επομένως, θα διαδραματίζουν ένα σημαντικό ρόλο στην αποδοτικότερη χρησιμοποίηση του περιορισμένου αυτού πόρου. Επίσης, καίτοι οι δικαιούχοι έχουν το δικαίωμα να ενοικιάζουν ή να μοιράζονται το αδειοδοτημένο φάσμα δεν υπάρχει σχετική υποχρέωση από κάποια ρυθμιστική αρχή.

Η δεύτερη προσέγγιση προτάθηκε από το ευρωπαϊκό πρόγραμμα DRiVE με στόχο τη βελτίωση της αποτελεσματικότητας της χρήσης του φάσματος μέσω

δυναμικής εκχώρησης, αξιοποιώντας τα χωρικά και χρονικά στατιστικά κίνησης των διαφόρων υπηρεσιών. Σε συγκεκριμένη περιοχή και συγκεκριμένο χρονικό διάστημα, το φάσμα διατίθεται σε υπηρεσίες για αποκλειστική χρήση. Βέβαια η εκχώρηση αυτή διαφέρει της τρέχουσας πολιτικής σε πολύ μεγαλύτερη κλίμακα.

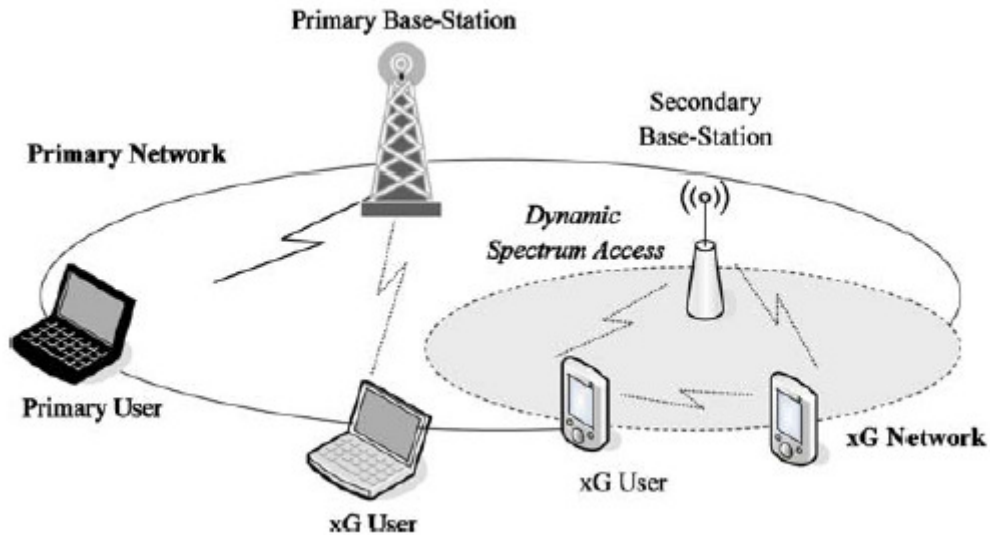
Ωστόσο, με βάση μόνο το μοντέλο αποκλειστικής χρήσης δεν είναι δυνατό να αντιμετωπιστούν και να αξιοποιηθούν τα λευκά φασματικά διαστήματα (white spaces) που προκύπτουν από την εκρηκτική (bursty) φύση της ασύρματης τηλεπικοινωνιακής κίνησης.

1.2.2 Μοντέλο ανοιχτού επιμερισμού (Open Sharing Model)

Αναφέρεται επίσης και ως μοντέλο κοινού φάσματος (spectrum common model) και χρησιμοποιεί την ανοικτή ανταλλαγή φάσματος μεταξύ ομότιμων χρηστών ως βάση για τη διαχείριση μιας φασματικής περιοχής. Στο μοντέλο ανοιχτού επιμερισμού οι χρήστες έχουν το ίδιο δικαίωμα πρόσβασης στο φάσμα. Οι υποστηρικτές του μοντέλου αυτού τάσσονται υπέρ της χρησιμοποίησής του λόγω της αδιαμφισβήτητης επιτυχίας των ασύρματων υπηρεσιών (πχ WiFi) που λειτουργούν στη μη αδειοδοτημένη βιομηχανική, επιστημονική και ιατρική ζώνη ραδιοσυχνοτήτων (ISM - Industrial, Scientific and Medical radio band). Συγκεντρωτικές και κατανεμημένες στρατηγικές κατανομής φάσματος έχουν αρχικά διερευνηθεί για την αντιμετώπιση των τεχνολογικών προκλήσεων στο πλαίσιο του συγκεκριμένου μοντέλου διαχείρισης ραδιοφάσματος.

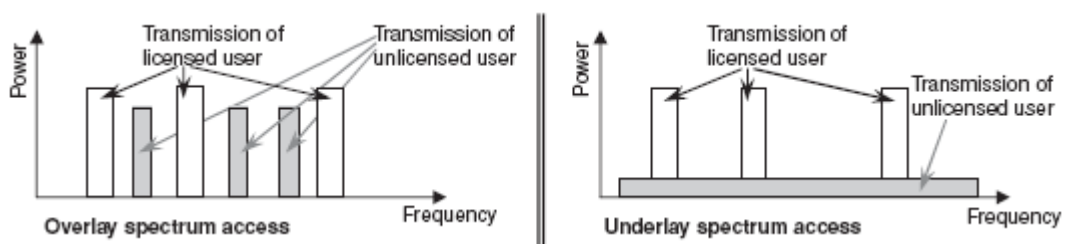
1.2.3 Μοντέλο ιεραρχικής πρόσβασης

Το μοντέλο αυτό υιοθετεί μια ιεραρχική δομή πρόσβασης με πρωτεύοντες και δευτερεύοντες χρήστες. Συγκεκριμένα, οι δευτερεύοντες χρήστες μπορούν χρησιμοποιούν ευκαιριακά το φάσμα όταν αυτό δεν είναι κατειλημμένο ή δεν αξιοποιείται πλήρως από τους πρωτεύοντες χρήστες (σχήμα 1.9).



Σχήμα 1.9: Απεικόνιση δικτύων που ακολουθούν το μοντέλο ιεραρχικής πρόσβασης

Η βασική ιδέα είναι η δυνατότητα χρησιμοποίησης του αδειοδοτημένου φάσματος από δευτερεύοντες χρήστες εφόσον περιορίζονται επαρκώς οι παρεμβολές που προκαλούνται στους πρωτογενείς χρήστες (π.χ. επιβάλλεται η πιθανότητα σύγκρουσης να παραμένει κάτω από ένα ορισμένο επίπεδο). Έχουν εξεταστεί δύο προσεγγίσεις καταμερισμού του φάσματος μεταξύ κύριων και δευτερευόντων χρηστών: η τεχνική φασματικής υπόστρωσης (spectrum underlay) και η τεχνική φασματικής επίστρωσης (spectrum overlay), (σχήμα 1.10).



Σχήμα 1.10: Απεικόνιση φασματικής επίστρωσης και φασματικής υπόστρωσης

1.2.3.1 Τεχνική φασματικής υπόστρωσης (spectrum underlay)

Σε αυτήν την περίπτωση ένας δευτερογενής χρήστης μπορεί να εκπέμπει ταυτόχρονα με ένα πρωτογενή χρήστη. Ωστόσο, η προσέγγιση φασματικής υπόστρωσης επιβάλλει αυστηρούς περιορισμούς ως προς την ισχύ εκπομπής των δευτερογενών χρηστών έτσι ώστε η παρεμβολή που προκαλούν στους πρωτογενείς

χρήστες να παραμένει κάτω από το ανεκτό επίπεδο θορύβου. Με τη φασματική εξάπλωση των μεταδιδόμενων σημάτων σε μια ευρεία ζώνη ραδιοσυχνοτήτων, οι δευτερογενείς χρήστες έχουν τη δυνατότητα να επιτυγχάνουν υψηλούς ρυθμούς μετάδοσης δεδομένων με εξαιρετικά χαμηλή ισχύ εκπομπής. Η τεχνική της φασματικής υπόστρωσης μπορεί να χρησιμοποιηθεί για δυναμική πρόσβαση φάσματος χρησιμοποιώντας CDMA ή UWB τεχνολογία.

1.2.3.2 Τεχνική φασματικής επίστρωσης (spectrum overlay)

Η τεχνική της φασματικής επίστρωσης επινοήθηκε αρχικά από το Mitola [8] υπό τον όρο συγκέντρωση φάσματος (spectrum pooling) και, στη συνέχεια, ερευνήθηκε από τον οργανισμό προηγμένων ερευνητικών έργων του υπουργείου άμυνας των Ηνωμένων Πολιτειών (DARPA) και το πρόγραμμα «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 στην ισχύ εκπομπής των δευτερογενών χρηστών αλλά καθορίζει πότε και πού μπορούν να εκπέμπουν. Συγκεκριμένα, σύμφωνα με την τεχνική της φασματικής επίστρωσης, ένας δευτερογενής χρήστης μπορεί να αποκτήσει πρόσβαση στο φάσμα όταν αυτό δεν χρησιμοποιείται από πρωτογενή χρήστη. Συνεπώς, προκειμένου να αποκτήσει πρόσβαση σε μια ζώνη συχνοτήτων, ο δευτερογενής χρήστης πρέπει να ανιχνεύει διαρκώς το φάσμα. Αν προσδιοριστεί κενό φάσματος, ο δευτερογενής χρήστης μπορεί να το χρησιμοποιήσει. Η τεχνική φασματικής επίστρωσης μπορεί να χρησιμοποιηθεί για δυναμική πρόσβαση φάσματος σε FDMA, TDMA ή OFDM ασύρματα συστήματα.

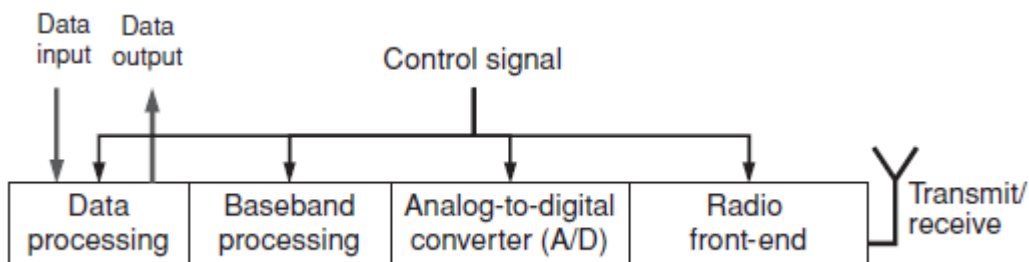
Συγκρινόμενο με τα μοντέλα της δυναμικής αποκλειστικής χρήσης και της ανοικτής ανταλλαγής, το μοντέλο ιεραρχικής πρόσβασης είναι ίσως η πλέον συμβατή προσέγγιση εφαρμογής της δυναμικής εκχώρησης συχνοτήτων, δεδομένης της τρέχουσας πολιτικής διαχείρισης του ραδιοφάσματος και της κληρονομιάς των ασύρματων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

1.3 Ραδιοεπικοινωνίες λογισμικού (Software-defined radio – SDR)

Ένα σύστημα ραδιοεπικοινωνιών λογισμικού (SDR) είναι ένα διαρκώς ρυθμιζόμενο σύστημα ασύρματης επικοινωνίας, στο οποίο οι παράμετροι εκπομπής (συχνότητα λειτουργίας, διαμόρφωση και πρωτόκολλο) ελέγχονται δυναμικά. Αυτή η προσαρμοστικότητα επιτυγχάνεται μέσω αλγορίθμων επεξεργασίας του σήματος που ελέγχονται από λογισμικό. Λόγω των χαρακτηριστικών τους, οι ραδιοεπικοινωνίες λογισμικού αποτέλεσαν τη βάση για την ανάπτυξη των γνωστικών ραδιοεπικοινωνιών. Οι βασικές λειτουργίες του SDR είναι [9]:

- Λειτουργία πολλαπλών ζωνών (multiband operation): οι ραδιοεπικοινωνίες λογισμικού υποστηρίζουν ασύρματη μετάδοση δεδομένων μέσω διαφορετικών ζωνών συχνοτήτων που χρησιμοποιούνται από διαφορετικά συστήματα ασύρματης πρόσβασης (π.χ. ζώνη κινητής τηλεφωνίας, ISM).
- Υποστήριξη πολλαπλών προτύπων (Multistandard support): υποστηρίζουν διαφορετικά πρότυπα όπως GSM, WCDMA, WiMAX, WiFi. Επίσης, υποστηρίζονται διαφορετικές διεπαφές (interfaces) στο πλαίσιο του ίδιου προτύπου (π.χ. IEEE 802.11a, 802.11b, 802.11g, ή 802.11n για το WiFi).
- Υποστήριξη πολλαπλών υπηρεσιών: υποστηρίζουν διάφορους τύπους υπηρεσιών, όπως, για παράδειγμα, υπηρεσίες κινητής τηλεφωνίας ή ασύρματου ευρυζωνικού internet.
- Πολυκαναλική υποστήριξη: μπορούν να λειτουργούν ταυτόχρονα σε διαφορετικές ζώνες συχνοτήτων.

Η γενική δομή ενός SDR πομποδέκτη φαίνεται στο σχήμα 1.11.



Σχήμα 1.11: Πομποδέκτης SDR

Παρά το γεγονός ότι τα περισσότερα μέρη των ραδιοεπικοινωνιών λογισμικού (π.χ. επεξεργασία σήματος, μετατροπή αναλογικού σε ψηφιακό και επεξεργασία βασικής ζώνης) είναι παρόμοια με αυτά που συναντώνται στους παραδοσιακούς πομποδέκτες, η διαφορά είναι ότι έκαστο εξ αυτών μπορεί να ελέγχεται από τα πρωτόκολλα των ανωτέρων στρωμάτων ή μπορεί να αναδιαρθρώνεται από τη μονάδα γνωστικής ραδιοεπικοινωνίας. Η λειτουργία έχει ως εξής: Αρχικά το αναλογικό σήμα λήψης διέρχεται από ένα ζωνοπερατό φίλτρο για να αποκτηθεί η επιθυμητή συχνότητα και ενισχύεται. Στη συνέχεια, παράγεται ένα σήμα ίδιας φάσης (in-phase) I και ένα ορθογωνιο σήμα (quadrature) με διαφορά φάσης $-\pi/2$ που μετατρέπονται σε ψηφιακά δεδομένα. Ο ρυθμός δειγματοληψίας για τη μετατροπή του αναλογικού σήματος σε ψηφιακό επιλέγεται ώστε να ικανοποιεί το κριτήριο Nyquist [10]. Ο ρυθμός δειγματοληψίας, οι παράμετροι των αναλογικών και ψηφιακών φίλτρων καθώς και οι αλγόριθμοι επεξεργασίας σήματος μπορούν να επαναπροσδιορίζονται σύμφωνα με τη συχνότητα λειτουργίας και τη διεπαφή της ασύρματης επικοινωνίας.

1.4 Γνωστικά συστήματα ραδιοεπικοινωνιών - Cognitive radios

1.4.1 Ορισμός

Οι γνωστικές ραδιοεπικοινωνίες (cognitive radios) αποτελούν εξέλιξη των ραδιοεπικοινωνιών λογισμικού και έχουν προταθεί ως μέσο αποδοτικότερης και αποτελεσματικότερης χρησιμοποίησης του φάσματος με στόχο την αξιοποίηση των αχρησιμοποίητων συχνοτήτων σε δυναμικά μεταβαλλόμενο περιβάλλον. Δηλαδή, προσφέρουν τη δυνατότητα ευκαιριακής κατάληψης ζωνών ραδιοσυχνοτήτων από δευτερεύοντες χρήστες στους οποίους δεν έχει εκχωρηθεί σταθερή ζώνη συχνοτήτων. Συνεπώς, οι γνωστικές ραδιοεπικοινωνίες αποτελούν μια νέα φιλοσοφία σχεδίασης που περιλαμβάνει έξυπνη ανίχνευση φασματικών κενών και, στη συνέχεια, καθορισμό και προσαρμογή των παραμέτρων μετάδοσης μιας ομάδας δευτερογενών χρηστών.

Οι όροι γνωστικές ραδιοεπικοινωνίες και συγκέντρωση φάσματος (spectrum pooling) εισήχθησαν αρχικά στην εργασία [11], όπου προτάθηκε ο εξής ορισμός: «ο όρος γνωστικές ραδιοεπικοινωνίες ταυτίζεται με το σημείο εκείνο όπου τα

ασύρματα PDAs (*personal digital assistants, προσωπικοί 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 να παρέχουν ραδιοπόρους και ασύρματες υπηρεσίες κατάλληλες να ικανοποιήσουν τις ανάγκες αυτές». Έτσι, λοιπόν, οι γνωστικές ραδιοεπικοινωνίες είναι ικανές να επιλέγουν αυτόματα την καλύτερη υπηρεσία και να καθυστερούν ή να προωθούν άμεσα ορισμένες ασύρματες μεταδόσεις σε αντιστοιχία με τους διαθέσιμους ή προβλεπόμενους πόρους.

Αργότερα, η ομοσπονδιακή επιτροπή επικοινωνιών των Ηνωμένων Πολιτειών (FCC – Federal Communication Commission) πρότεινε ένα λιγότερο γενικό ορισμό [12], όπου κάθε τηλεπικοινωνιακό σύστημα με ικανότητα μάθησης του φασματικού περιβάλλοντος και προσαρμογής σε αυτό αναφέρεται με τον όρο γνωστικό σύστημα ραδιοεπικοινωνιών. Ειδικότερα, «ως σύστημα γνωστικών ραδιοεπικοινωνιών καλείται κάθε σύστημα που μπορεί να αλλάξει τις παραμέτρους εκπομπής του βασιζόμενο στην αλληλεπίδραση με το περιβάλλον στο οποίο λειτουργεί. Η πλειοψηφία των γνωστικών συστημάτων ραδιοεπικοινωνιών βασίζεται σε σύστημα SDR, αλλά ούτε η ύπαρξη λογισμικού ούτε η ύπαρξη προγραμματισμού αποτελούν αναγκαίες απαιτήσεις των γνωστικών ραδιοεπικοινωνιών».

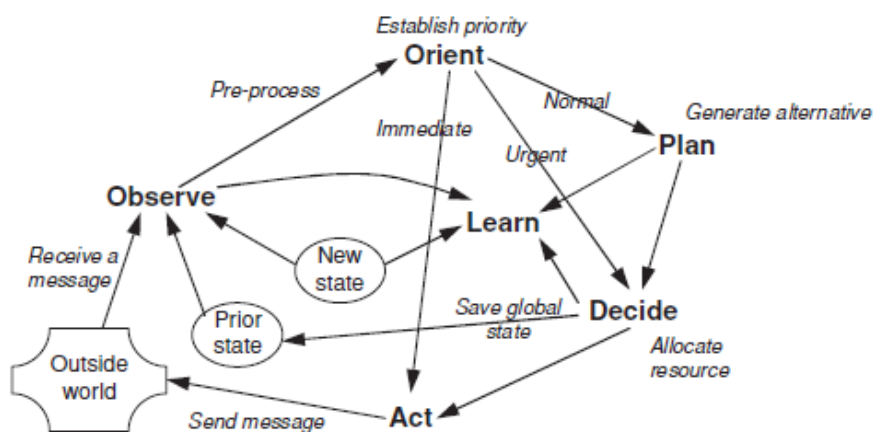
Ο Simon Haykin σε μια δημοσίευσή του [7] δίνει τον εξής ορισμό: « Σύστημα γνωστικών ραδιοεπικοινωνιών είναι ένα έξυπνο ασύρματο σύστημα επικοινωνιών που έχει γνώση του περιβάλλοντός του και χρησιμοποιεί τη μέθοδο “κατανόησης οικοδομώντας” (*understanding-by-building*) για να μάθει από το περιβάλλον και να προσαρμόσει τις εσωτερικές του λειτουργίες στις στατιστικές μεταβολές των εισερχόμενων RF διεγέρσεων, κάνοντας αντίστοιχες αλλαγές σε συγκεκριμένες λειτουργικές παραμέτρους (π.χ. στην ισχύ εκπομπής, τις φέρουσες συχνότητες και το σχήμα διαμόρφωσης) σε πραγματικό χρόνο και έχοντας δύο κύριους στόχους:

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

Τέλος, το 2007 η ομάδα εργασίας IEEE 1900.1, που δημιουργήθηκε για το καθορισμό ορολογιών και εννοιών, πρότεινε τον επόμενο ορισμό [9] για τις γνωστικές ραδιοεπικοινωνίες: «ένα είδος ασύρματου συστήματος ραδιοεπικοινωνιών που μπορεί να ανιχνεύσει και να κρίνει αυτόνομα το περιβάλλον του και προσαρμόζεται ανάλογα σε αυτό. Το ασύρματο αυτό σύστημα επικοινωνιών θα μπορούσε να χρησιμοποιεί αναπαράσταση γνώσης, αυτοματοποιημένη κρίση και μηχανισμό μάθησης για την εγκατάσταση, πραγματοποίηση και τερματισμό επικοινωνίας με άλλα ασύρματα συστήματα επικοινωνιών. Τα γνωστικά συστήματα ραδιοεπικοινωνιών μπορούν δυναμικά και αυτόνομα να προσαρμόζουν τις λειτουργικές παραμέτρους τους».

1.4.2 Λειτουργίες συστημάτων γνωστικών ραδιοεπικοινωνιών

Οι βασικές λειτουργίες των γνωστικών ραδιοεπικοινωνιών που απαιτούνται για την προσαρμογή των παραμέτρων μετάδοσης στο συνεχώς μεταβαλλόμενο ραδιοεπικοινωνιακό περιβάλλον μπορούν να συνοψιστούν σε ένα κύκλο γνώσης (cognitive cycle), όπως αυτός παρουσιάστηκε από τον J.Mitola και φαίνεται στο σχήμα 1.12.



Σχήμα 1.12: Γνωστικός κύκλος

Όπως φαίνεται από το σχήμα 1.12 σύμφωνα με τον κύκλο γνώσης, το τηλεπικοινωνιακό σύστημα αποκτά πληροφορίες για το εξωτερικό περιβάλλον (Outside world) μέσω άμεσων παρατηρήσεων ή μέσω σηματοδότησης. Στη συνέχεια,

οι πληροφορίες αυτές εκτιμώνται (Orient) ώστε να διαπιστωθεί η σπουδαιότητά τους. Με την αποτίμηση των πληροφοριών γίνεται ο σχεδιασμός των εναλλακτικών στρατηγικών (Plan) και, στη συνέχεια, με κριτήριο τη βελτιστοποίηση του τελικού στόχου, επιλέγεται (Decide) και εκτελείται (Act) η κατάλληλη ενέργεια. Τα αποτελέσματα των ενεργειών αυτών αντικατοπτρίζονται στις επιδόσεις του συστήματος και στις ενδεχόμενες ανεπιθύμητες παρεμβολές που παρουσιάζονται στο εξωτερικό περιβάλλον. Τέλος, το ασύρματο σύστημα επικοινωνιών χρησιμοποιεί τα αποτελέσματα και τις εκάστοτε αποφάσεις κατά τη διαδικασία μάθησης (Learn), με σκοπό τη βελτίωση της λειτουργίας του, δημιουργώντας νέες μοντελοποιημένες καταστάσεις και παράγοντας νέες εναλλακτικές στρατηγικές που προσαρμόζονται καλύτερα στις συνθήκες του περιβάλλοντος λειτουργίας.

Οι βασικές λειτουργίες ενός συστήματος γνωστικών ραδιοεπικοινωνιών είναι οι εξής:

- *Ανίχνευση φάσματος (spectrum sensing)*: Στόχος της ανίχνευσης φάσματος είναι ο καθορισμός της κατάστασης του ραδιοφάσματος και της δραστηριότητας των αδειοδοτημένων χρηστών μέσω τις περιοδικής ανίχνευσης της επιθυμητής ζώνης συχνοτήτων. Συγκεκριμένα, ο γνωστικός πομποδέκτης εντοπίζει το αχρησιμοποίητο φάσμα (φασματική τρύπα – spectrum hole) και καθορίζει τον τρόπο πρόσβασης σε αυτό (ισχύς εκπομπής, διάρκεια πρόσβασης) χωρίς να παρεμβάλλει την εκπομπή ενός αδειοδοτημένου χρήστη.

Η ανίχνευση φάσματος μπορεί να είναι είτε συγκεντρωτική (centralized) είτε κατακεκομημένη (distributed). Στη συγκεντρωτική ανίχνευση φάσματος, ένας ελεγκτής ανίχνευσης (σημείο πρόσβασης ή σταθμός βάσης) ανιχνεύει την επιθυμητή ζώνη συχνοτήτων και, στη συνέχεια, μοιράζεται την πληροφορία που αποκτά με άλλους κόμβους του δικτύου. Με τη συγκεντρωτική ανίχνευση μειώνεται η πολυπλοκότητα των τερματικών των χρηστών καθώς όλες οι λειτουργίες ανίχνευσης πραγματοποιούνται στον ελεγκτή ανίχνευσης (sensing controller). Ωστόσο, αντιμετωπίζει πρόβλημα λόγω της ποικιλομορφίας των περιοχών. Για παράδειγμα, ο ελεγκτής ανίχνευσης είναι πιθανό να μην μπορέσει να εντοπίσει ένα μη

αδειοδοτημένο χρήστη αν αυτός βρίσκεται στα άκρα της περιοχής κάλυψης. Κατά την κατανομημένη ανίχνευση φάσματος οι μη αδειοδοτημένοι χρήστες ανιχνεύουν το φάσμα ανεξάρτητα και τα αποτελέσματα της ανίχνευσης μπορούν χρησιμοποιηθούν ατομικά από τους χρήστες (μη συνεργατική ανίχνευση – non cooperative sensing) ή να γνωστοποιηθούν και σε άλλους δευτερεύοντες χρήστες (συνεργατική ανίχνευση – cooperative sensing). Μολονότι η συνεργατική ανίχνευση επιφέρει ένα επιπλέον κόστος επικοινωνίας και επεξεργασίας, παρέχει μεγαλύτερη ακρίβεια ανίχνευσης φάσματος από τη μη συνεργατική ανίχνευση.

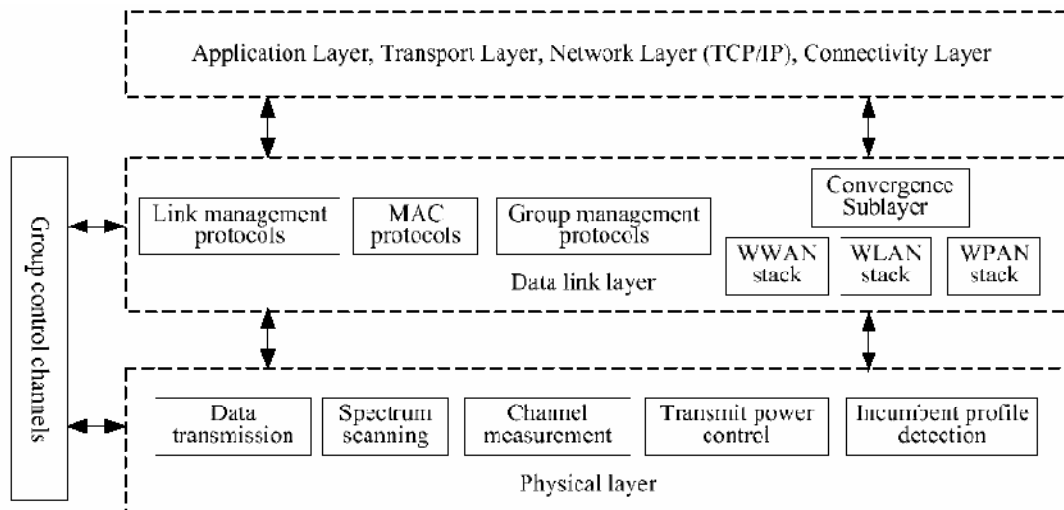
- *Ανάλυση φάσματος (spectrum analysis)*: Η πληροφορία που αποκτάται μέσω της φασματικής ανίχνευσης χρησιμοποιείται για τον προγραμματισμό και την οργάνωση της πρόσβασης στο φάσμα από τους δευτερογενείς χρήστες. Επίσης, λαμβάνονται υπόψη οι απαιτήσεις επικοινωνίας των δευτερογενών χρηστών προκειμένου να βελτιστοποιηθούν οι παράμετροι εκπομπής. Κατά την ανάλυση φάσματος η πληροφορία από την ανίχνευση αρχικά χρησιμοποιείται για την απόκτηση γνώσης σχετικά με τα φασματικά κενά (εκτίμηση παρεμβολής, διάρκεια διαθεσιμότητας, πιθανότητα σύγκρουσης λόγω λανθασμένης ανίχνευσης). Στη συνέχεια, λαμβάνεται η απόφαση για πρόσβαση στο φάσμα (συχνότητα λειτουργίας, εύρος ζώνης, διαμόρφωση, ισχύς εκπομπής, τοποθεσία, χρονική διάρκεια) με στόχο τη βέλτιστη λειτουργία του συστήματος δεδομένων των επιθυμητών στόχων (π.χ. μεγιστοποίηση της ρυθμαπόδοσης – throughput) υπό διάφορους περιορισμούς (π.χ. διατήρηση του επιπέδου παρεμβολής στους αδειοδοτημένους χρήστες κάτω από ένα συγκεκριμένο επίπεδο)
- *Πρόσβαση στο φάσμα (spectrum access)*: Αφού ληφθεί η απόφαση για πρόσβαση στο φάσμα, οι μη αδειοδοτημένοι χρήστες επιχειρούν πρόσβαση στο αχρησιμοποίητο φάσμα. Υπεύθυνο για την πρόσβαση στο φάσμα είναι ένα γνωστικό πρωτόκολλο πρόσβασης μέσου (cognitive MAC protocol), που αποσκοπεί στην αποφυγή συγκρούσεων τόσο με τους πρωτογενείς όσο και με τους δευτερογενείς χρήστες. Ο πομπός είναι επίσης υπεύθυνος για την διαπραγμάτευση και το συγχρονισμό με τον παραλήπτη προκειμένου να

εξασφαλιστεί η επιτυχής μετάδοση και λήψη των δεδομένων. Η λειτουργία ενός γνωστικού πρωτόκολλου MAC μπορεί να βασίζεται είτε σε MAC πρωτόκολλα σταθερής εκχώρησης (π.χ. FDMA, TDMA, CDMA) είτε σε MAC πρωτόκολλα τυχαίας πρόσβασης (π.χ. ALOHA, CSMA/CA).

- *Κινητικότητα φάσματος (spectrum mobility):* Η κινητικότητα φάσματος είναι λειτουργία που σχετίζεται με την αλλαγή της ζώνης συχνοτήτων λειτουργίας των cognitive χρηστών. Όταν ένας αδειοδοτημένος χρήστης ξεκινά πρόσβαση σε ένα κανάλι που χρησιμοποιείται από ένα μη αδειοδοτημένο χρήστη, ο τελευταίος μεταπηδά σε άλλο κανάλι που εκείνη τη στιγμή είναι αχρησιμοποίητο. Κατά τη διάρκεια της αλλαγής συχνότητας λειτουργίας, οι παράμετροι πρωτοκόλλων στις διάφορες βαθμίδες του OSI πρέπει να προσαρμόσουν τη λειτουργία τους στις απαιτήσεις της νέας ζώνης συχνοτήτων. Κατά τη μετάβαση αυτή, πρέπει να είναι εγγυημένο με κάποια πιθανότητα ότι η μετάδοση δεδομένων θα συνεχιστεί και στη νέα ζώνη συχνοτήτων.

1.4.3 Αρχιτεκτονική πρωτοκόλλων για συστήματα γνωστικών ραδιοεπικοινωνιών

Η διαχείριση του φάσματος σχετίζεται πρωτίστως με τα δύο κατώτερα επίπεδα πρωτοκόλλων: το φυσικό επίπεδο και το επίπεδο ζεύξης δεδομένων. Τα ανώτερα επίπεδα σχετίζονται με ειδικότερες εφαρμογές και βελτιστοποίηση άλλων παραμέτρων του δικτύου π.χ. προσαρμογή και εύρεση βέλτιστης δρομολόγησης. Στο σχήμα 1.13 παρουσιάζονται γραφικά οι λειτουργίες των συστημάτων γνωστικών ραδιοεπικοινωνιών.



Σχήμα 1.13: Ιεραρχία πρωτοκόλλων λειτουργίας σε Cognitive Radio

Όπως παρατηρείται το φυσικό στρώμα περιλαμβάνει τις διαδικασίες:

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

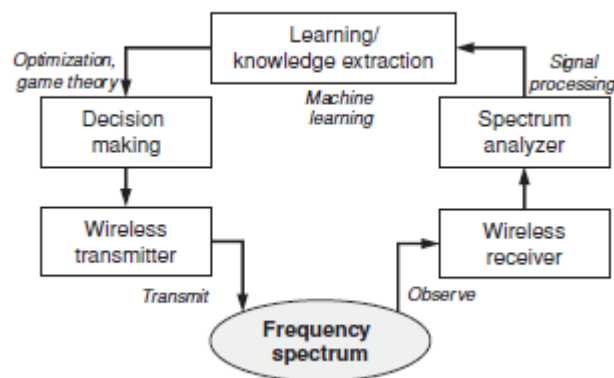
Όσον αφορά το στρώμα ζεύξης δεδομένων, περιλαμβάνονται:

- *Πρωτόκολλα Διαχείρισης Ζεύξης (Link Management Protocols)* που είναι υπεύθυνα για τη ζεύξη μεταξύ δύο χρηστών τεχνολογίας “Cognitive radio”.
- *Πρωτόκολλα MAC* που συμβάλλουν στην κατάλληλη επιλογή συχνοτήτων για την ομαλή επικοινωνία των κόμβων του δικτύου και την επίλυση προβλημάτων κρυμμένου ή εκτεθειμένου τερματικού.
- *Πρωτόκολλα Διαχείρισης Ομάδων (Group Management Protocols)* για το συντονισμό των χρηστών που ανήκουν στην ίδια ομάδα (υποδίκτυο).

- Υποστρώμα Σύγκλισης (Convergence Sublayer) που παρέχει τη δυνατότητα στο γνωστικό σύστημα να λειτουργεί σε εντελώς διαφορετικά ασύρματα περιβάλλοντα όπως π.χ. Ασύρματα Τοπικά Δίκτυα (WLANs), Ασύρματα Προσωπικά Δίκτυα (WPANs) και Ασύρματα Δίκτυα Ευρείας Περιοχής (WWANs).

1.4.5 Βασικά στοιχεία συστημάτων γνωστικών ραδιοεπικοινωνιών

Τα βασικά μέρη ενός γνωστικού πομποδέκτη που εκτελούν τις λειτουργίες που αναφέρθηκαν προηγουμένως φαίνονται στο σχήμα 1.14.



Σχήμα 1.14: Στοιχεία γνωστικού πομποδέκτη

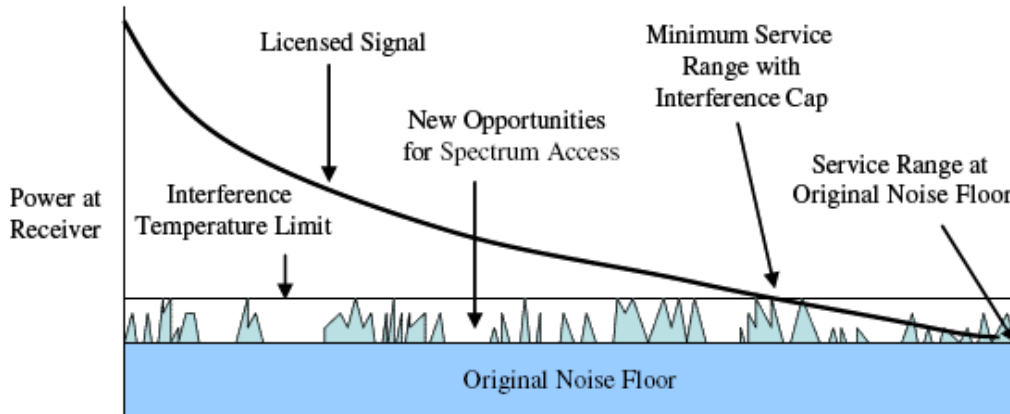
- Πομπός/δέκτης (transmitter/receiver): Βασικό στοιχείο ενός συστήματος γνωστικών ραδιοεπικοινωνιών είναι ένας πομποδέκτης ραδιοεπικοινωνιών λογισμικού (software defined radio transceiver) που χρησιμοποιείται για τη μετάδοση και λήψη των σημάτων δεδομένων. Οι παράμετροι λειτουργίας του πομποδέκτη μπορούν να μεταβάλλονται δυναμικά σύμφωνα με τις υποδείξεις των πρωτοκόλλων των ανώτερων στρωμάτων. Επιπλέον, συχνά χρησιμοποιείται και ένας ακόμα δέκτης ο οποίος παρακολουθεί τη δραστηριότητα στο φάσμα συχνοτήτων (ανίχνευση φάσματος).
- Αναλυτής φάσματος: Ο αναλυτής φάσματος χρησιμοποιεί τα ανιχνευμένα σήματα για να εξαγάγει συμπεράσματα ως προς τη χρήση του φάσματος. Για παράδειγμα, ο αναλυτής φάσματος είναι υπεύθυνος για τον εντοπισμό του

αποτυπώματος ενός σήματος που ανήκει σε αδειοδοτημένο χρήστη καθώς και για την εύρεση κενών φάσματος. Ο αναλυτής φάσματος πρέπει να εξασφαλίζει ότι η μετάδοση ενός αδειοδοτημένου χρήστη δεν θα παρεμβάλλεται από ένα μη αδειοδοτημένο χρήστη που αποφασίζει να χρησιμοποιήσει το φάσμα.

- Εξαγωγή γνώσης/ εκμάθηση (Knowledge extraction/learning): Οι λειτουργίες της μάθησης και της εξαγωγής γνώσης χρησιμοποιούν την πληροφορία από τη φασματική χρήση προκειμένου να κατανοήσουν το RF περιβάλλον (π.χ. τη συμπεριφορά των αδειοδοτημένων χρηστών). Στο πλαίσιο αυτό δημιουργείται και διατηρείται μια γνωστική βάση δεδομένων που χρησιμοποιείται για τη βελτιστοποίηση και προσαρμογή των παραμέτρων μετάδοσης προκειμένου να επιτευχθούν οι επιθυμητοί στόχοι δεδομένων των περιορισμών του περιβάλλοντος. Για αποδοτικότερη μάθηση μπορούν να χρησιμοποιηθούν αλγόριθμοι ενισχυτικής μάθησης (reinforcement learning algorithms) από τον τομέα της τεχνικής νοημοσύνης όπως θα δούμε στη συνέχεια (π.χ. Q-learning, value iteration, policy iteration αλγόριθμοι).
- Διαδικασία απόφασης (Decision making): Αφότου η γνώση για τη χρήση του φάσματος είναι διαθέσιμη, λαμβάνεται η απόφαση για πρόσβαση ή όχι στο φάσμα. Η βέλτιστη απόφαση εξαρτάται από το περιβάλλον, που σημαίνει ότι εξαρτάται από τη συνεργατική ή ανταγωνιστική συμπεριφορά των μη αδειοδοτημένων χρηστών. Για την εύρεση της βέλτιστης πολιτικής μπορούν να χρησιμοποιηθούν διάφορες τεχνικές. Για παράδειγμα, η θεωρία της βελτιστοποίησης μπορεί να χρησιμοποιηθεί όταν το σύστημα μπορεί να περιγραφεί ως μία οντότητα με ένα κοινό σκοπό. Αντίθετα, μοντέλα από τη θεωρία παιγνίων μπορούν να χρησιμοποιηθούν όταν το σύστημα αποτελείται από πολλαπλές οντότητες εκάστη των οποίων έχει το δικό της σκοπό. Τέλος, μπορεί να εφαρμοστεί στοχαστική βελτιστοποίηση όταν οι καταστάσεις του συστήματος είναι τυχαίες.

1.5 Θερμοκρασία παρεμβολής

Η υπηρεσία πολιτικής φάσματος της FCC πρόσφατα πρότεινε μία νέα μέθοδο μέτρησης των παρεμβολών[13]. Στο πλαίσιο αυτό εισήγαγε και ένα νέο μέτρο παρεμβολής, τη θερμοκρασία παρεμβολής I_T για την ποσοτικοποίηση και τη διαχείριση των πηγών παρεμβολής όπως είναι, για παράδειγμα, οι εκπομπές από μη αδειοδοτημένους χρήστες (σχήμα 1.15).



Σχήμα 1.15: Μοντέλο θερμοκρασίας παρεμβολής

Με βάση το μέτρο αυτό, το όριο θερμοκρασίας παρεμβολής I_T χρησιμοποιείται για την παρεμβολή που προκαλείται στους αδειοδοτημένους χρήστες σε συγκεκριμένη ζώνη συχνοτήτων σε μια συγκεκριμένη περιοχή. Η έννοια της θερμοκρασίας παρεμβολής είναι παρόμοια με αυτή της θερμοκρασίας θορύβου, αλλά έχει επεκταθεί ώστε να περιλαμβάνει τόσο το θόρυβο όσο και τη ντετερμινιστική παρεμβολή από άλλες πηγές σήματος. Η θερμοκρασία παρεμβολής μετράται στο δέκτη και αντικατοπτρίζει την ποιότητα της λήψης σήματος. Επίσης, προσφέρει ακριβή και αποτελεσματική πληροφορία σχετικά με την παρεμβολή σε μια ζώνη συχνοτήτων ώστε ένας μη αδειοδοτημένος χρήστης να μπορεί να αξιολογήσει πόσο εφικτή είναι η πρόσβαση σε αυτή.

Η θερμοκρασία παρεμβολής μετράται σε βαθμούς Kelvin και μπορεί να εκφραστεί μέσω της σχέσης:

$$I_T(f_c, W) = \frac{P_I(f_c, W)}{KW}, \quad (1.1)$$

όπου $P_I(f_c, W)$ είναι η μέση ισχύς παρεμβολής σε εύρος W περί τη συχνότητα f_c και K η σταθερά Boltzmann ($K = 1.38 \times 10^{-23} \text{W/Hz/}^\circ\text{K}$)

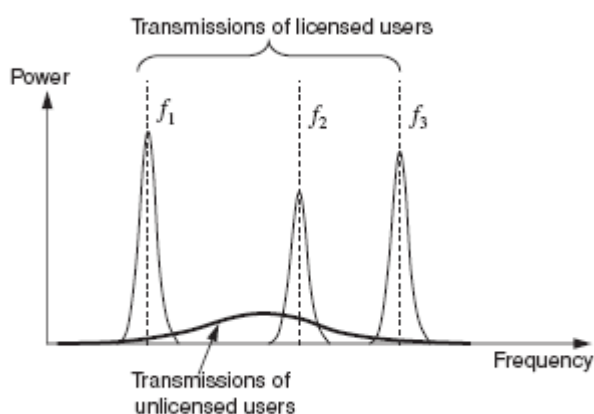
Στη συνέχεια, αναλύονται δύο μοντέλα θερμοκρασίας παρεμβολής, το ιδανικό και το γενικό.

1.5.1 Ιδανικό μοντέλο θερμοκρασίας παρεμβολής

Έστω P η ισχύς εκπομπής ενός μη αδειοδοτημένου χρήστη σε συγκεκριμένη ζώνη συχνοτήτων όπου λειτουργούν n αδειοδοτημένοι χρήστες και W_i το εύρος ζώνης του χρήστη i περί τη συχνότητα f_i . Η εκπομπή ενός μη αδειοδοτημένου χρήστη πρέπει να εγγυάται το ακόλουθο όριο θερμοκρασίας παρεμβολής για τον αδειοδοτημένο δέκτη [15]:

$$I_T(f_i, W_i) + \frac{M_i P}{K W_i} \leq I_L(f_i) \quad i = 1, 2, \dots, n \quad (1.2)$$

Όπου $0 \leq M_i \leq 1$ ένας παράγοντας εξασθένησης που οφείλεται στις απώλειες διάδοσης και σκίασης του διαύλου επικοινωνίας μεταξύ του μη αδειοδοτημένου πομπού και του αδειοδοτημένου δέκτη. Ένα παράδειγμα για $n = 3$ φαίνεται στο σχήμα 1.16.



Σχήμα 1.16: Θερμοκρασία παρεμβολής σε αδειοδοτημένους χρήστες

Σε αυτό το ιδανικό μοντέλο θερμοκρασίας παρεμβολής, το σήμα από ένα μη αδειοδοτημένο χρήστη πρέπει να διαχωρίζεται από το σήμα ενός αδειοδοτημένου χρήστη. Επιπλέον, η μέτρηση της παρεμβολής βασίζεται στην παρουσία ενός

αδειοδοτημένου σήματος. Για το λόγο αυτό πρέπει να είναι γνωστή η κυματομορφή του σήματος των αδειοδοτημένων χρηστών για τη μέτρηση της παρεμβολής τη σωστή χρονική στιγμή. Συγκεκριμένα, η παρεμβολή πρέπει να μετράται όταν ένας αδειοδοτημένος χρήστης εκπέμπει. Αν η παρεμβολή μετρηθεί όταν δεν εκπέμπεται σήμα από κάποιο αδειοδοτημένο χρήστη, η θερμοκρασία παρεμβολής θα είναι λάθος.

1.5.2 Γενικευμένο μοντέλο θερμοκρασίας παρεμβολής

Στο γενικευμένο μοντέλο θερμοκρασίας παρεμβολής η πληροφορία για το σήμα των αδειοδοτημένων χρηστών δεν είναι διαθέσιμη· συνεπώς, το σήμα τους δεν μπορεί να διαχωριστεί από αυτό των μη αδειοδοτημένων χρηστών. Το γενικευμένο μοντέλο ορίζει την παρεμβολή για το συνολικό εύρος συχνοτήτων W και ο περιορισμός για το όριο της θερμοκρασίας παρεμβολής μπορεί να οριστεί ως εξής [15]:

$$I_T(f_c, W) + \frac{MP}{KW} \leq I_L(f_c) \quad (1.3)$$

Σε αυτήν την περίπτωση ο περιορισμός εκφράζεται σε όρους f_c και W , όπου f_c είναι η κεντρική συχνότητα και W το εύρος ζώνης που χρησιμοποιεί ο μη αδειοδοτημένος χρήστης.

1.5.3 Χωρητικότητα διαύλου κάτω από το όριο παρεμβολής

Το όριο θερμοκρασίας παρεμβολής ορίζει το άνω φράγμα για την ισχύ εκπομπής ενός μη αδειοδοτημένου χρήστη σε συγκεκριμένη ζώνη συχνοτήτων. Δεδομένου αυτού του άνω φράγματος η χωρητικότητα διαύλου (spectrum capacity) r για ένα μη αδειοδοτημένο χρήστη είναι:

$$r = W \log \left(1 + \frac{S}{\sigma^2 + P_1} \right), \quad (1.4)$$

όπου W είναι το εύρος της ζώνης συχνοτήτων, S η ισχύς του σήματος λήψης του μη αδειοδοτημένου χρήστη, σ^2 η ισχύς του θορύβου και P_1 η ισχύς παρεμβολής στο

δέκτη του μη αδειοδοτημένου χρήστη εξαιτίας της εκπομπής του αδειοδοτημένου χρήστη [16].

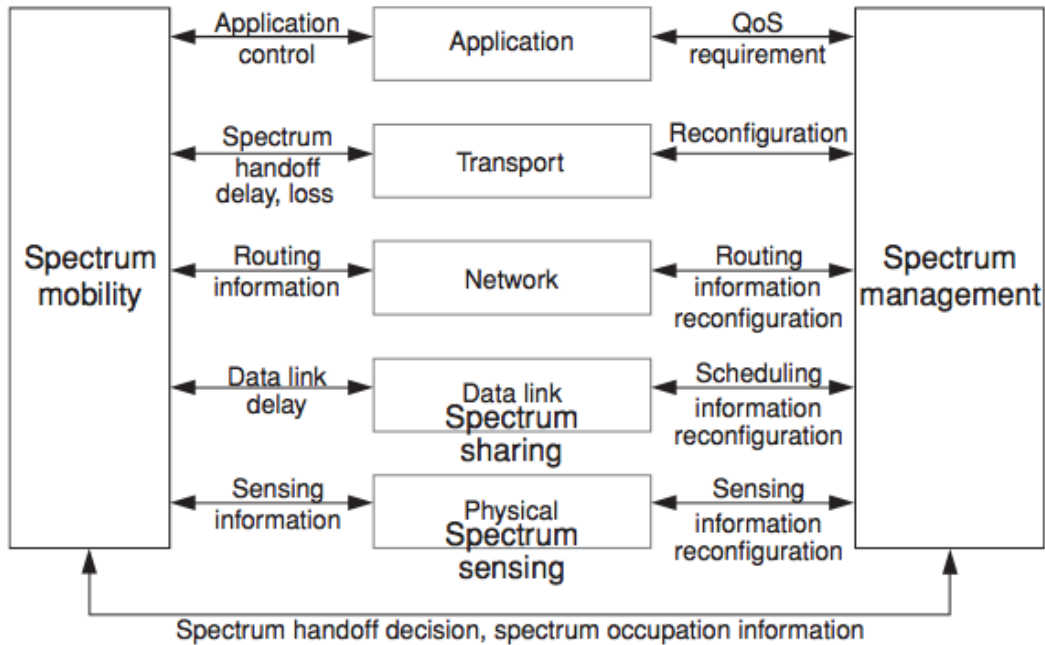
1.6 Αρχιτεκτονικές γνωστικών ραδιοεπικοινωνιών για συστήματα επόμενης γενιάς

Προκειμένου να βελτιώσει την υπάρχουσα αξιοποίηση του φάσματος, ο οργανισμός προηγμένων ερευνητικών έργων του υπουργείου άμυνας των Ηνωμένων Πολιτειών (DARPA-Defense Advanced Research Projects Agency) έθεσε σε λειτουργία το “πρόγραμμα επόμενης γενιάς” (Next Generation program-XG). Η βασική ιδέα είναι η χρήση του μηχανισμού δυναμικής πρόσβασης στο φάσμα μέσω γνωστικών ραδιοεπικοινωνιών για XG δίκτυα. Σε ένα τέτοιο δίκτυο, ένας χρήστης γνωστικού συστήματος πρέπει να είναι σε θέση να ανιχνεύει το διαθέσιμο φάσμα καθώς και να εντοπίζει την παρουσία αδειοδοτημένων χρηστών στην επιθυμητή ζώνη συχνοτήτων ώστε να απελευθερώνει το φάσμα όταν εντοπίζονται αδειοδοτημένοι χρήστες. Οι λειτουργίες αυτές αναφέρονται ως ανίχνευση φάσματος, διαχείριση φάσματος, καταμερισμός φάσματος και κινητικότητα φάσματος, αντίστοιχα. Οι λειτουργίες αυτές και η ένταξή τους στη στοίβα πρωτοκόλλων ενός δικτύου XG φαίνονται στο σχήμα 1.17.

Το σχήμα 1.18 δείχνει τη γενική αρχιτεκτονική ενός XG δικτύου. Στην αρχιτεκτονική αυτή υπάρχουν δύο βασικές ομάδες χρηστών, οι αδειοδοτημένοι και οι μη αδειοδοτημένοι χρήστες που ανήκουν στο πρωτεύον και το XG δίκτυο, αντίστοιχα.

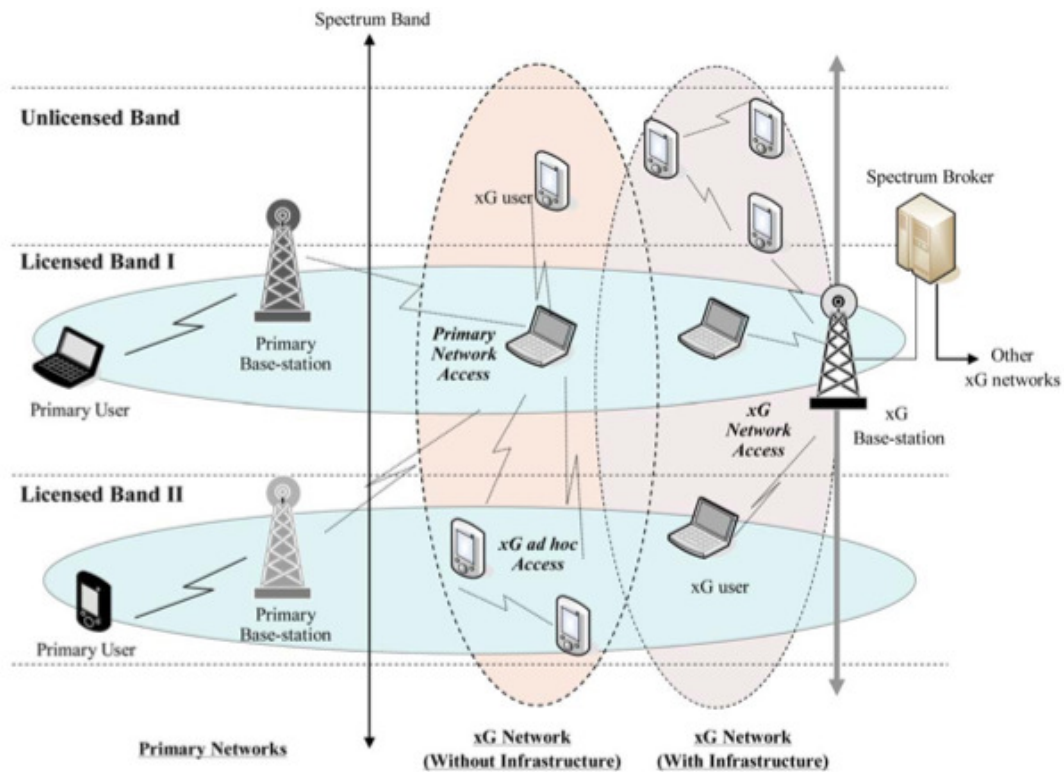
Το πρωτεύον δίκτυο περιλαμβάνει την υπάρχουσα υποδομή και κατέχει αποκλειστικό δικαίωμα πρόσβασης σε συγκεκριμένη ζώνη συχνοτήτων. Παραδείγματα τέτοιων δικτύων αποτελούν τα δίκτυα κινητής τηλεφωνίας και ψηφιακής τηλεόρασης. Το πρωτεύον δίκτυο αποτελείται από τους πρωτεύοντες (αδειοδοτημένους) χρήστες και τον πρωτεύοντα σταθμό βάσης. Οι πρωτεύοντες χρήστες έχουν το αποκλειστικό δικαίωμα χρήσης μιας ζώνης συχνοτήτων. Η πρόσβαση στο φάσμα ελέγχεται από τον πρωτεύοντα σταθμό βάσης και δεν πρέπει να επηρεάζεται από τη λειτουργία οποιουδήποτε μη αδειοδοτημένου χρήστη.

Επιπλέον, το πρωτεύον δίκτυο δεν χρειάζεται τροποποιήσεις στον τρόπο λειτουργίας του για να συνυπάρξει με ένα XG δίκτυο. Ο πρωτεύων σταθμός βάσης χρησιμοποιείται για τον έλεγχο της πρόσβασης των αδειοδοτημένων χρηστών χωρίς να έχει δυνατότητα παροχής φάσματος στους XG χρήστες.



Σχήμα 1.17: Στοίβα πρωτοκόλλων δικτύων νέας γενιάς

Το XG δίκτυο ή αλλιώς μη αδειοδοτημένο δίκτυο δεν έχει δικαίωμα λειτουργίας σε κάποια ζώνη συχνοτήτων, οπότε η πρόσβαση στο φάσμα πραγματοποιείται κατά ευκαιριακό τρόπο. Το μη αδειοδοτημένο σύστημα μπορεί να είναι είτε δομημένο είτε αδόμητο. Στην πρώτη περίπτωση ένας XG σταθμός βάσης χρησιμοποιείται για τον έλεγχο της πρόσβασης των μη αδειοδοτημένων χρηστών. Στην περίπτωση αδόμητου συστήματος δεν υπάρχει κεντρικός ελεγκτής και οι μη αδειοδοτημένοι χρήστες αποφασίζουν για πρόσβαση στο φάσμα βασισμένοι μόνο σε τοπική πληροφορία.



Σχήμα 1.18: Αρχιτεκτονική δικτύου νέας γενιάς

Στη συνέχεια, παρατίθενται μερικά από τα πλέον αντιπροσωπευτικά παραδείγματα αρχιτεκτονικών δικτύων επόμενης γενιάς:

CORVUS: Είναι μια προσέγγιση γνωστικών επικοινωνιών για τη χρήση μη αδειοδοτημένου φάσματος [17], σύμφωνα με την οποία, ανάλογα με την τοπική ανίχνευση φάσματος, ο εντοπισμός αδειοδοτημένων χρηστών και η κατανομή φάσματος πραγματοποιούνται συνεργατικά. Για να επιτευχθεί η συνεργασία μεταξύ τους, οι δευτερογενείς χρήστες οργανώνονται σε ομάδες. Κάθε μέλος μιας ομάδας ανιχνεύει ένα τμήμα φάσματος που διαιρείται σε υποκανάλια. Ένα κοινό κανάλι ελέγχου χρησιμοποιείται από όλες τις ομάδες για συντονισμό και ξεχωριστά κανάλια ελέγχου από τα μέλη κάθε ομάδας για την ανταλλαγή της πληροφορίας ανίχνευσης και την καθιέρωση δευτερογενών διαύλων επικοινωνίας.

DIMSUNnet: Προτάθηκε ως μια ευφυής αρχιτεκτονική δυναμικής διαχείρισης φάσματος για πανταχού κινητά δίκτυα [18]. Για τη βελτίωση της αξιοποίησης φάσματος, το DIMSUMnet χρησιμοποιεί στατιστικά πολυπλεγμένη

πρόσβαση(statistically multiplexed access-SMA), ενώ για τη βελτίωση της αποδοτικότητας και της δικαιοσύνης έχει εισαχθεί η έννοια της συντονισμένης ζώνης πρόσβασης (coordinated access band-CAB). Η συντονισμένη ζώνη πρόσβασης αποτελείται από ένα τμήμα φάσματος που έχει παραχωρηθεί σε ρυθμιστικές αρχές. Ένας μεσίτης φάσματος(spectrum broker) κατέχει το φάσμα αυτό και το μισθώνει ανάλογα με τη ζήτηση.

DRiVE/OverDRiVE: Το Ευρωπαϊκό πρόγραμμα DRiVE (Dynamic Radio for IP Services in Vehicular Environments), που ξεκίνησε τον Απρίλιο του 2000 υπό τη συνεργασία δεκαπέντε κορυφαίων εταιριών των τηλεπικοινωνιών, των μέσων ενημέρωσης και της αυτοκινητοβιομηχανίας, έχει ως στόχο τη δυναμική κατανομή φάσματος σε ετερογενή δίκτυα (GSM, GPRS, UMTS, DAB, DVB-T) θεωρώντας ένα κοινό κανάλι συντονισμού[19]. Με βάση χωρικά και χρονικά στατιστικά στοιχεία κίνησης, βελτιώνεται η αποτελεσματικότητα εκχώρησης φάσματος (π.χ. το φάσμα που κατανέμεται σε συστήματα UMTS και DVB-T διαφέρει από περιοχή σε περιοχή εξαρτάται από το χρόνο), δηλαδή χρησιμοποιείται το μοντέλο δυναμικής αποκλειστικής χρήσης. Ερευνώνται δύο πλευρές της δυναμικής πρόσβασης φάσματος, η χρονική και η χωρική. Στην περίπτωση της χρονικής δυναμικής πρόσβασης στο φάσμα ένα δίκτυο ασύρματης πρόσβασης μπορεί να χρησιμοποιήσει το φάσμα όταν αυτό δεν χρησιμοποιείται από κάποιο άλλο δίκτυο. Αντίθετα, η χωρική δυναμική πρόσβαση επιτρέπει την προσαρμογή της κατανομής φάσματος ανάλογα με τις τοπικές εναλλαγές στις απαιτήσεις τηλεπικοινωνιακής κίνησης. Παρά το γεγονός ότι τα μοντέλα αυτά έδειξαν αξιοσημείωτες δυνατότητες για αύξηση της φασματικής αξιοποίησης, η εφαρμογή του συστήματος αναδιαμόρφωσης για χρονική και χωρική δυναμική πρόσβαση παραμένει ακόμα μεγάλη πρόκληση.

1.7 Πρότυπα γνωστικών επικοινωνιών

1.7.1 IEEE SCC 41

Η ομάδα IEEE SCC 41 δημιουργήθηκε για την εξέταση των θεμάτων σχετικά με την ανάπτυξη, την εφαρμογή και την εξέλιξη των ασύρματων δικτύων επόμενης

γενιάς μέσω μιας σειράς σχετικών προτύπων που ονομάστηκε IEEE 1900 [21]. Η IEEE SCC αποτελείται από τέσσερις ομάδες εργασίας και μία ομάδα μελέτης. Εκάστη των ομάδων αυτών είναι υπεύθυνη για την έναρξη των διαδικασιών προτυποποίησης για τα διάφορα θέματα ενός συστήματος γνωστικών ραδιοεπικοινωνιών. Τα βασικά πρότυπα της σειράς IEEE 1900 είναι τα εξής:

- IEEE 1900.1: Κύριος στόχος του προτύπου αυτού είναι η αναγνώριση και η επεξήγηση ενός λεξιλογίου όρων και εννοιών που σχετίζονται με τη διαχείριση φάσματος, τις ραδιοεπικοινωνίες λογισμικού, τις προσαρμοστικές ραδιοεπικοινωνίες και άλλες σχετικές τεχνολογίες. Επίσης, περιγράφει την αλληλοσυσχέτιση μεταξύ όρων και εννοιών. Το πρότυπο IEEE 1900.1 χρησιμοποιείται ως σύνδεση με τις υπόλοιπες ομάδες εργασίας της IEEE SSC 41 καθώς όλες οι ομάδες εργασίας θα στηρίζονται στους ορισμούς του προτύπου αυτού.
- IEEE 1900.2: Το πρότυπο αυτό σχετίζεται με την παρεμβολή και τη συνύπαρξη συστημάτων γνωστικών ραδιοεπικοινωνιών. Στις γνωστικές ραδιοεπικοινωνίες πολλές ασύρματες συσκευές και υπηρεσίες επιτρέπεται να συνυπάρχουν στην ίδια περιοχή την ίδια χρονική στιγμή. Η βελτιστοποίηση των παραμέτρων λειτουργίας των συσκευών και υπηρεσιών αυτών αποτελεί κρίσιμη συνιστώσα για την αποφυγή των παρεμβολών. Σκοπός του προτύπου αυτού είναι να προτείνει τα κριτήρια για την ανάλυση της παρεμβολής. Επίσης έχει ως στόχο τη δημιουργία ενός πλαισίου για τη μέτρηση, την εξερεύνηση και την ανάλυση της παρεμβολής μεταξύ ασύρματων συσκευών και υπηρεσιών. Το IEEE 1900.2 αναπτύσσει ένα κοινό πρότυπο για την επίλυση οποιασδήποτε σύγκρουσης που μπορεί να προκύψει σε ένα σύστημα γνωστικών ραδιοεπικοινωνιών (π.χ. μεταξύ συσκευών που λειτουργούν στην ίδια φασματική περιοχή).
- IEEE 1900.3: Το πρότυπο αυτό σχετίζεται με τη σύσταση ομάδας εργασίας για την αξιολόγηση της συμμόρφωσης των μονάδων λογισμικού των συστημάτων SDR (software defined radio). Αυτή η ομάδα εργασίας ασχολείται με την ανάπτυξη και τον ορισμό των μεθόδων ελέγχου για την αξιολόγηση της συμμόρφωσης των στοιχείων λογισμικού σε ασύρματες

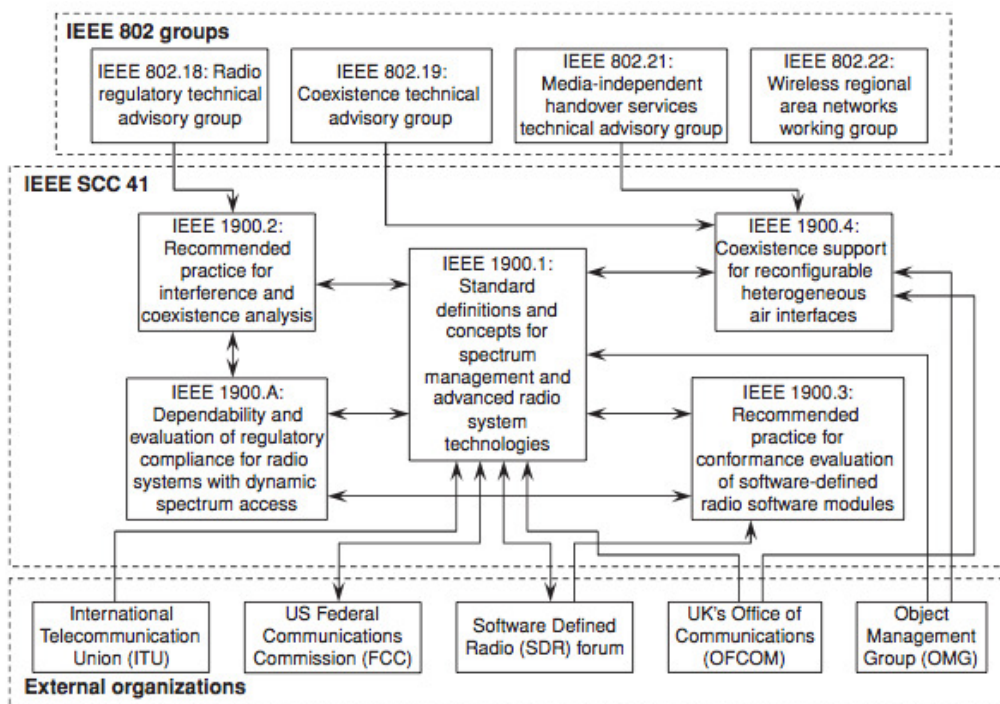
συσκευές που χρησιμοποιούν ραδιοεπικοινωνίες λογισμικού. Βασικός στόχος του IEEE 1900.3 είναι η εξασφάλιση της συμβατότητας σε ό,τι αφορά το λογισμικό. Αυτή η συμμόρφωση είναι απαραίτητη για την επικύρωση και την πιστοποίηση των τελικών προϊόντων γνωστικών ραδιοεπικοινωνιών.

- IEEE 1900.4: Το πρότυπο αυτό σχετίζεται με την υποστήριξη της συνύπαρξης μεταξύ ετερογενών ασύρματων διεπαφών (interfaces). Στα δίκτυα επόμενης γενιάς, μια ασύρματη συσκευή πρέπει να μπορεί να χρησιμοποιεί διαφορετικές ασύρματες τεχνολογίες ταυτοχρόνως. Συνεπώς, η δυναμική πρόσβαση φάσματος πρέπει να υποστηρίζει επιλογή διαφορετικών υπηρεσιών. Σκοπός αυτής της ομάδας εργασίας είναι ο ορισμός μίας ολικής αρχιτεκτονικής συστήματος καθώς και ο ορισμός των βασικών λειτουργιών του.
- IEEE 1900.A: Αυτή η ομάδα μελέτης είναι υπεύθυνη για την πιστοποίηση των συσκευών που στηρίζονται στη δυναμική πρόσβαση φάσματος. Η ομάδα αποσκοπεί στη συμμόρφωση των ασύρματων συσκευών δυναμικής πρόσβασης σύμφωνα με τις ρυθμιστικές αρχές. Νέες μεθοδολογίες και διαδικασίες έλεγχου πρέπει να αναπτυχθούν προκειμένου οι πιστοποιημένες συσκευές να μην παρεμβάλλουν στην εκπομπή αδειοδοτημένων συσκευών.

Υπάρχουν και άλλα πρότυπα της IEEE που σχετίζονται με τα ασύρματα δίκτυα επόμενης γενιάς και τις γνωστικές ραδιοεπικοινωνίες όπως για παράδειγμα τα IEEE 802.18, IEEE 802.19, IEEE 802.21 και IEEE 802.22. Η ομάδα εργασίας του IEEE 802.18 είναι μια ρυθμιστική-συμβουλευτική ομάδα υπεύθυνη για τη συμμετοχή και την παρακολούθηση της εξέλιξης των ρυθμιστικών ενεργειών για διάφορα προγράμματα. Η ομάδα αυτή σχολιάζει και συμβουλεύει τις ρυθμιστικές αρχές. Η ομάδα εργασίας του 802.19 ασχολείται με θέματα συνύπαρξης μεταξύ μη αδειοδοτημένων ασύρματων δικτύων που χρησιμοποιούν IEEE 802 πρότυπα (π.χ. IEEE 802.11 Bluetooth). Στην περίπτωση αυτή, όταν εισάγεται ένα νέο πρότυπο για μη αδειοδοτημένα ασύρματα δίκτυα, η ομάδα του 802.19 θα αξιολογήσει τις διαβεβαιώσεις συνύπαρξης του νέου προτύπου προκειμένου να εγγυηθεί την συνύπαρξη με τις ήδη υπάρχουσες ασύρματες τεχνολογίες που λειτουργούν στην ίδια φασματική περιοχή. Το 802.21 είναι ένα νέο πρότυπο για τη διαχείριση της

κινητικότητας σε ομογενή και ετερογενή δίκτυα. Πρόκειται να αποτελέσει θεμελιώδες πρότυπο για τα δίκτυα επόμενης γενιάς στα οποία ένας κινητός χρήστης θα μπορεί να χρησιμοποιεί πολλαπλές ασύρματες τεχνολογίες ταυτόχρονα. Το 802.22 είναι μια νέα ασύρματη τεχνολογία που υποστηρίζει τη μετάδοση δεδομένων σε μια ευρεία περιοχή κάλυψης. Η τεχνολογία αυτή πρόκειται να λειτουργήσει στις ζώνες συχνοτήτων της τηλεόρασης που παραμένουν ακρησιμοποιήτες σε μεγάλο βαθμό. Ωστόσο, εφαρμόζεται δυναμική πρόσβαση στο φάσμα για να εξασφαλιστεί ότι η αρχική υπηρεσία δεν θα επηρεαστεί από τις συσκευές τεχνολογίας 802.22.

Οι σχέσεις μεταξύ των διαφόρων στοιχείων της IEEE SCC 41 φαίνονται στο σχήμα 1.19.



Σχήμα 1.19: Σχέση μεταξύ των στοιχείων της IEEE SCC 41

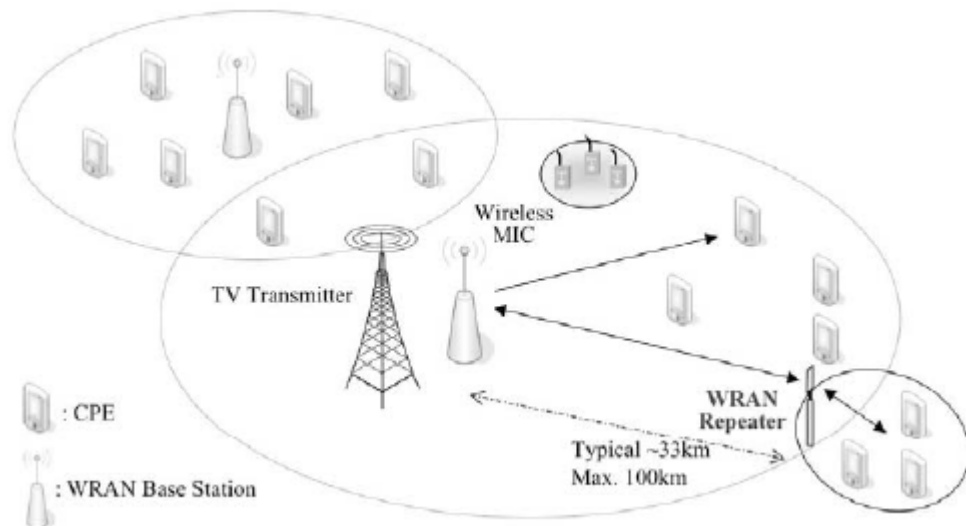
1.7.2 IEEE 802.22

Έχει παρατηρηθεί ότι ορισμένες ζώνες συχνοτήτων τηλεόρασης σε κάποιες περιοχές παραμένουν ανενεργές για μεγάλο χρονικό διάστημα. Για τη βελτίωση της χρησιμοποίησης των ζωνών αυτών, η ομοσπονδιακή επιτροπή επικοινωνιών των

Ηνωμένων Πολιτειών (FCC – Federal Communication Commission) πρότεινε μια αλλαγή στην πολιτική χρήσης του φάσματος ώστε να επιτρέπεται η λειτουργία μη αδειοδοτημένων υπηρεσιών σε φασματικές ζώνες τηλεόρασης κάτω των 900 MHz. Επειδή αυτές οι ζώνες έχουν χαμηλές συχνότητες, τα χαρακτηριστικά διάδοσης είναι κατάλληλα για εκπομπές μεγάλης απόστασης.

Το πρότυπο IEEE 802.22 προτάθηκε ως νέα τεχνολογία ασύρματης περιφερειακής δικτύωσης (Wireless Regional Network-WRAN) που σχεδιάστηκε για να λειτουργήσει στη φασματική περιοχή των τηλεοπτικών υπηρεσιών (54 MHz – 862 MHz) και να συνυπάρξει με αυτές. Το IEEE 802.22 αναμένεται να εξυπηρετεί ασύρματους χρήστες σε κυψέλη με περιοχή κάλυψης μέχρι και 100 km και να κάνει χρήση γνωστικών ραδιοεπικοινωνιών για την αποφυγή παρεμβολής με τις επικείμενες αδειοδοτημένες υπηρεσίες τηλεόρασης.

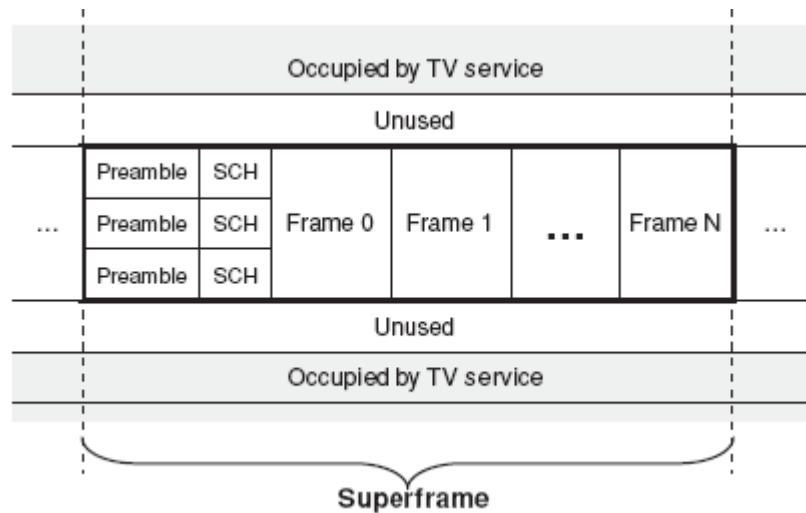
Η αρχιτεκτονική ενός 802.22 WRAN (σχήμα 1.20) είναι παρόμοια με αυτή των σύγχρονων ευρυζωνικών δικτύων ασύρματης πρόσβασης (Broadband Wireless Access networks – BWA) του 802.16 WiMAX. Συγκεκριμένα, τα δίκτυα 802.22 WRAN λειτουργούν σε τοπολογία σταθερού σημείου-προς-πολλαπλά σημεία, όπου ο σταθμός βάσης μιας κυψέλης ελέγχει τις επικοινωνίες των τερματικών συσκευών καταναλωτών (CPEs – Consumer Premise Equipments). Ο σταθμός βάσης διατηρεί τον έλεγχο μιας κατανεμημένης στρατηγικής ανίχνευσης σημάτων πρωτογενών χρηστών και αποφασίζει πότε και πώς πρέπει να έχουν πρόσβαση στο φάσμα οι δευτερογενείς χρήστες (WRAN). Για παράδειγμα, η πρόσβαση στο φάσμα από τις τερματικές συσκευές καταναλωτών και η εμβέλεια που δίνεται στους δευτερογενείς χρήστες καθορίζεται από το σταθμό βάσης. Τέλος, για την αύξηση της περιοχής κάλυψης ενός WRAN μπορούν να χρησιμοποιηθούν σταθμοί αναμετάδοσης.



Σχήμα 1.20: Αρχιτεκτονική δικτύου 802.2

1.7.2.1 Φυσικό στρώμα

Στο φυσικό στρώμα οι εκπομπές των 802.22 συσκευών χρησιμοποιούν ορθογωνική πολλαπλή πρόσβαση διαίρεσης συχνότητας (Orthogonal Frequency-Division Multiple Access - OFDMA). Πολλαπλές ζώνες συχνοτήτων μπορούν να χρησιμοποιηθούν ταυτόχρονα για τη μετάδοση μιας ροής δεδομένων χρησιμοποιώντας τεχνικές ομαδοποίησης καναλιών που παρέχει η OFDM. Παρόμοιες τεχνικές άθροισης καναλιών χρησιμοποιούνται και από τα συστήματα 802.11n που επίσης βασίζονται στην OFDM. Το πρότυπο αναμένεται να υποστηρίζει διάφορους τρόπους διαμόρφωσης και κωδικοποίησης με τη φασματική απόδοση να κυμαίνεται από 0.5 bps/Hz μέχρι 5 bps/Hz. Αν θεωρηθεί μια μέση τιμή 3 bps/Hz αυτή αντιστοιχεί σε ρυθμό δεδομένων 18Mbps σε ένα κανάλι τηλεόρασης εύρους 6MHz. Ένα δίκτυο 802.22 αναμένεται να υποστηρίζει ταυτοχρόνως περισσότερους από 10 χρήστες ανά ζώνη και η αθροιστική ρυθμαπόδοση (throughput) για downlink και uplink θα είναι 1.5 Mbps και 384 kbps, αντίστοιχα.



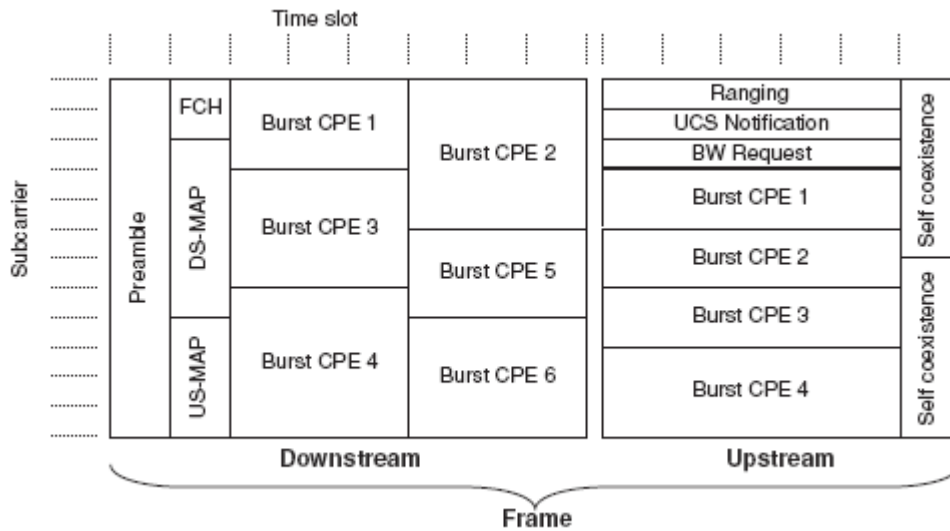
Σχήμα 1.21: Δομή υπερπλασίου του 802.22

1.7.2.2 Στρώμα MAC

Στο σχήμα 1.21 παρουσιάζεται η δομή υπερπλασίου του 802.22 στρώματος MAC. Στην αρχή κάθε υπερπλασίου, το προοίμιο και η επικεφαλίδα ελέγχου υπερπλασίου (Superframe Control Header – SCH) εκπέμπονται από το σταθμό βάσης σε όλες τις διαθέσιμες ζώνες συχνοτήτων. Το προοίμιο χρησιμοποιείται για την προστασία της αδειοδοτημένης υπηρεσίας ενώ η επικεφαλίδα ελέγχου υπερπλασίου χρησιμοποιείται από τις τερματικές συσκευές για συγχρονισμό με το σταθμό βάσης και περιέχει όλες τις πληροφορίες που χρειάζεται μια τερματική συσκευή για να ξεκινήσει μια σύνδεση.

Ένα υπερπλαίσιο αποτελείται από πολλαπλά πλαίσια εκπομπής, η δομή των οποίων φαίνεται στο σχήμα 1.22. Η εκπομπή πλαισίων γίνεται κατά χρονοσχισμές και στηρίζεται σε OFDM υποφέρουσες. Σε κάθε πλαίσιο υπάρχουν υποπλαίσια καθόδου (downstream) και ανόδου (upstream). Ένα πλαίσιο καθόδου αποτελείται από ένα προοίμιο για την ένδειξη της αρχής του πλαισίου. Στη συνέχεια, οι ετικέτες US-MAP και DS-MAP χρησιμοποιούνται για να υποδείξουν τη δομή των υποπλασίων ανόδου και καθόδου, αντίστοιχα. Στο υποπλαίσιο ανόδου υπάρχουν διαστήματα περιεχομένου σχετικά με την αρχικοποίηση των τερματικών συσκευών, την ανακοίνωση επείγουσας κατάστασης συνύπαρξης και τον εντοπισμό άλλων 802.22 δικτύων. Και στα δύο πλαίσια δεσμεύεται χώρος για τη μετάδοση

δεδομένων από και προς τις τερματικές συσκευές. Η δομή των υπερπλαισίων, πλαισίων, υποπλαισίων καθώς και το μέγεθος εκπομπής των τερματικών σταθμών ελέγχονται από το σταθμό βάσης.



Σχήμα 1.22: Δομή ενός 802.22 πλαισίου

Για το ξεκίνημα μιας σύνδεσης με το σταθμό βάσης, η τερματική συσκευή καταναλωτή ανιχνεύει και αναγνωρίζει τις διαθέσιμες ζώνες συχνοτήτων και με βάση αυτές ψάχνει για την εκπεμπόμενη από το σταθμό βάσης επικεφαλίδα ελέγχου υπερπλαisiού. Αφού λάβει την επικεφαλίδα ξεκινά τη σύνδεση στέλνοντας μήνυμα αίτησης αποστολής (request to send - RTS) στο σταθμό βάσης. Ο σταθμός βάσης δεσμεύει χώρο στο πλαίσιο και επιτρέπει στην τερματική συσκευή να εκπέμψει ανάλογα με τη διαθεσιμότητα του φάσματος.

1.8 Πιθανές εφαρμογές γνωστικών ραδιοεπικοινωνιών

Οι γνωστικές ραδιοεπικοινωνίες μπορούν να εφαρμοστούν σε μια ποικιλία δικτύων ασύρματων επικοινωνιών, μερικά από τα οποία περιγράφονται στη συνέχεια.

- *Ασύρματα δίκτυα επόμενης γενιάς:* Οι γνωστικές ραδιοεπικοινωνίες αναμένεται να αποτελέσουν τεχνολογία κλειδί για τα επόμενης γενιάς ετερογενή ασύρματα δίκτυα καθώς θα μπορούν να προσφέρουν αυξημένη

ευφυΐα στον εξοπλισμό των χρηστών όσο και του παρόχου, ώστε να επιτυγχάνεται καλύτερη αξιοποίηση των πόρων του δικτύου. Από την πλευρά του χρήστη, μία κινητή συσκευή με πολλαπλές διεπαφές αέρος μπορεί να παρατηρεί την κατάσταση των ασύρματων δικτύων πρόσβασης και να αποφασίζει σε ποιο δίκτυο θα συνδεθεί. Από την πλευρά του παρόχου, βελτιστοποιείται η αξιοποίηση των φασματικών πόρων για το δεδομένο σύνολο χρηστών και τις αντίστοιχες απαιτήσεις υπηρεσιών τους (QoS). Στηριζόμενοι στα μοντέλα κινητικότητας των χρηστών μπορούν να ενσωματωθούν οι κατάλληλοι μηχανισμοί στην υποδομή του παρόχου προκειμένου να κατανέμεται η κίνηση σε πολλαπλά διαθέσιμα δίκτυα και να αποφεύγεται η συμφόρηση.

- *Συνύπαρξη με διαφορετικές ασύρματες τεχνολογίες:* Νέες ασύρματες τεχνολογίες (π.χ. 802.22 WRAN [25]) αναπτύσσονται για την αναχρησιμοποίηση του φάσματος συχνοτήτων που είναι κατανεμημένο σε άλλες ασύρματες υπηρεσίες (π.χ. υπηρεσίες τηλεόρασης). Οι γνωστικές ραδιοεπικοινωνίες αποτελούν μία λύση προκειμένου να επιτευχθεί συνύπαρξη μεταξύ αυτών των διαφορετικών τεχνολογιών και ασύρματων υπηρεσιών. Για παράδειγμα, οι χρήστες του 802.22 WRAN μπορούν να χρησιμοποιούν ευκαιριακά τις ζώνες συχνοτήτων που έχουν παραχωρηθεί για υπηρεσίες τηλεόρασης, εφόσον δεν εκπέμπει στην περιοχή τηλεοπτικός σταθμός [26].
- *Ηλεκτρονικές υπηρεσίες υγείας:* Για την καλύτερη εξυπηρέτηση των ασθενών, οι υπηρεσίες ιατρικής περίθαλψης έχουν υιοθετήσει διάφορους τύπους ασύρματων τεχνολογιών. Ωστόσο, η χρήση συσκευών ασύρματης επικοινωνίας στην ιατρική περίθαλψη περιορίζεται από τις απαιτήσεις για ηλεκτρομαγνητική παρεμβολή και ηλεκτρομαγνητική συμβατότητα με τον ιατρικό εξοπλισμό. Από τη στιγμή όπου ο ιατρικός εξοπλισμός και οι ανιχνευτές βιολογικού σήματος είναι ευαίσθητοι στις ηλεκτρομαγνητικές παρεμβολές, η ισχύς εκπομπής των ασύρματων συσκευών πρέπει να είναι ελεγχόμενη. Επιπλέον, διάφορες συσκευές βιοιατρικής πραγματοποιούν εκπομπές RF. Η φασματική χρήση αυτών των συσκευών πρέπει να

επιλέγεται προσεκτικά έτσι ώστε να αποφεύγεται η μεταξύ τους παρεμβολή. Στην περίπτωση αυτή, συνεπώς, υπάρχει δυνατότητα εφαρμογής στοιχείων ενισχυτικής μάθησης. Για παράδειγμα, πολλοί ασύρματοι αισθητήρες έχουν σχεδιαστεί για να λειτουργούν στην ζώνη ISM(industrial, scientific, and medical), με αποτέλεσμα να μπορούν να χρησιμοποιήσουν γνωστικές ραδιοεπικοινωνίες για την επιλογή των κατάλληλων καναλιών εκπομπής και την αποφυγή παρεμβολών.

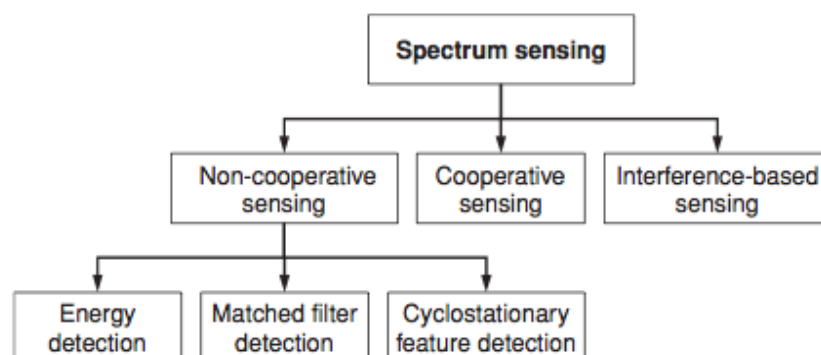
- *Δίκτυα επείγουσας ανάγκης:* Τα δίκτυα δημόσιας ασφάλειας και επείγουσας ανάγκης μπορούν να επωφεληθούν από τις γνωστικές ραδιοεπικοινωνίες για την παροχή ευέλικτης και αξιόπιστης ασύρματης επικοινωνίας. Για παράδειγμα, σε ένα σενάριο καταστροφής, οι τηλεπικοινωνιακές υποδομές μπορεί να τεθούν εκτός λειτουργίας. Συνεπώς, για την υποστήριξη των απαραίτητων αναγκών επικοινωνίας μπορεί να χρειαστεί ένα προσαρμοστικό σύστημα ασύρματης επικοινωνίας. Ένα τέτοιο δίκτυο μπορεί να βασίζεται σε γνωστικές ραδιοεπικοινωνίες προκειμένου να επιτρέπει ασύρματη μετάδοση και λήψη επί ενός μεγάλου τμήματος του φάσματος συχνοτήτων.
- *Στρατιωτικά δίκτυα:* Μέσω των γνωστικών ραδιοεπικοινωνιών, οι παράμετροι ασύρματης επικοινωνίας μπορούν να προσαρμόζονται δυναμικά με βάση την ώρα, την τοποθεσία καθώς και το είδος της αποστολής των στρατιωτών. Για παράδειγμα, όταν κάποιες συχνότητες είναι κατειλημμένες ή θορυβώδεις, ο γνωστικός πομποδέκτης θα αναζητήσει εναλλακτικές ζώνες για επικοινωνία. Επίσης, ένα γνωστικό ασύρματο σύστημα μπορεί να ελέγχει την ισχύ εκπομπής σε μία περιοχή ώστε να μην προκαλεί παρεμβολή στα στρατιωτικά τηλεπικοινωνιακά συστήματα υψηλής προτεραιότητας [27].

ΚΕΦΑΛΑΙΟ 2: ΑΝΙΧΝΕΥΣΗ ΦΑΣΜΑΤΟΣ

Οι γνωστικές ραδιοεπικοινωνίες έχουν προταθεί ως λύση για τη βελτίωση της χρησιμοποίησης του φάσματος. Συνεπώς, βασική προϋπόθεση των δικτύων γνωστικών ραδιοεπικοινωνιών είναι η δυνατότητα εντοπισμού των φασματικών κενών (spectrum white spaces). Οι γνωστικές ραδιοεπικοινωνίες έχουν σχεδιαστεί ώστε να αντιλαμβάνονται τις αλλαγές στο περιβάλλον τους. Η λειτουργία ανίχνευσης φάσματος δίνει τη δυνατότητα στα συστήματα γνωστικών ραδιοεπικοινωνιών να εντοπίζουν τα φασματικά κενά και να προσαρμόζονται στις αλλαγές του περιβάλλοντός τους.

Ο αποδοτικότερος τρόπος για τον εντοπισμό του αχρησιμοποίητου φάσματος είναι ο εντοπισμός των πρωτογενών χρηστών που είναι ενεργοί εντός της περιοχής κάλυψης του δευτερογενούς χρήστη. Στην πράξη, ωστόσο, είναι δύσκολο για ένα δευτερογενή χρήστη να έχει άμεση αντίληψη του διαύλου μεταξύ ενός πρωτογενούς δέκτη και ενός πρωτογενούς πομπού. Επομένως, η πλέον πρόσφατη έρευνα στον τομέα της ανίχνευσης φάσματος εστιάζει στον εντοπισμό των πρωτογενών πομπών με βάση τις τοπικές παρατηρήσεις των δευτερογενών χρηστών.

Όπως φαίνεται από το σχήμα 2.1 οι τεχνικές ανίχνευσης φάσματος μπορούν να κριθούν σε τρεις βασικές κατηγορίες συγκεκριμένα στη μη συνεργατική ανίχνευση, τη συνεργατική ανίχνευση και την ανίχνευση παρεμβολής.



Σχήμα 2.1: Κατηγορίες τεχνικών ανίχνευσης φάσματος

2.1 Μη συνεργατική ανίχνευση πομπού

Η μη συνεργατική ανίχνευση φάσματος χρησιμοποιείται από ένα μη αδειοδοτημένο χρήστη για τον εντοπισμό του σήματος που μεταδίδει ένας πρωτογενής πομπός και στηρίζεται σε τοπικές μετρήσεις και παρατηρήσεις. Το μοντέλο ανίχνευσης σήματος τη χρονική στιγμή t μπορεί να περιγραφεί ως εξής [29]:

$$x(t) = \begin{cases} n(t), & H_0, \\ h \times s(t) + n(t), & H_1, \end{cases} \quad (2.1)$$

όπου $x(t)$ είναι το σήμα που λαμβάνει ένας μη αδειοδοτημένος χρήστης, $s(t)$ είναι το εκπεμπόμενο σήμα του αδειοδοτημένου χρήστη, $n(t)$ είναι ο προσθετικός λευκός θόρυβος τύπου Gauss (additive white Gaussian noise, AWGN) και h είναι το κέρδος του διαύλου. Οι μεταβλητές H_0 , H_1 εκφράζουν την υπόθεση του να μην υπάρχει και να υπάρχει, αντίστοιχα, σήμα πρωτογενούς χρήστη στην δεδομένη ζώνη συχνοτήτων.

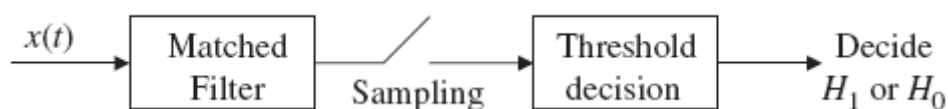
Η επίδοση μιας τεχνικής ανίχνευσης φάσματος αποτιμάται μέσω της πιθανότητας επιτυχημένης ανίχνευσης (P_d), πιθανότητας λανθασμένου συναγερμού (P_f) και πιθανότητας αστοχίας (P_m), όπου $P_d = \text{Prob} \{ \text{απόφαση} = H_1 | H_1 \}$, $P_f = \text{Prob} \{ \text{απόφαση} = H_1 | H_0 \}$ και $P_m = \text{Prob} \{ \text{απόφαση} = H_0 | H_1 \}$.

Τρεις βασικές τεχνικές ανίχνευσης πομπού είναι η ανίχνευση προσαρμοσμένου φίλτρου, η ανίχνευση ενέργειας και η κυκλοστατική ανίχνευση.

2.1.1 Ανίχνευση προσαρμοσμένου φίλτρου

Το προσαρμοσμένο φίλτρο χρησιμοποιείται ευρέως για τον εντοπισμό ενός σήματος συγκρίνοντας ένα γνωστό σήμα (πρότυπο) με το σήμα εισόδου. Επίσης, η ανίχνευση προσαρμοσμένου φίλτρου μεγιστοποιεί το λαμβανόμενο σηματοθορυβικό λόγο (SNR) λήψης του υπό μέτρηση σήματος. Συνεπώς, εφόσον είναι γνωστή η πληροφορία του σήματος (π.χ. διαμόρφωση, διάταξη πακέτου) που εκπέμπεται από ένα αδειοδοτημένο χρήστη, το προσαρμοσμένο φίλτρο αποτελεί βέλτιστο τρόπο ανίχνευσης σε ένα περιβάλλον στατικού θορύβου Gauss [30]. Επειδή

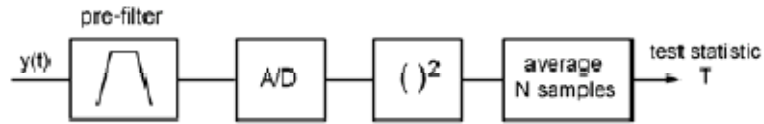
ο εντοπισμός γίνεται με χρήση προτύπου, το προσαρμοσμένο φίλτρο απαιτεί μόνο ένα μικρό χρονικό διάστημα για να λειτουργήσει. Ωστόσο, όταν το πρότυπο δεν είναι διαθέσιμο ή είναι εσφαλμένο η απόδοση της ανίχνευσης μειώνεται σημαντικά. Η ανίχνευση με προσαρμοσμένο φίλτρο είναι κατάλληλη όταν το σήμα που εκπέμπει ένας αδειοδοτημένος χρήστης έχει πιλότο, προοίμιο, λέξη συγχρονισμού ή κώδικες διασποράς έτσι ώστε να είναι δυνατή η κατασκευή προτύπου για την ανίχνευση φάσματος. Ωστόσο, ένα σημαντικό μειονέκτημα του προσαρμοσμένου φίλτρου είναι το κόστος του και κυρίως το ότι η συσκευή γνωστικών ραδιοεπικοινωνιών πρέπει να έχει ένα αφιερωμένο δέκτη για κάθε κατηγορία πρωτεύοντος χρήστη, απαίτηση που έχει και το μειονέκτημα της μεγάλης κατανάλωσης ισχύος. Η υλοποίηση του συστήματος ανίχνευσης με χρήση προσαρμοσμένου φίλτρου φαίνεται στο σχήμα 2.2.



Σχήμα 2.2: Υλοποίηση ανιχνευτή με προσαρμοσμένο φίλτρο

2.1.2 Ανιχνευτής ενέργειας πομπού

Η ανίχνευση ενέργειας είναι ο βέλτιστος τρόπος ανίχνευσης του φάσματος όταν δεν είναι γνωστά τα χαρακτηριστικά του σήματος (π.χ. διαμόρφωση, διάταξη πακέτου) του αδειοδοτημένου χρήστη[30]. Στην περίπτωση αυτή, ένας βέλτιστος ανιχνευτής βασισμένος σε προσαρμοσμένο φίλτρο δεν αποτελεί επιλογή, καθώς απαιτεί γνώση της μορφής των δεδομένων για σύμφωνη αποδιαμόρφωση. Ένας τυπικός ανιχνευτής ενέργειας αποτελείται από ένα βαθυπερατό φίλτρο (έχει βέβαια 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, ένα κύκλωμα τετραγωνικού νόμου και τέλος ένα ολοκληρωτή (σχήμα 2.3). Το σήμα εξόδου από το φίλτρο εύρους W υψώνεται στο τετράγωνο και ολοκληρώνεται στο διάστημα παρατήρησης. Ένας αλγόριθμος παρατήρησης συγκρίνει την έξοδο του ολοκληρωτή, T , με ένα κατώφλι λ προκειμένου να διαπιστώσει αν υπάρχει ή όχι αδειοδοτημένος χρήστης. Αν $T > \lambda$, υπάρχει πρωτογενές σήμα, διαφορετικά δεν υπάρχει.



Σχήμα 2.3: Υλοποίηση ανιχνευτή ενέργειας με βαθυπερατό φίλτρο

Αν ο εντοπισμός ενέργειας μπορεί να εφαρμοστεί σε ένα περιβάλλον χωρίς εξασθένιση όπου h το κέρδος διαύλου της (2.1), η πιθανότητα εντοπισμού P_d και η πιθανότητα λανθασμένου συναγερμού P_f υπολογίζονται μέσω των σχέσεων:

$$P_d = P\{Y > \lambda | H_1\} = Q_m(\sqrt{2\gamma}, \sqrt{\lambda}) \quad (2.2)$$

$$P_f = P\{Y > \lambda | H_0\} = \frac{\Gamma(m, \lambda/2)}{\Gamma(m)} \quad (2.3)$$

όπου γ είναι το SNR του λαμβανόμενου σήματος, $u = TW$ το γινόμενο του διαστήματος παρατήρησης επί το εύρος του φίλτρου, Γ η συνάρτηση γάμμα και $Q_m()$ η γενικευμένη συνάρτηση Q που δίνεται από τη σχέση

$$Q(x, y) = \int_y^\infty e^{-\frac{(x^2+w^2)}{2}} I_0(xw) w dw \quad (2.4)$$

όπου

$$I_0(x) = \sum_{k=0}^{\infty} \frac{(-\frac{1}{4}x^2)^k}{k! \Gamma(k+1)} \quad (2.5)$$

Οι συναρτήσεις Γ δίνονται από τις σχέσεις

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (2.6)$$

$$\Gamma(a, x) = \frac{1}{\Gamma(a)} \int_x^\infty e^{-t} t^{a-1} dt, \quad a > 0 \quad (2.7)$$

Χαμηλές τιμές της P_d οδηγούν σε αυξημένη παρεμβολή στον πρωτογενή χρήστη καθώς η πιθανότητα να μην γίνει αντιληπτός από ένα δευτερογενή χρήστη είναι μεγάλη. Αντίθετα υψηλές τιμές P_f έχουν ως αποτέλεσμα τη χαμηλή χρησιμοποίηση του φάσματος καθώς οι λανθασμένοι συναγερμοί ύπαρξης πρωτογενούς σήματος αυξάνουν το πλήθος των χαμένων ευκαιριών πρόσβασης. Το γεγονός ότι ο ανιχνευτής ενέργειας είναι εύκολα υλοποιήσιμος έχει ως αποτέλεσμα

την ευρεία υιοθέτηση του από δευτερογενή συστήματα γνωστικών ραδιοεπικοινωνιών.

Αν ληφθούν υπόψη οι παράγοντες εξασθένησης και πολυδιαδρομικής διάδοσης, η P_f είναι ανεξάρτητη της συνάρτησης Γ και η πιθανότητα εντοπισμού δίνεται από τη σχέση

$$P_d = \int Q_m(\sqrt{2\gamma}, \sqrt{\lambda}) f_\gamma(x) dx \quad (2.8)$$

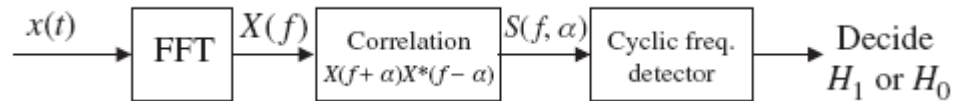
όπου $f_\gamma(x)$ είναι η συνάρτηση κατανομής πιθανότητας του SNR.

Παρά το γεγονός ότι ο ανιχνευτής ενέργειας είναι εύκολος στη υλοποίηση, χαρακτηρίζεται από αρκετά μειονεκτήματα, όπως για παράδειγμα η σχετικά αργή ταχύτητα ανίχνευσης. Επιπλέον, μειονεκτήματα του ανιχνευτή ενέργειας αποτελούν το ότι επηρεάζεται από τη μεταβλητότητα του θορύβου (ύπαρξη ελάχιστου SNR κάτω από τον οποίο το σήμα δεν μπορεί να ανιχνευθεί) και το ότι μπορεί να εντοπίσει την παρουσία ενός σήματος αλλά δεν μπορεί να αναγνωρίσει τη φύση του (π.χ. διαμορφωμένο σήμα, θόρυβος ή παρεμβολή). Αυτό σημαίνει ότι δεν μπορεί να διαχωρίσει σήματα δευτερευόντων χρηστών που χρησιμοποιούν το ίδιο κανάλι με αυτό του πρωτεύοντος χρήστη και έχει ως αποτέλεσμα η πιθανότητα λανθασμένου εντοπισμού να αυξάνεται με την παρουσία και άλλων πηγών σήματος εκτός του πρωτεύοντος χρήστη. Τέλος, δεν μπορεί να χρησιμοποιηθεί για σήματα απλωμένου φάσματος (spread spectrum signals) καθώς δεν μπορεί να τα διαχωρίσει από το θόρυβο.

2.1.3 Ανίχνευση στοιχείου κυκλοστατικότητας

Το σήμα που εκπέμπει ένας αδειοδοτημένος χρήστης παρουσιάζει, στη γενική περίπτωση, περιοδικότητα που αναφέρεται ως κυκλοστατικότητα και μπορεί να χρησιμοποιηθεί για τον εντοπισμό της παρουσίας πρωτογενών χρηστών[32]. Ένα σήμα θεωρείται κυκλοστατικό όταν η συνάρτηση αυτοσυσχετίσής του είναι περιοδική. Με βάση την περιοδικότητα αυτή το σήμα πρωτογενούς εκπομπής 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 σχήμα 2.4.



Σχήμα 2.4: Υλοποίηση ανιχνευτή στοιχείου κυκλοστατικότητας

Προκειμένου να βελτιωθεί η συνολική απόδοση της ανίχνευσης φάσματος, σε ένα μη αδειοδοτημένο σύστημα μπορούν να συνδυαστούν διάφορες μέθοδοι ανίχνευσης. Για παράδειγμα, η ανίχνευση ενέργειας μπορεί να χρησιμοποιηθεί για τον εντοπισμό των φασματικών ζωνών που παρουσιάζουν υψηλή πυκνότητα ισχύος (π.χ. λόγω της εκπομπής των αδειοδοτημένων χρηστών). Στη συνέχεια, η κυκλοστατική ανίχνευση μπορεί να χρησιμοποιηθεί στις εναπομείνουσες ζώνες για την ανίχνευση πρωτογενών χρηστών.

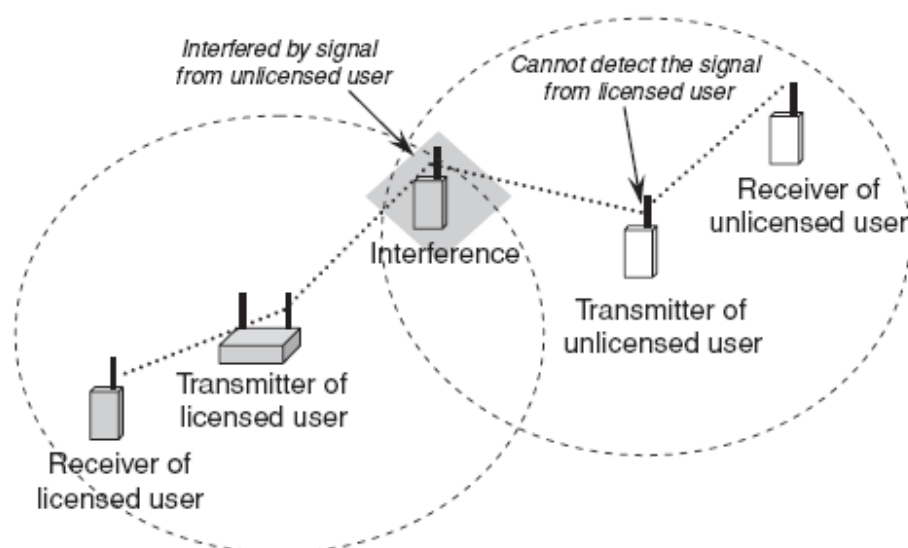
Τεχνική ανίχνευσης	Πλεονέκτημα	Μειονέκτημα
Προσαρμοσμένο φίλτρο	Βέλτιστη επίδοση	Απαιτείται προηγούμενη γνώση του πρωτεύοντος σήματος
Ανιχνευτής ενέργειας	Δεν απαιτείται προηγούμενη γνώση και χαμηλό κόστος	Αδυναμία λειτουργίας υπό χαμηλό SNR λαμβανόμενου σήματος
Ανιχνευτής στοιχείου κυκλοστατικότητας	Δεν επηρεάζεται από το επίπεδο του θορύβου και την ύπαρξη παρεμβολών	Απαιτείται μερική γνώση των χαρακτηριστικών εκπομπής του πρωτεύοντος χρήστη, υψηλό υπολογιστικό κόστος
Ανιχνευτής κυματομορφής	Αποτελεσματική για σήματα ευρείας ζώνης	Αδυναμία χρησιμοποίησης για σήματα απλωμένου φάσματος, υψηλό κόστος υπολογισμού

Πίνακας 2.1

2.2 Συνεργατική ανίχνευση

Ένας μη αδειοδοτημένος χρήστης μπορεί να μην είναι πάντα σε θέση να εντοπίζει το σήμα ενός αδειοδοτημένου πομπού λόγω γεωγραφικού διαχωρισμού και διαλείψεων. Για παράδειγμα, όπως φαίνεται στο σχήμα 2.5, ο πομπός και ο δέκτης του μη αδειοδοτημένου δικτύου αδυνατούν να εντοπίσουν το σήμα του αδειοδοτημένου πομπού καθώς βρίσκονται έξω από την περιοχή κάλυψης του. Όταν ο μη αδειοδοτημένος πομπός εκπέμπει, θα παρεμβάλει τον αδειοδοτημένο δέκτη. Η περίπτωση αυτή είναι γνωστή ως το πρόβλημα του κρυμμένου τερματικού (hidden terminal problem).

Για την επίλυση του προβλήματος του κρυμμένου τερματικού μπορεί να χρησιμοποιηθεί συνεργατική ανίχνευση. Κατά τη συνεργατική ανίχνευση πολλαπλοί δευτερογενείς χρήστες ανταλλάσσουν μεταξύ τους πληροφορία ανίχνευσης προκειμένου να εντοπίσουν την παρουσία αδειοδοτημένων χρηστών. Με την ανταλλαγή πληροφορίας για ανίχνευση φάσματος αντιμετωπίζεται το πρόβλημα του κρυμμένου τερματικού ενώ βελτιώνεται αισθητά και η πιθανότητα εντοπισμού σε περιβάλλον σκίασης.



Σχήμα 2.5: Πρόβλημα κρυμμένου τερματικού

Οι δυσκολίες που θέτει η συνεργατική ανίχνευση περιλαμβάνουν την επιβάρυνση του πρωτοκόλλου συνεργασίας και την αυξημένη πολυπλοκότητα του

συστήματος. Τα μειονεκτήματα και τα πλεονεκτήματα της τοπικής και της συνεργατικής ανίχνευσης αναφέρονται στον Πίνακα 2.2.

Σε αρχιτεκτονικές συνεργατικής ανίχνευσης, ο δίαυλος ελέγχου μπορεί να υλοποιηθεί χρησιμοποιώντας διαφορετικές μεθοδολογίες. Αυτές περιλαμβάνουν αφιερωμένη ζώνη συχνότητας, μη αδειοδοτημένη ζώνη όπως η ζώνη ISM, και υποκείμενο UWB (underlay Ultra WideBand) σύστημα το οποίο αποτελεί μια διαμόρφωση απλωμένου φάσματος σε πολύ μεγάλο εύρος ζώνης με πολύ χαμηλή ισχύ λειτουργίας. Ανάλογα με τις απαιτήσεις κάθε συστήματος επιλέγεται μία από αυτές τις μεθόδους. Επιπλέον, μπορούν να εφαρμοστούν διάφορες τεχνικές για το συνδυασμό των αποτελεσμάτων της ανίχνευσης. Όσον αφορά τη δικτύωση, ο αλγόριθμος συνεργασίας πρέπει να έχει μειωμένη επιβάρυνση πρωτοκόλλου (επιπλέον πακέτα ελέγχου) και να είναι ανθεκτικός σε αλλαγές και βλάβες στο δίκτυο. Επίσης, ο αλγόριθμος συνεργασίας πρέπει να εισάγει τη μικρότερη δυνατή καθυστέρηση, κάτι που είναι πολύ δύσκολο να επιτευχθεί.

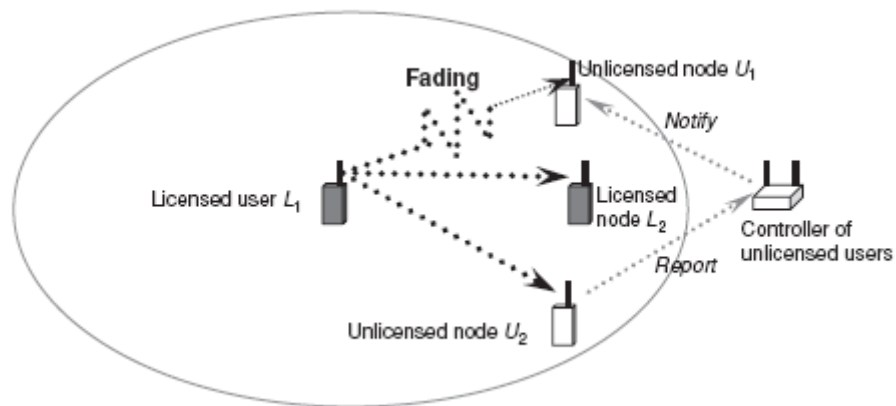
Μέθοδος ανίχνευσης	Πλεονεκτήματα	Μειονεκτήματα
Μη συνεργατική (τοπική)	Απλή υλοποίηση και χαμηλή πολυπλοκότητα	Πρόβλημα κρυμμένου τερματικού
Συνεργατική	Μεγαλύτερη ακρίβεια, αποφυγή προβλήματος σκίασης και κρυμμένου τερματικού	Αυξημένη πολυπλοκότητα ανιχνευτή και συστήματος συνεργασίας, επιβάρυνση πρωτοκόλλου συνεργασίας, απαίτηση για δίαυλο ελέγχου

Πίνακας 2.2

Η συνεργατική ανίχνευση μπορεί να είναι κεντρικά ελεγχόμενη ή καταναμημένη όπως περιγράφεται στη συνέχεια.

2.2.1 Κεντρικά ελεγχόμενη συνεργατική ανίχνευση

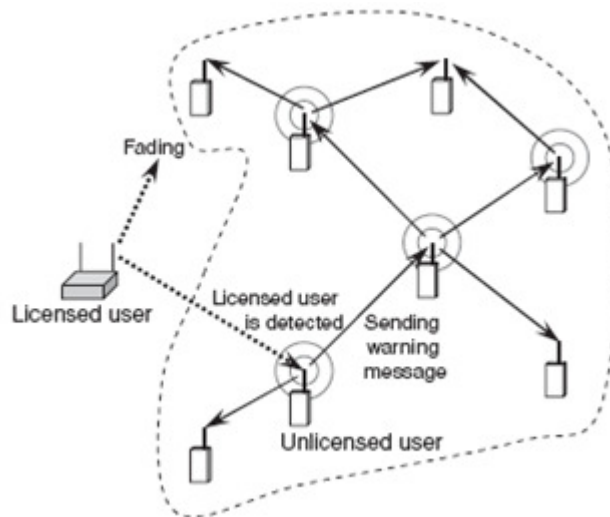
Κατά την κεντρικά ελεγχόμενη ανίχνευση, μια κεντρική μονάδα (συνήθως το σημείο πρόσβασης – Access Point) συλλέγει τις πληροφορίες ανίχνευσης από συσκευές γνωστικών ραδιοεπικοινωνιών, ταυτοποιεί το διαθέσιμο φάσμα και κάνει ευρυεκπομπή της πληροφορίας αυτής σε άλλες δευτερεύουσες συσκευές ή ελέγχει απευθείας την κίνηση που μεταδίδεται μεταξύ των δευτερευόντων συσκευών (σχήμα 2.6). Στόχος είναι η ελάττωση των επιδράσεων των διαλείψεων του διαύλου και η αύξηση της αξιοπιστίας και της ταχύτητας της ανίχνευσης.



Σχήμα 2.6: Παράδειγμα συνεργατικής κεντρικά ελεγχόμενης ανίχνευσης

2.2.2 Κατανεμημένη συνεργατική ανίχνευση

Στην περίπτωση της κατανεμημένης ανίχνευσης, οι συσκευές γνωστικών ραδιοεπικοινωνιών ανταλλάσσουν πληροφορίες μεταξύ τους αλλά λαμβάνουν μόνες τους τις αποφάσεις για τη ζώνη του φάσματος που θα χρησιμοποιήσουν. Η κατανεμημένη ανίχνευση πλεονεκτεί υπό την έννοια ότι δεν υπάρχει η ανάγκη για υποστηρικτική (backbone) υποδομή (σχήμα 2.7).



Σχήμα 2.7: Παράδειγμα συνεργατικής κεντρικά ελεγχόμενης ανίχνευσης

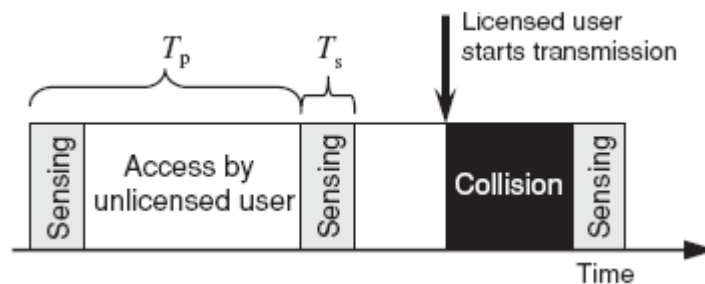
2.3 Ανίχνευση παρεμβολής

Η ανίχνευση παρεμβολής (interference based sensing) αποτελεί μια νέα μέθοδο ανίχνευση που προτάθηκε από την ομοσπονδιακή επιτροπή επικοινωνιών των Ηνωμένων Πολιτειών (FCC – Federal Communications Commission)[35]. Στη περίπτωση αυτή, ο αλγόριθμος ανίχνευσης μετρά τα επίπεδα θορύβου/παρεμβολής (από όλες τις πηγές σήματος) στο δέκτη του πρωτογενούς χρήστη. Η πληροφορία αυτή χρησιμοποιείται από ένα μη αδειοδοτημένο χρήστη προκειμένου να ελέγχει τη πρόσβασή του στο φάσμα έτσι ώστε να μην παραβιάζεται το επίπεδο της θερμοκρασίας παρεμβολής(σχήμα 1.19). Εναλλακτικά, ένας μη αδειοδοτημένος πομπός μπορεί να παρακολουθεί το σήμα ανάδρασης από ένα αδειοδοτημένο δέκτη για να αποκτά γνώση του επιπέδου παρεμβολής.

2.4 Περίοδος ανίχνευσης

Για την προστασία ενός αδειοδοτημένου χρήστη από παρεμβολές, ένας δευτερογενής χρήστης πρέπει να ανιχνεύει περιοδικά το φάσμα ανά T_p χρονικές στιγμές. Όταν διαπιστωθούν φασματικά κενά, ο δευτερογενής χρήστης αποκτά πρόσβαση στο φάσμα. Ωστόσο, κατά τη διάρκεια της εκπομπής, ο αδειοδοτημένος χρήστης δεν έχει πληροφόρηση για τη δράση του αδειοδοτημένου χρήστη μέχρι την

επόμενη περίοδο ανίχνευσης. Με τη χρήση ενός μόνο πομποδέκτη δεν είναι δυνατή η ταυτόχρονη ανίχνευση και χρήση του φάσματος από ένα μη αδειοδοτημένο χρήστη. Αυτό σημαίνει ότι αν ένας αδειοδοτημένος χρήστης αρχίσει να εκπέμπει μεταξύ δύο περιόδων ανίχνευσης (σχήμα 2.8), θα παρεμβληθεί από το δευτερεύοντα χρήστη. Συνεπώς, η QoS (Quality of service) τόσο των αδειοδοτημένων όσο και των μη αδειοδοτημένων χρηστών εξαρτάται από την περίοδο ανίχνευσης T_p . Συγκεκριμένα, αν η περίοδος ανίχνευσης είναι μεγάλη, η χρονική διάρκεια πιθανής παρεμβολής σε ένα αδειοδοτημένο χρήστη είναι μεγάλη. Από την άλλη πλευρά, η διάρκεια της περιόδου ανίχνευσης εισάγει ένα αντιστάθμισμα (tradeoff) επίδοσης. Όσο μικρότερη είναι η περίοδος ανίχνευσης T_p τόσο μικρότερη είναι η διάρκεια παρεμβολής στον αδειοδοτημένο χρήστη, αλλά και τόσο χαμηλότερη η ρυθμαπόδοση (throughput) του μη αδειοδοτημένου χρήστη. Για παράδειγμα, με διάρκεια ανίχνευσης $T_s = 1\text{ ms}$ και περίοδο ανίχνευσης $T_p = 2\text{ ms}$ ο συνολικός χρόνος εκπομπής του μη αδειοδοτημένου χρήστη σε ένα χρονικό διάστημα 10 ms θα είναι μόνο 5 ms καθώς ο δευτερογενής χρήστης ανιχνεύει κάθε 2 ms και η διάρκεια ανίχνευσης είναι 1 ms . Ωστόσο, εάν $T_p = 10\text{ ms}$ ο χρόνος εκπομπής γίνεται 9 ms καθώς ο δευτερογενής χρήστης ανιχνεύει κάθε 9 ms και η διάρκεια ανίχνευσης είναι 1 ms . Αυτό το αντιστάθμισμα ανάμεσα στην επίδοση των αδειοδοτημένων και των μη αδειοδοτημένων χρηστών πρέπει να λαμβάνεται υπόψη κατά τον ορισμό της περιόδου ανίχνευσης.



Σχήμα 2.8: Περίοδος ανίχνευσης φάσματος

Για την επίλυση του προηγούμενου προβλήματος μπορούν να χρησιμοποιηθούν συστήματα που διαθέτουν ένα πομποδέκτη για εκπομπή δεδομένων και ένα ξεχωριστό δέκτη που αναλαμβάνει την ανίχνευση φάσματος.

Με τον τρόπο αυτό γίνεται εφικτή η ταυτόχρονη εκπομπή και ανίχνευση φάσματος. Βέβαια και σε αυτή την περίπτωση πρέπει να οριστεί μία περίοδος ανίχνευσης καθώς η συνεχής ανίχνευση οδηγεί σε υψηλή κατανάλωση ισχύος.

2.5 Ανίχνευση φάσματος σε σύγχρονα πρότυπα επικοινωνιών

Πρόσφατα ασύρματα πρότυπα έχουν αρχίσει να υιοθετούν χαρακτηριστικά γνωστικών ραδιοεπικοινωνιών. Στη συνέχεια αναφέρονται πρότυπα που χρησιμοποιούν κάποιας μορφής ανίχνευση φάσματος για προσαρμογή ή για δυναμική επιλογή συχνότητας (Dynamic Frequency Selection - DFS).

2.5.1 IEEE 802.11k

Το πρότυπο IEEE 802.11k περιλαμβάνει πολλά χαρακτηριστικά μετρήσεων και διαχείρισης καναλιού. Μερικές από τις μετρήσεις αυτές περιλαμβάνουν φορτίο καναλιού, αναφορά κίνησης καναλιού, ιστόγραμμα θορύβου και στατιστικά επίδοσης σταθμού[37]. Το ιστόγραμμα θορύβου παρέχει μεθόδους για τη μέτρηση του επιπέδου της παρεμβολής που προέρχεται από όλη την ηλεκτρομαγνητική ακτινοβολία που δεν σχετίζεται με το 802.11. Σύμφωνα με το πρότυπο αυτό, ένα σημείο πρόσβασης (Access Point – AP) μπορεί να συλλέγει και να επεξεργάζεται τη διαυλική πληροφορία κάθε τερματικού σταθμού. Στη συνέχεια, τα αποτελέσματα της επεξεργασίας μπορούν να χρησιμοποιηθούν από το σημείο πρόσβασης για τον έλεγχο της πρόσβασης των χρηστών. Για παράδειγμα, αν παρατηρηθεί ότι τα επίπεδα παρεμβολής σε ένα κανάλι είναι υψηλότερα από το ορισμένο κατώφλι, ένας σταθμός πρόσβασης μπορεί να αποφασίσει για όλους τους τερματικούς σταθμούς να μεταβούν σε ένα κανάλι με χαμηλότερα επίπεδα παρεμβολής. Επιπλέον, όταν κάποιο σημείο πρόσβασης είναι υπερφορτωμένο, ένας νέος τερματικός σταθμός μπορεί να οδηγηθεί σε σύνδεση με ένα άλλο σημείο πρόσβασης με μικρότερο φορτίο[38]. Στην περίπτωση αυτή, η πληροφορία για τη μέση καθυστέρηση πρόσβασης μπορεί να χρησιμοποιηθεί για τον υπολογισμό του τηλεπικοινωνιακού φορτίου του δικτύου. Επιπλέον, η σηματοδότηση με βάση το 802.11k μπορεί να χρησιμοποιηθεί για τον καθορισμό της περιοχής κάλυψης ενός σταθμού πρόσβασης και των γειτόνων του[39].

2.5.2 Bluetooth

Το Bluetooth λειτουργεί στη φασματική ζώνη ISM (μη αδειοδοτημένη ζώνη των 2.4 GHz)· επομένως, η αποφυγή παρεμβολών από και με άλλες ασύρματες συσκευές (π.χ. συσκευές 802.11b/g, ασύρματα τηλέφωνα, φούρνοι μικροκυμάτων) που μοιράζονται το ίδιο φάσμα, είναι απαραίτητη για την επίτευξη υψηλής επίδοσης. Για το σκοπό αυτό το Bluetooth χρησιμοποιεί Προσαρμοστική Αναπήδηση Συχνότητας (Adaptive Frequency Hopping - AFH)[40] η οποία αναγνωρίζει τις μεταδόσεις στην ISM ζώνη και αποφεύγει τις συχνότητές τους. Έτσι, η παρεμβολή στενής ζώνης μπορεί να αποφευχθεί και να επιτευχθεί είτε καλύτερη απόδοση ως προς το ποσοστό λαθών (Bit Error Rate, BER), είτε μείωση της ισχύος εκπομπής των τερματικών.

Η ανίχνευση φάσματος αποτελεί μέρος του αλγορίθμου προσαρμοστικής αναπήδησης συχνότητας για τη μέτρηση των στατιστικών του διαύλου με βάση τα οποία καθορίζεται αν ένα κανάλι είναι κατειλημμένο ή όχι. Τα στατιστικά αυτά μπορεί να είναι το ποσοστό λανθασμένων πακέτων (Packet Error Rate, PER), το ποσοστό λανθασμένων ψηφίων (Bit Error Rate, BER), 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.

2.5.3 IEEE 802.22

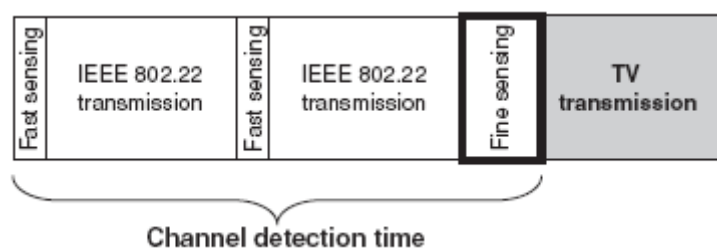
Το πρότυπο IEEE 802.22 WRAN (Wide Regional Area Network) είναι γνωστό ως το πρώτο πρότυπο γνωστικών επικοινωνιών, λόγω των γνωστικών χαρακτηριστικών που διαθέτει. Όπως αναφέρθηκε και στο Κεφάλαιο 1, το πρότυπο βρίσκεται ακόμα σε στάδιο ανάπτυξης και αναφέρεται στην ευκαιριακή χρήση του τηλεοπτικού φάσματος από δευτερεύον σύστημα μετάδοσης δεδομένων για παροχή ευρυζωνικών υπηρεσιών κυρίως σε αγροτικές περιοχές. Η κατάταξη των τηλεοπτικών υπηρεσιών που ως πρωτεύουσας υπηρεσίας έχει γίνει ώστε το πρότυπο να επωφελείται από τις πολύ καλές ιδιότητες διάδοσης σημάτων σε τηλεοπτικές συχνότητες UHF που επιτρέπουν την κάλυψη μεγάλων αποστάσεων.

Η ανίχνευση φάσματος αποτελεί θεμελιώδες τμήμα του IEEE 802.22 για τον εντοπισμό των σημάτων από τους πομπούς τηλεόρασης. Το πρότυπο αναμένεται να υποστηρίζει διάφορες τεχνικές για ταχεία και αποδοτική ανίχνευση. Ανάμεσα σε αυτές είναι η ανίχνευση ενέργειας, η ανίχνευση κυματομορφής, ανίχνευση χαρακτηριστικού κυκλοστατικότητας και το προσαρμοσμένο φίλτρο.

Στο IEEE 802.22, τόσο ο σταθμός βάσης όσο και οι τερματικές συσκευές πραγματοποιούν περιοδική ανίχνευση φάσματος. Για την επικοινωνία με το σταθμό βάσης χρησιμοποιείται κατευθυντική κεραία, ενώ για τη φασματική ανίχνευση ομοιοκατευθυντική. Ο σταθμός βάσης αναθέτει σε κάθε τερματικό σταθμό ανίχνευση καναλιών και αυτοί αποστέλλουν τα αποτελέσματα της ανίχνευσης πίσω στον σταθμό βάσης για τη δημιουργία ενός χάρτη φασματικής χρήσης για κάθε κυψέλη. Το πρωτόκολλο MAC πρέπει να έχει σχεδιαστεί έτσι ώστε να υποστηρίζει τις απαραίτητες λειτουργίες διαχείρισης φάσματος όπως, για παράδειγμα, η εναλλαγή καναλιών (channel switching), η παύση/συνέχιση ενός καναλιού και η περίληψη ή αφαίρεση καναλιών από τη λίστα καναλιών προς εκπομπή.

Η ανίχνευση σημάτων πρωτογενών χρηστών γίνεται σε δύο στάδια: την ταχεία (fast) και τη λεπτομερή (fine) ανίχνευση όπως φαίνεται στο σχήμα 2.9. Η ταχεία ανίχνευση (π.χ. ανίχνευση ενέργειας) πραγματοποιείται για ένα μικρό χρονικό διάστημα προκειμένου να περιοριστεί η παρεμβολή στην εκπομπή δεδομένων. Τα αποτελέσματα της ανίχνευσης αυτής χρησιμοποιούνται για την ανάλυση της χρήσης των καναλιών. Αν σε ένα κανάλι εντοπιστεί υψηλή ενέργεια εκπομπής, εφαρμόζεται η λεπτομερής ανίχνευση. Η λεπτομερής ανίχνευση περιλαμβάνει ενδελεχή ανίχνευση, κατά την οποία χρησιμοποιούνται ισχυρότερες μέθοδοι. Αρκετές τεχνικές που έχουν προταθεί περιλαμβάνουν ανίχνευση κυματομορφής, ανίχνευση χαρακτηριστικού κυκλοστατικότητας και προσαρμοσμένο φίλτρο. Ο αλγόριθμος λεπτομερούς ανίχνευσης απαιτεί περισσότερο χρόνο (μέχρι 25 ms) για την αναγνώριση των στοιχείων συγκεκριμένου σήματος της πρωτογενούς υπηρεσίας. Οι μηχανισμοί ταχείας και λεπτομερούς ανίχνευσης πρέπει να καλύπτουν τις απαιτήσεις λειτουργίας του 802.22, που σημαίνει ότι πρέπει να μπορούν να εντοπίζουν τηλεοπτικά σήματα ισχύος τουλάχιστον 116dBm με πιθανότητα τουλάχιστον 0.9. Επιπλέον, μια πρωτογενής

συσκευή πρέπει να εντοπίζεται εντός δύο δευτερολέπτων από τη στιγμή που ξεκινά να χρησιμοποιεί ένα κανάλι.



Σχήμα 2.9: Μηχανισμοί γρήγορης και λεπτομερούς ανίχνευσης του IEEE 802.22

Σύμφωνα με το IEEE 802.22 η ανίχνευση φάσματος μπορεί να είναι είτε κεντρική είτε κατανεμημένη. Στην κεντρική προσέγγιση οι σταθμοί βάσης είναι εφοδιασμένοι με GPS (Global Positioning System) για να μπορούν να αναφέρουν τη θέση τους. Δεδομένης της γνώσης της θέσης, ένας κεντρικός ελεγκτής μπορεί να υπολογίζει τις απαιτήσεις εκπομπής για κάθε σταθμό βάσης και να κατανέμει τις φασματικές ζώνες κατάλληλα. Εναλλακτικά, κατά την κατανεμημένη προσέγγιση, κάθε σταθμός βάσης ανιχνεύει το φάσμα και εντοπίζει τις διαθέσιμες φασματικές ζώνες τοπικά.

2.6 Πεδία έρευνας στην ανίχνευση φάσματος

Οι υπάρχουσες τεχνικές ανίχνευσης φάσματος εμφανίζουν συγκεκριμένες αδυναμίες. Συνεπώς, εξακολουθούν να υπάρχουν πολλά ανοικτά θέματα προς διερεύνηση προκειμένου να επιτευχθεί αξιόπιστη ανίχνευση φάσματος.

- *Μέτρηση θερμοκρασίας παρεμβολής:* Η δυσκολία αυτού του μοντέλου εντοπισμού δέκτη οφείλεται στην ανάγκη αποδοτικής μέτρησης της θερμοκρασίας παρεμβολής. Ένας γνωστικός χρήστης γνωρίζει τα επίπεδα ισχύος εκπομπής του καθώς και την ακριβή του τοποθεσία με τη χρήση GPS (Global Positioning System). Παρά το γεγονός αυτό, μπορεί να προκαλέσει σημαντική παρεμβολή σε ένα γειτονικό δέκτη που λειτουργεί στην ίδια

συχνότητα. Επιπλέον, δεν υπάρχει πρακτικός τρόπος για ένα γνωστικό χρήστη να μετρήσει την θερμοκρασία παρεμβολής σε γειτονικούς πρωτογενείς δέκτες. Επίσης, επειδή οι δέκτες συνήθως είναι παθητικά στοιχεία ένας γνωστικός χρήστης δεν μπορεί να γνωρίζει την ακριβή τοποθεσία τους. Αυτό αποτελεί σημαντική αδυναμία ανίχνευσης καθώς, αν οι δευτερογενείς χρήστες δεν μπορούν να υπολογίσουν την επίδραση της παρεμβολής τους σε όλους τους πρωτογενείς δέκτες, οι μετρήσεις θερμοκρασίας παρεμβολής χάνουν τη χρησιμότητά τους.

- *Ανίχνευση φάσματος σε δίκτυα πολλαπλών χρηστών:* Συνήθως τα δίκτυα γνωστικών ραδιοεπικοινωνιών εδρεύουν σε περιβάλλοντα πολλαπλών χρηστών τα οποία αποτελούνται από πολλαπλούς δευτερογενείς και πρωτογενείς χρήστες. Επιπλέον, τα δευτερογενή δίκτυα μπορεί να συνυπάρχουν με άλλα δευτερογενή δίκτυα και να ανταγωνίζονται για την ίδια φασματική περιοχή. Τα υπάρχοντα μοντέλα ανίχνευσης φάσματος δεν λαμβάνουν υπόψη την επίδραση της συνύπαρξης πολλαπλών δευτερογενών χρηστών. Στη περίπτωση αυτή, γίνεται δυσκολότερη η ανίχνευση των πρωτογενών χρηστών και ο υπολογισμός της πραγματικής παρεμβολής. Συνεπώς, πρέπει να αναπτυχθούν λειτουργίες ανίχνευσης που θα λαμβάνουν υπόψη την ύπαρξη πολλαπλών δευτερογενών χρηστών.
- *Δυνατότητα εντοπισμού:* Μία από τις κύριες απαιτήσεις των δικτύων νέας γενιάς είναι ο εντοπισμός των πρωτογενών χρηστών σε σύντομο χρονικό διάστημα. Ένας τρόπος μείωσης του συνολικού χρόνου ανίχνευσης είναι η ανίχνευση πολλαπλών φερουσών (multi-carrier sensing) η οποία καθίσταται δυνατή με τη χρήση OFDM. Αν σε μια φέρουσα εντοπιστεί ένας πρωτογενής χρήστης, η ανίχνευση στις υπόλοιπες δεν είναι απαραίτητη.

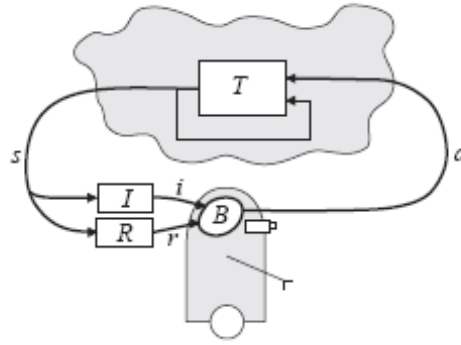
ΚΕΦΑΛΑΙΟ 3: ΕΝΙΣΧΥΤΙΚΗ ΜΑΘΗΣΗ

Ενισχυτική μάθηση (reinforcement learning) ορίζεται το πρόβλημα που αντιμετωπίζει ένας λογικός πράκτορας¹ προκειμένου να εκπαιδευτεί σε μια συγκεκριμένη συμπεριφορά μέσω αλληλεπιδράσεων δοκιμής-λάθους με ένα δυναμικό περιβάλλον. Αποτελεί βασικό τρόπο εκπαίδευσης πρακτόρων σε προβλήματα συμπεριφοράς που βασίζονται στις παρατηρήσεις και την αλληλεπίδραση με το περιβάλλον και στηρίζεται στην διαρκή ενίσχυση του πράκτορα μέσω ανταμοιβών και ποινών ώστε να βελτιστοποιεί δυναμικά τη λειτουργία του προσαρμοζόμενος στις μεταβολές του περιβάλλοντος του.

3.1 Κλασικό μοντέλο ενισχυτικής μάθησης

Σύμφωνα με το κλασικό μοντέλο ενισχυτικής μάθησης, ένας πράκτορας αλληλεπιδρά με το περιβάλλον του μέσω αντιλήψεων και ενεργειών όπως φαίνεται στο σχήμα 3.1. Σε κάθε βήμα αλληλεπίδρασης με το περιβάλλον ο πράκτορας λαμβάνει ως είσοδο, i , κάποια ένδειξη της κατάστασης του περιβάλλοντος και, στη συνέχεια, επιλέγει μια ενέργεια, a , η οποία αποτελεί και την έξοδό του στο περιβάλλον. Η ενέργεια που επιλέγει ο πράκτορας μεταβάλλει την κατάσταση του περιβάλλοντος και η αξία που προκύπτει από τη μεταβολή της κατάστασης μεταβιβάζεται στον πράκτορα μέσω τιμών ενός βαθμωτού ενισχυτικού σήματος r . Η συμπεριφορά του πράκτορα, B , πρέπει να επιλέγει τις ενέργειες που οδηγούν, μακροπρόθεσμα, στη μεγιστοποίηση του αθροίσματος των τιμών ενίσχυσης (ανταμοιβών). Ο πράκτορας μαθαίνει με το πέρασμα του χρόνου μέσω δοκιμών και λαθών.

¹ Πράκτορας ή ορθολογικός πράκτορας (rational agent) είναι ένα σύστημα που ενεργεί ώστε να επιτύχαινε το καλύτερο αποτέλεσμα ή, όταν υπάρχει αβεβαιότητα, το καλύτερο αναμενόμενο αποτέλεσμα.



Σχήμα 3.1: Γενικό μοντέλο ενισχυτικής μάθησης

Σύμφωνα με τα προηγούμενα 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 των πραγματικών αριθμών

Το σχήμα 3.1 περιλαμβάνει και μια συνάρτηση εισόδου I , η οποία καθορίζει πώς ο πράκτορας αντιλαμβάνεται την κατάσταση του περιβάλλοντος. Η αντίληψη που έχει ο πράκτορας για το περιβάλλον του μπορεί να είναι είτε μερική είτε πλήρης, όπως φαίνεται ακολούθως. Ένας διαισθητικός τρόπος ώστε να γίνει αντιληπτή η σχέση του πράκτορα και του περιβάλλοντος του φαίνεται από τον υποθετικό διάλογο του σχήματος 3.2.

Environment: You are in state 65. You have 4 possible actions.
Agent: I'll take action 2.
Environment: You received a reinforcement of 7 units. You are now in state 15. You have 2 possible actions.
Agent: I'll take action 1.
Environment: You received a reinforcement of -4 units. You are now in state 65. You have 4 possible actions.
Agent: I'll take action 2.
Environment: You received a reinforcement of 5 units. You are now in state 44. You have 5 possible actions.
 ⋮ ⋮

Σχήμα 3.2: Υποθετικός διάλογος πράκτορα και περιβάλλοντος

Σύμφωνα το διάλογο του σχήματος 3.2, ο πράκτορας ενημερώνεται από το περιβάλλον για την κατάσταση του και τις διαθέσιμες ενέργειες. Στη συνέχεια επιλέγει μια από αυτές και λαμβάνει την αντίστοιχη ανταμοιβή ενώ παράλληλα ενημερώνεται για τη νέα κατάσταση που βρίσκεται και τις διαθέσιμες ενέργειες που αντιστοιχούν σε αυτή.

Στόχος του πράκτορα είναι, αντιστοιχίζοντας ενέργειες σε καταστάσεις, να προσδιορίσει μία πολιτική(στρατηγική) π η οποία μεγιστοποιεί κάποιο μακροπρόθεσμο μέτρο ενίσχυσης. Γενικά θεωρείται ότι το περιβάλλον στο οποίο βρίσκεται ο πράκτορας είναι στοχαστικό και ότι η εκτέλεση της ίδιας ενέργειας στην ίδια κατάσταση σε δύο διαφορετικές περιπτώσεις μπορεί να οδηγήσει σε διαφορετικές επόμενες καταστάσεις ή διαφορετικές τιμές ενίσχυσης. Αυτό φαίνεται και από το παράδειγμα του σχήματος 3.2, όπου σε δύο περιπτώσεις που ο πράκτορας βρίσκεται στην κατάσταση 65 και επιλέγει την ενέργεια 2, λαμβάνει διαφορετικές ανταμοιβές και μεταβαίνει σε διαφορετικές επόμενες καταστάσεις. Ωστόσο, θεωρείται στατικό περιβάλλον, δηλαδή οι πιθανότητες μεταβάσεων κατάστασης ή λήψης συγκεκριμένων ενισχυτικών σημάτων δεν μεταβάλλονται με το χρόνο.

Αξίζει να σημειωθεί ότι η ενισχυτική μάθηση διαφέρει από την ευρύτερα διαδεδομένη επιβλεπόμενη μάθηση (supervised learning). Η σημαντικότερη διαφορά είναι ότι στη ενισχυτική μάθηση δεν υπάρχουν ζευγάρια εισόδου – εξόδου. Αντιθέτως, ύστερα από την επιλογή μια ενέργειας ο πράκτορας μαθαίνει ποιά θα είναι η άμεση ανταμοιβή και η ακόλουθη κατάσταση αλλά όχι ποιά ενέργεια θα ήταν η καλύτερη μακροπρόθεσμα. Είναι υποχρεωτικό για τον πράκτορα να αποκτήσει αρκετή εμπειρία σχετικά με τις πιθανές καταστάσεις του συστήματος, τις ενέργειες, τις μεταβιβάσεις και τις αντίστοιχες ανταμοιβές ώστε να λειτουργήσει ιδεατά.

3.2 Στοιχεία ενισχυτικής μάθησης

Εκτός του περιβάλλοντος και του πράκτορα, διακρίνονται τέσσερα βασικά στοιχεία ενός συστήματος ενισχυτικής μάθησης: η πολιτική, η συνάρτηση

ανταμοιβής, η συνάρτηση χρησιμότητας και προαιρετικά, το μοντέλο του περιβάλλοντος.

Η πολιτική (*policy*) $\pi(s)$ ορίζει τον τρόπο συμπεριφοράς του πράκτορα σε δεδομένη στιγμή. Αποτελεί μια αντιστοίχιση των καταστάσεων που αντιλαμβάνεται ο πράκτορας σε ενέργειες που θα επιλεγούν. Σε ορισμένες περιπτώσεις η τακτική είναι μια απλή συνάρτηση ή ένας πίνακας αναζήτησης, ενώ σε άλλες περιπτώσεις συνεπάγεται εκτεταμένους υπολογισμούς. Η τακτική ή πολιτική αποτελεί βασικό στοιχείο του πράκτορα ενισχυτικής μάθησης, υπό την έννοια ότι αρκεί μόνη της για τον καθορισμό της συμπεριφοράς του. Γενικά, η πολιτική που ακολουθείται από ένα πράκτορα δεν είναι ντετερμινιστική αλλά στοχαστική διαδικασία.

Η συνάρτηση ανταμοιβής (*reward function*) $R(s)$ ορίζει το στόχο του προβλήματος ενισχυτικής μάθησης. Αντιστοιχίζει κάθε ζεύγος κατάστασης-ενέργειας σε μια τιμή ανταμοιβής που ποσοτικοποιεί την επίπτωση της συγκεκριμένης επιλογής. Ο πράκτορας μαθαίνει να εκτελεί εκείνες τις ενέργειες που μακροπρόθεσμα μεγιστοποιούν το άθροισμα των ανταμοιβών. Η συνάρτηση ανταμοιβής καθορίζει τις θετικές και αρνητικές επιπτώσεις για τον πράκτορα, μπορεί να λαμβάνει θετικές ή αρνητικές τιμές και δεν τροποποιείται από τον ίδιο τον πράκτορα. Μπορεί, όμως, να αποτελεί βάση για αλλαγή της τακτικής που ακολουθείται.

Αν και η συνάρτηση ανταμοιβής υποδεικνύει τι είναι καλό σε μια άμεση ενέργεια, η συνάρτηση χρησιμότητας (*value function*) $V\pi(s)$ προσδιορίζει τι θα είναι καλό μακροπρόθεσμα, λαμβάνοντας υπόψη τις καταστάσεις και τις ανταμοιβές που πιθανόν να ακολουθήσουν. Δηλαδή η χρησιμότητα μιας κατάστασης είναι το συνολικό ποσό ανταμοιβών που αναμένεται να συγκεντρωθεί στο μέλλον, ξεκινώντας από τη συγκεκριμένη κατάσταση. Για παράδειγμα, μια κατάσταση ενδεχομένως να αποδίδει χαμηλή τιμή ανταμοιβής αλλά να έχει υψηλή χρησιμότητα καθώς ακολουθείται από καταστάσεις που αποδίδουν υψηλές ανταμοιβές, ή το αντίστροφο. Προφανώς, οι ακολουθίες καταστάσεων εξαρτώνται από την τακτική (πολιτική) που εκτελείται και για το λόγο αυτό η χρησιμότητα ορίζεται σε σχέση με συγκεκριμένη πολιτική. Γενικά, οι επιλογές των ενεργειών

πραγματοποιούνται με βάση κριτήρια χρησιμότητας. Αναζητούνται ενέργειες που θα φέρουν καταστάσεις υψηλής χρησιμότητας (και όχι αναγκαστικά υψηλής ανταμοιβής), αφού τέτοιες ενέργειες οδηγούν μακροπρόθεσμα σε βέλτιστο αποτέλεσμα. Δυστυχώς, είναι πολύ δυσκολότερο να καθοριστεί η χρησιμότητα μιας κατάστασης σε σχέση με τον προσδιορισμό της άμεσης ανταμοιβής της. Οι ανταμοιβές καθορίζονται άμεσα από το περιβάλλον, ενώ οι χρησιμότητες πρέπει να υπολογιστούν και να επαναυπολογιστούν βάσει σειράς παρατηρήσεων που ο πράκτορας κάνει σε όλη τη διάρκεια της λειτουργίας του. Στην πραγματικότητα, το σημαντικότερο συστατικό της ενισχυτικής μάθησης είναι ο τρόπος αποτελεσματικού υπολογισμού των χρησιμοτήτων.

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 υπάρχει ένας σταθερός χρόνος N μετά τον οποίο δεν έχει τίποτε σημασία, για παράδειγμα ο πράκτορας σταματά να 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 του πεπερασμένου ορίζοντα η βέλτιστη ενέργεια σε μια 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

$$V([s_0, s_1, s_2, \dots]) = R(s_0) + R(s_1) + R(s_2) + \dots \quad (3.1)$$

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

$$V([s_0, s_1, s_2, \dots]) = R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \dots \quad (3.2)$$

όπου ο παράγοντας ελάττωσης (discount factor) γ είναι ένας αριθμός μεταξύ του 0 και του 1. Ο παράγοντας ελάττωσης περιγράφει την προτίμηση του πράκτορα για βραχυπρόθεσμες ανταμοιβές έναντι μακροπρόθεσμων ανταμοιβών. Ειδικά όταν το

γ βρίσκεται κοντά στο 0, οι ανταμοιβές του μακρινού μέλλοντος θεωρούνται αμελητέες.

Το τέταρτο και τελευταίο στοιχείο ενός συστήματος ενισχυτικής μάθησης είναι το μοντέλο (*model*), που ουσιαστικά προσομοιώνει τη συμπεριφορά του περιβάλλοντος. Για παράδειγμα, δεδομένης μιας κατάστασης και μιας ενέργειας, το μοντέλο μπορεί να προβλέψει την επόμενη κατάσταση και την επόμενη ανταμοιβή. Τα μοντέλα χρησιμοποιούνται για προγραμματισμό, δηλαδή για την απόφαση μιας σειράς ενεργειών λαμβάνοντας υπόψη πιθανές μελλοντικές καταστάσεις. Υπάρχουν τεχνικές ενισχυτικής μάθησης που στηρίζονται στη χρήση μοντέλου (*model-based*) και άλλες που στηρίζονται αποκλειστικά στην έρευνα μέσω δοκιμών και λαθών.

3.3 Μέτρηση επίδοσης μάθησης

Τα κριτήρια που αναφέρθηκαν προηγουμένως μπορούν να χρησιμοποιηθούν για την αξιολόγηση των πολιτικών που προέκυψαν από ένα δεδομένο αλγόριθμο. Για την αξιολόγηση της ποιότητας της εκπαίδευσης χρησιμοποιούνται διαφορετικά κριτήρια. Μερικά από τα κριτήρια αυτά αναφέρονται ακολούθως.

3.3.1 Τελική σύγκλιση στη βέλτιστη λύση

Υπάρχουν αλγόριθμοι που εγγυώνται ασυμπτωτική σύγκλιση στη βέλτιστη συμπεριφορά. Αυτό αποτελεί εχέγγυο για την εύρεση της βέλτιστης συμπεριφοράς αλλά πολλές φορές είναι πρακτικά περιττό. Ένας πράκτορας που προσεγγίζει γρήγορα ένα επίπεδο της τάξης του 99% της βέλτιστης συμπεριφοράς είναι πιθανό σε πολλές εφαρμογές να είναι προτιμότερος από ένα πράκτορα που εγγυάται πλήρη σύγκλιση στη βέλτιστη συμπεριφορά αλλά με ένα αργό αρχικό ρυθμό μάθησης.

3.3.2 Ταχύτητα σύγκλισης προς τη βέλτιστη λύση

Η βέλτιστη λύση είναι συνήθως ένα ασυμπτωτικό αποτέλεσμα και συνεπώς η ταχύτητα σύγκλισης αποτελεί ένα ασαφές μέτρο επίδοσης. Ένα περισσότερο πρακτικό μέτρο επίδοσης είναι η ταχύτητα σύγκλισης στη σχεδόν βέλτιστη λύση (speed of convergence to near-optimality). Το μέτρο αυτό εξαρτάται από τον καθορισμό της ικανοποιητικής προσέγγισης στη βέλτιστη λύση. Ένα άλλο σχετικό μέτρο είναι το επίπεδο επίδοσης μετά από ένα ορισμένο χρονικό διάστημα (level of performance after a given time) που εξαρτάται επίσης από τον καθορισμό του χρονικού διαστήματος.

Τα μέτρα που σχετίζονται με την ταχύτητα εκμάθησης έχουν μια πρόσθετη αδυναμία. Ένας αλγόριθμος που απλώς προσπαθεί να πετύχει τη βέλτιστη λύση όσο το δυνατόν ταχύτερα μπορεί να οδηγήσει σε περιττές μεγάλες ποινές κατά τη διάρκεια της εκμάθησης. Μία λιγότερο επιθετική στρατηγική που απαιτεί περισσότερο χρόνο για την επίτευξη της βέλτιστης συμπεριφοράς αλλά οδηγεί σε μεγαλύτερη συνολική ενίσχυση του πράκτορα κατά τη διάρκεια της εκμάθησης μπορεί να είναι προτιμητέα.

3.3.3 Κόστος

Ένα καταλληλότερο μέτρο επίδοσης είναι η αναμενόμενη μείωση της ανταμοιβής που οφείλεται στην εκτέλεση του αλγορίθμου ενισχυτικής μάθησης έναντι της βέλτιστης συμπεριφοράς από την αρχή. Το μέτρο αυτό είναι γνωστό ως κόστος (regret). Μειονέκτημα αυτού του μέτρου επίδοσης αποτελεί το γεγονός ότι είναι δύσκολο να αποκτηθούν τα αποτελέσματα που αφορούν τα κόστη των αλγορίθμων.

3.4 Αντιστάθμιση εκμετάλλευσης με έρευνα

Μία από τις προκλήσεις που προκύπτουν από την ενισχυτική μάθηση είναι η αντιστάθμιση μεταξύ έρευνας (exploration) και εκμετάλλευσης (exploitation). Προκειμένου να πετύχει υψηλή ανταμοιβή, ο πράκτορας πρέπει να προτιμά

ενέργειες που έχουν δοκιμαστεί στο παρελθόν και αποδείχθηκαν αποδοτικές. Ωστόσο, για να ανακαλυφθούν τέτοιες ακολουθίες ενεργειών, πρέπει να δοκιμάζονται νέες ενέργειες που στο παρελθόν δεν έχουν επιλεγεί. Δηλαδή ο πράκτορας οφείλει να εκμεταλλεύεται γνώσεις του παρελθόντος για την απόκτηση υψηλής ανταμοιβής αλλά, ταυτόχρονα, πρέπει να ερευνά έτσι ώστε να πραγματοποιήσει ακόμα καλύτερες ενέργειες στον μέλλον. Το πρόβλημα είναι ότι ούτε η εξερεύνηση ούτε η εκμετάλλευση θα οδηγήσουν στη βέλτιστη λύση, αν επιδιωχθούν από μόνες τους αποκλειστικά. Ο πράκτορας πρέπει να δοκιμάζει μια ποικιλία ενεργειών και με την πάροδο να προτιμά αυτές που αποδίδουν καλύτερα. Συμπερασματικά, στην ενισχυτική μάθηση οι πράκτορες αντιμετωπίζουν δύο αντικρουόμενους στόχους. Πρέπει να επιδιώκουν ως πρωταρχικό στόχο την όσο το δυνατόν υψηλότερη ανταμοιβή σε δεδομένο περιβάλλον. Ωστόσο, αν ο πράκτορας επιλέγει ενέργειες που θεωρεί βέλτιστες, χωρίς να εξετάζει άλλες επιλογές για πιθανή υψηλότερη ανταμοιβή, είναι πιθανό να οδηγηθεί σε χαμηλότερη της δυνατής απόδοση.

3.5 Στρατηγικές ενισχυτικής μάθησης

- *Άπληστη στρατηγική:* Σύμφωνα με την άπληστη στρατηγική, ο πράκτορας επιλέγει πάντα την ενέργεια με τη μεγαλύτερη αναμενόμενη ανταμοιβή. Το μειονέκτημα της στρατηγικής αυτής είναι ότι, σε περίπτωση ατυχούς αρχικής παρατήρησης του περιβάλλοντος, η ανταμοιβή της βέλτιστης ενέργειας μπορεί να φαίνεται μικρότερη από την ανταμοιβή μιας υποβέλτιστης ενέργειας. Έτσι, ο πράκτορας θα επιλέγει πάντα την υποβέλτιστη ενέργεια χωρίς να ανακαλύψει ποτέ τη βέλτιστη ενέργεια. Για την αποφυγή τέτοιων αποτελεσμάτων ο πράκτορας πρέπει να παρατηρεί το περιβάλλον διαρκώς.
- *Τυχαιοποιημένη στρατηγική:* Μια ακόμη απλή στρατηγική εξερεύνησης είναι η λήψη της ενέργειας με τη μεγαλύτερη αναμενόμενη ανταμοιβή ως προεπιλογή, αλλά με πιθανότητα p να επιλέγεται μια ενέργεια στην τύχη. Προκειμένου να ενθαρρύνουν την αρχική εξερεύνηση του περιβάλλοντος, κάποιες εκδοχές της στρατηγικής αυτής ξεκινούν με μεγάλη τιμή p η οποία σταδιακά μειώνεται. Μία ένσταση για τη στρατηγική αυτή είναι ότι, όταν ο

πράκτορας πειραματίζεται με μη άπληστο τρόπο, είναι εξίσου πιθανό να επιλεγεί μια υποσχόμενη εναλλακτική με το να επιλεγεί μια μη υποσχόμενη. Μια ελαφρώς εξελιγμένη μέθοδος που αντισταθμίζει το πρόβλημα είναι η εξερεύνηση Boltzmann (Boltzmann exploration). Στην περίπτωση αυτή η αναμενόμενη ανταμοιβή για την ενέργεια a , $ER(a)$ χρησιμοποιείται για την επιλογή ενέργειας ακολουθώντας την κατανομή

$$P(a) = \frac{e^{ER(a)/T}}{\sum_{a' \in A} e^{ER(a')/T}} \quad (3.3)$$

Η παράμετρος T μπορεί να μειωθεί με το πέρασμα του χρόνου για να μειωθεί η εξερεύνηση. Η μέθοδος αυτή λειτουργεί καλά όταν η βέλτιστη ενέργεια είναι ξεκάθαρα διαχωρισμένη από τις υπόλοιπες, αλλά δεν αποδίδει καλά όταν οι χρησιμότητες των ενεργειών παρουσιάζουν μικρές διαφορές. Επίσης μπορεί να συγκλίνει με καθυστέρηση εάν δεν ρυθμιστεί κατάλληλα η παράμετρος T .

3.6 Καθυστερημένη ανταμοιβή

Στη γενική περίπτωση του προβλήματος της ενισχυτικής μάθησης, οι ενέργειες του πράκτορα δεν καθορίζουν απλώς την άμεση ανταμοιβή αλλά και την επόμενη κατάσταση του περιβάλλοντος (τουλάχιστον πιθανοτικά). Έτσι, ο πράκτορας πρέπει να λαμβάνει υπόψη του εκτός από την άμεση ανταμοιβή, και την επόμενη κατάσταση του περιβάλλοντος, όταν αποφασίζει ποιά ενέργεια θα επιλέξει. Ο τρόπος με τον οποίο λαμβάνεται υπόψη η αξία του μέλλοντος καθορίζεται από το μοντέλο μακροπρόθεσμης βελτιστοποίησης που ακολουθείται. Ο πράκτορας οφείλει να μαθαίνει από καθυστερημένη ενίσχυση (delayed reinforcement): ενδεχομένως μπορεί να χρειαστεί μια μακρά ακολουθία ενεργειών, όπου θα λαμβάνει ασήμαντη ενίσχυση (δηλαδή μικρή ανταμοιβή), μέχρι τελικά να φθάσει σε κατάσταση υψηλής ενίσχυσης.

3.7 Διαδικασία απόφασης Markov (MDP)

Τα προβλήματα με καθυστερημένη ενίσχυση για ένα πλήρως παρατηρήσιμο περιβάλλον περιγράφονται αποτελεσματικά ως διαδικασίες απόφασης Markov (Markov Decision process-MDP). Μια διαδικασία απόφασης Markov ορίζεται από τις ακόλουθες συνιστώσες:

- ένα σύνολο καταστάσεων² S
- ένα σύνολο ενεργειών A
- μία συνάρτηση ανταμοιβής $R : S \times A \rightarrow \mathbb{R}$, και
- και μία συνάρτηση μετάβασης καταστάσεων $T : S \times A \rightarrow \Pi(S)$

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

3.8 Εύρεση πολιτικής για δεδομένο μοντέλο

Θεωρώντας ότι για τη λήψη αποφάσεων υπάρχει άπειρος ορίζοντας (infinite-horizon) και ότι η χρησιμότητα μιας ακολουθίας καταστάσεων γίνεται μέσω προεξοφλημένων ανταμοιβών (discounted rewards), η βέλτιστη χρησιμότητα μιας κατάστασης γράφεται

$$V^*(s) = \max_{\pi} E(\sum_{t=0}^{\infty} \gamma^t r_t) \quad (3.4)$$

Η βέλτιστη συνάρτηση χρησιμότητας είναι μοναδική και μπορεί να οριστεί ως λύση της επόμενης σχέσης που εκφράζει ότι η βέλτιστη χρησιμότητα μιας κατάστασης s είναι ίση με το άθροισμα της αναμενόμενης άμεσης ανταμοιβής και της αναμενόμενης χρησιμότητας της επόμενης κατάστασης s' , εκτελώντας τη βέλτιστη ενέργεια.

² Αν και γενικά μπορεί να υπάρχει άπειρος αριθμός καταστάσεων, θα συζητηθεί κυρίως η επίλυση προβλημάτων πεπερασμένου αριθμού καταστάσεων.

$$V^*(s) = \max_a (R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V^*(s')), \forall s \in S \quad (3.5)$$

Δεδομένης συνεπώς της βέλτιστης συνάρτησης χρησιμότητας, η βέλτιστη τακτική προσδιορίζεται από τη σχέση

$$\pi^*(s) = \operatorname{argmax}_a (R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V^*(s')) \quad (3.6)$$

3.8.1 Επανάληψη αξιών (value iteration)

Ένας τρόπος προσδιορισμού της βέλτιστης πολιτικής είναι ο υπολογισμός της βέλτιστης συνάρτησης χρησιμότητας με τη βοήθεια του αλγορίθμου επανάληψη αξιών. Η βασική ιδέα είναι να υπολογιστεί η χρησιμότητα κάθε κατάστασης με βάση τη σχέση (3.5), που ονομάζεται και εξίσωση Bellman, 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. Επειδή ο τελεστής “max” δεν είναι γραμμικός, το σύστημα εξισώσεων είναι μη γραμμικό και μπορεί να λυθεί μέσω επαναληπτικής διαδικασίας προσέγγισης που τερματίζεται όταν επιτευχθεί ισορροπία στις τιμές δύο διαδοχικών συναρτήσεων χρησιμότητας. Η περιγραφή του αλγορίθμου μέσω ψευδοκώδικα δίνεται ακολούθως.

```
Initialize  $V(s)$  arbitrarily
loop until policy good enough
  loop for  $s \in S$ 
    loop for  $a \in A$ 
       $Q(s, a) := R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V(s')$ 
     $V(s) := \max_a Q(s, a)$ 
  end loop
end loop
```

Δεν είναι φανερό πότε πρέπει να σταματήσει ο αλγόριθμος επανάληψη αξιών. Ένα αποτελεσματικό κριτήριο διακοπής του αλγορίθμου προτάθηκε από τους 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 που μεγιστοποιεί την αναμενόμενη προεξοφλημένη ανταμοιβή, δεν διαφέρει από τη χρησιμότητα της βέλτιστης πολιτικής περισσότερο από $2\epsilon\gamma/(1-\gamma)$ σε οποιαδήποτε κατάσταση. Ένας εναλλακτικός τρόπος ανανέωσης της $Q(s, a)$ αντί της χρήσης της σχέσης $Q(s, a) := R(s, a) + \gamma \sum_{s' \in S} T(s, a, s')V(s')$ (full backups) είναι η χρήση της σχέσης $Q(s, a) := Q(s, a) + \alpha(r + \gamma \max_{a'} Q(s', a') - Q(s, a))$ (sample backup). Ωστόσο, η εφαρμογή της σχέσης sample backup προϋποθέτει ότι τα ζεύγη τιμών a, s ανανεώνονται συνεχώς και ανά τακτά χρονικά διαστήματα, η s' δειγματοληπτείται από την κατανομή $T(s, a, s')$, η r δειγματοληπτείται με μέση τιμή τη συνάρτηση ανταμοιβής $R(s, a)$ και φραγμένη διασπορά, ενώ ο ρυθμός μάθησης α μειώνεται σταδιακά. Αποδεικνύεται ότι η τιμή χρησιμότητας συγκλίνει με πιθανότητα 1 στη βέλτιστη χρησιμότητα $V^*(s)$, εφόσον κάθε ενέργεια εκτελείται σε κάθε κατάσταση άπειρες φορές σε ένα απείρου ορίζοντα πρόβλημα και η παράμετρος α μειώνεται σταδιακά.

Η υπολογιστική πολυπλοκότητα του αλγορίθμου επανάληψης αξιών με full backups, ανά επανάληψη, είναι τετραγωνική όσο αφορά το πλήθος των καταστάσεων και γραμμική όσο αφορά το πλήθος των ενεργειών. Συνήθως ο πίνακας πιθανοτήτων μετάβασης $T(s, a, s')$ είναι αραιός. Αν, κατά μέσο όρο, υπάρχει ένας σταθερός αριθμός επόμενων καταστάσεων με μη μηδενική πιθανότητα μετάβασης, το κόστος κάθε επανάληψης είναι γραμμικό ως προς το πλήθος των καταστάσεων καθώς και το πλήθος των ενεργειών. Ο απαιτούμενος αριθμός επαναλήψεων για 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 μεγαλώνει πολυωνυμικά σε $1/(1-\gamma)$ και ο ρυθμός σύγκλισης μειώνεται αισθητά καθώς ο παράγοντας προεξόφλησης πλησιάζει το 1.

3.8.2 Επανάληψη πολιτικών (policy iteration)

Ένας εναλλακτικός τρόπος εύρεσης βέλτιστων τακτικών (πολιτικών) είναι ο αλγόριθμος επανάληψης πολιτικών ο οποίος χειραγωγεί άμεσα την πολιτική χωρίς να κάνει χρήση της βέλτιστης συνάρτησης χρησιμότητας. Ξεκινώντας από μια αρχική

πολιτική π_0 , η βασική ιδέα του αλγορίθμου είναι η εναλλαγή των ακόλουθων δύο βημάτων.

- Αξιολόγηση πολιτικής (policy evaluation): Δεδομένης μιας πολιτικής π_i , υπολογίζεται το V_i , δηλαδή η χρησιμότητα για κάθε κατάσταση i αν πρόκειται να εκτελεστεί η πολιτική π_i .
- Βελτίωση πολιτικής: Υπολογίζεται μια νέα πολιτική π_{i+1} , χρησιμοποιώντας προεξέταση ενός βήματος με βάση τη V_i .

Ο αλγόριθμος τερματίζεται όταν το βήμα βελτίωσης της πολιτικής δεν παράγει μεταβολή των χρησιμοτήτων. Η περιγραφή του αλγορίθμου με ψευδοκώδικα δίνεται ακολούθως.

Choose an arbitrary policy π'

loop

$\pi := \pi'$

compute the value function of policy π :

solve the linear equations

$$V_{\pi}(s) = R(s, \pi(s)) + \gamma \sum_{s' \in S} T(s, \pi(s), s') V_{\pi}(s')$$

improve the policy at each state:

$$\pi'(s) := \operatorname{argmax}_a (R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V_{\pi}(s'))$$

until $\pi = \pi'$

Καθώς υπάρχουν το πολύ $|A|^{|S|}$ διαφορετικές πολιτικές και η ακολουθία των πολιτικών βελτιώνεται σε κάθε βήμα, ο αριθμός επαναλήψεων στη συνήθη περίπτωση είναι εκθετικός. Ωστόσο, παραμένει άγνωστο το πλήθος των επαναλήψεων που χρειάζεται ο αλγόριθμος επανάληψης πολιτικών στη χειρότερη περίπτωση.

Στην πράξη, ο αλγόριθμος επανάληψης αξιών εκτελεί συντομότερες χρονικά επαναλήψεις ενώ ο αλγόριθμος επανάληψης πολιτικών χρειάζεται λιγότερες επαναλήψεις. Για την επιτάχυνση των ανωτέρω αλγορίθμων έχουν προταθεί στη βιβλιογραφία διάφορες τροποποιημένες προσεγγιστικές τεχνικές³.

³ Modified policy iteration algorithm (Puterman & Shin, 1978), multigrid methods (Rude, 1993), state aggregation (Bertsekas & Castanon, 1989).

3.9 Μερικώς παρατηρήσιμες διαδικασίες απόφασης Markov (POMDP)

Κατά την περιγραφή των διαδικασιών αποφάσεων Markov που παρουσιάστηκε προηγουμένως έγινε η υπόθεση ότι το περιβάλλον είναι πλήρως παρατηρήσιμο (fully observable). Με βάση την υπόθεση αυτή ο πράκτορας γνωρίζει πάντα σε ποιά κατάσταση βρίσκεται. Σε συνδυασμό με την υπόθεση Markov για το μοντέλο μετάβασης, αυτό σημαίνει ότι η βέλτιστη πολιτική εξαρτάται μόνο από την τρέχουσα κατάσταση. Όμως, όταν το περιβάλλον είναι μερικώς παρατηρήσιμο (partially observable), η κατάσταση είναι λιγότερο σαφής. Ο πράκτορας δεν γνωρίζει απαραίτητως σε ποιά κατάσταση βρίσκεται και, έτσι, δεν μπορεί να πραγματοποιήσει την ενέργεια $\pi(s)$ που είναι η συνιστώμενη για αυτή τη κατάσταση. Επιπλέον, η χρησιμότητα μιας κατάστασης s και η βέλτιστη ενέργεια στη συγκεκριμένη κατάσταση δεν εξαρτώνται μόνο από την s αλλά και από το μέγεθος της γνώσης του πράκτορα όταν βρίσκεται στη συγκεκριμένη κατάσταση. Για τους λόγους αυτούς, οι μερικώς παρατηρήσιμες MDP (partially observable MDP ή POMDP) θεωρούνται συνήθως πολυπλοκότερες από τις κανονικές MDP.

Σε πολλά περιβάλλοντα του πραγματικού κόσμου, όπως και στην περίπτωση των γνωστικών ραδιοεπικοινωνιών, το σύστημα δεν είναι δυνατό να έχει πλήρη αντίληψη της κατάστασης του περιβάλλοντος. Ο πράκτορας βασίζεται σε παρατηρήσεις που δεν περιγράφουν πλήρως την κατάσταση του περιβάλλοντος και περιέχουν ενδεχομένως θόρυβο.

Μια διαδικασία POMDP έχει τα ίδια στοιχεία με μια MDP, δηλαδή το μοντέλο μετάβασης $T(s, a, s')$ και τη συνάρτηση ανταμοιβής $R(s)$. Όμως, έχει επίσης ένα μοντέλο παρατήρησης (observation model) $O(s, o)$ που προσδιορίζει την πιθανότητα να γίνει αντιληπτή η παρατήρηση o στην κατάσταση s . Στην πραγματικότητα, ο πράκτορας δεν αντιλαμβάνεται την κατάσταση του περιβάλλοντος αλλά λαμβάνει μια παρατήρηση που είναι συνάρτηση της τρέχουσας κατάστασης. Αν η παρατήρηση ταυτίζεται με την κατάσταση του περιβάλλοντος, το μοντέλο είναι αντίστοιχο με το πλήρως παρατηρήσιμο MDP. Το

μοντέλο POMDP χαρακτηρίζεται από τη συνάρτηση μετάβασης $P(S_{t+1}|S_t, A_t)$, τη συνάρτηση παρατήρησης $P(O_t|S_t, A_{t-1})$ και τη συνάρτηση ανταμοιβής $E(R_t|S_t, A_{t-1})$. Ο στόχος του πράκτορα είναι να μάθει την πολιτική π που αντιστοιχίζει το ιστορικό παρατηρήσεων σε ενέργειες A_t , μεγιστοποιώντας την αξία χρησιμότητας.

$$h_{1:t} = ((O_1, A_1, R_1), \dots, (O_{t-1}, A_{t-1}, R_{t-1}), (O_t, -, -)) \quad (3.7)$$

Δεδομένου ότι ο πράκτορας έχει περιορισμένη χωρητικότητα μνήμης, πρέπει με κάποιο τρόπο να συμπιεστεί το ιστορικό $h_{1:t}$ σε μια εσωτερική κατάσταση. Ορίζεται, συνεπώς, το διάνυσμα κατάστασης πεποίθησης (belief state) $b(t) = P(S_t|h_{1:t})$ που αποτελεί μια κατανομή πιθανοτήτων ως προς όλες τις δυνατές καταστάσεις. Στη συνέχεια, θα χρησιμοποιηθεί ο συμβολισμός $b(s)$ για την πιθανότητα που αποδίδεται στην πραγματική κατάσταση s από το διάνυσμα κατάστασης πεποίθησης $b(t)$. Αν $b(s)$ ήταν η προηγούμενη κατάσταση πεποίθησης και ο πράκτορας εκτέλεσε μια ενέργεια a λαμβάνοντας την παρατήρηση o , η νέα κατάσταση πεποίθησης είναι

$$b'(s') = nO(s', o) \sum_s T(s, a, s') b(s) \quad (3.8)$$

όπου n σταθερά κανονικοποίησης ώστε το άθροισμα των πιθανοτήτων των καταστάσεων πεποίθησης να είναι ίσο με 1.

Η ουσιώδης παρατήρηση που απαιτείται προκειμένου να γίνει κατανοητή η διαδικασία POMDP είναι ότι η βέλτιστη ενέργεια εξαρτάται μόνο από την τρέχουσα κατάσταση πεποίθησης του πράκτορα. Αυτό σημαίνει ότι η βέλτιστη τακτική μπορεί να περιγραφεί ως μια απεικόνιση $\pi^*(b)$ που απεικονίζει τις καταστάσεις πεποίθησης στις ενέργειες και δεν εξαρτάται από την πραγματική κατάσταση στην οποία βρίσκεται ο πράκτορας (αφού δεν την γνωρίζει πλήρως). Έτσι, ο κύκλος απόφασης ενός πράκτορα POMDP είναι ο εξής:

- Με δεδομένη την τρέχουσα κατάσταση πεποίθησης b , εκτέλεσε την ενέργεια $a = \pi^*(b)$.
- Λάβε την παρατήρηση o .

- Προσδιόρισε την τρέχουσα κατάσταση πεποίθησης μέσω της σχέσης

$$b'(s') = \eta O(s', o) \sum_s T(s, a, s') b(s) \text{ και επανάλαβε.}$$

Ο υπολογισμός της πιθανότητας «ο πράκτορας που βρίσκεται στην κατάσταση b να φθάσει στην κατάσταση b' μετά την εκτέλεση της ενέργειας a » δεν μπορεί άμεσα να γίνει, αφού δεν είναι ακόμη γνωστή η επακόλουθη παρατήρηση που θα ληφθεί. Η πιθανότητα ο πράκτορας να λάβει την παρατήρηση o με δεδομένο ότι εκτέλεσε την ενέργεια a ξεκινώντας από την κατάσταση πεποίθησης b , υπολογίζεται αθροίζοντας όλες τις πραγματικές καταστάσεις στις οποίες μπορεί να καταλήξει ο πράκτορας:

$$\begin{aligned} P(o|a, b) &= \sum_{s'} P(o|a, s', b) P(s'|a, b) \\ &= \sum_{s'} O(s', o) P(s'|a, b) \\ &= \sum_{s'} O(s', o) \sum_s T(s, a, s') b(s) \end{aligned} \quad (3.9)$$

Συνεπώς, η πιθανότητα μετάβασης από το b στο b' , δεδομένης της ενέργειας a , που συμβολίζεται με $\tau(b, a, b')$ προκύπτει από τη σχέση

$$\begin{aligned} \tau(b, a, b') &= P(b'|a, b) = \sum_o P(b'|o, a, b) P(o|a, b) \\ &= \sum_o P(b'|o, a, b) \sum_{s'} O(s', o) \sum_s T(s, a, s') b(s) \end{aligned} \quad (3.10)$$

όπου $P(b'|o, a, b)$ είναι 1 αν $b' = \eta O(s', o) \sum_s T(s, a, s') b(s)$ και 0 σε κάθε άλλη περίπτωση.

Η ανωτέρω εξίσωση μπορεί να θεωρηθεί ότι ορίζει ένα μοντέλο μετάβασης για το χώρο καταστάσεων πεποίθησης. Αν οριστεί, επίσης, μια συνάρτηση ανταμοιβής για τις καταστάσεις πεποίθησης

$$\rho(b) = \sum_s b(s) R(s) \quad (3.11)$$

φαίνεται ότι τα $\tau(b, a, b')$ και $\rho(b)$ μαζί ορίζουν μια παρατηρήσιμη MDP στο χώρο των καταστάσεων πεποίθησης. Δηλαδή η επίλυση μιας POMDP σε ένα φυσικό χώρο καταστάσεων μπορεί να αναχθεί στην επίλυση μιας MDP στον αντίστοιχο χώρο

καταστάσεων πεποίθησης. Αξίζει να αναφερθεί ότι η διαδικασία MDP που προκύπτει έχει συνεχή χώρο καταστάσεων. Ωστόσο, είναι δυνατό να αναπτυχθούν εκδοχές των αλγορίθμων επανάληψης αξιών και πολιτικών που είναι εφαρμόσιμες και σε MDP με συνεχείς καταστάσεις. Η βασική ιδέα είναι ότι μια πολιτική $\pi(b)$ μπορεί να παρασταθεί ως ένα σύνολο περιοχών στο χώρο των καταστάσεων πεποίθησης, εκάστη των οποίων σχετίζεται με συγκεκριμένη βέλτιστη ενέργεια. Η συνάρτηση χρησιμότητας συσχετίζει μια διακριτή γραμμική συνάρτηση του b με κάθε περιοχή. Κάθε βήμα της επανάληψης αξιών ή πολιτικών αποσαφηνίζει τα όρια μεταξύ των περιοχών και μπορεί να εισάγει νέες περιοχές.

3.10 Εύρεση πολιτικής χωρίς τη χρήση μοντέλου

Στα προηγούμενα έγινε αναφορά σε μεθόδους εύρεσης της βέλτιστης πολιτικής για διαδικασίες Markov (MDP) θεωρώντας δεδομένη την ύπαρξη ενός μοντέλου. Το μοντέλο περιλαμβάνει τη γνώση του μοντέλου μετάβασης $T(s, a, s')$ και της συνάρτησης ανταμοιβής $R(s, a)$. Όμως, ο βασικός στόχος της ενισχυτικής μάθησης είναι η εύρεση της βέλτιστης πολιτικής όταν ένα τέτοιο μοντέλο δεν είναι γνωστό εξ αρχής. Ο πράκτορας πρέπει να αλληλεπιδράσει με το περιβάλλον άμεσα προκειμένου να αποκτήσει πληροφορίες οι οποίες μέσω ενός κατάλληλου αλγορίθμου μπορούν να υποστούν επεξεργασία ώστε να βρεθεί η βέλτιστη πολιτική.

Με βάση τα ανωτέρω οι μέθοδοι ενισχυτικής μάθησης κρίνονται σε δύο κατηγορίες:

- Μέθοδοι με χρήση μοντέλου: όπου πρώτα δημιουργείται ένα μοντέλο και στη συνέχεια χρησιμοποιείται για την εκπαίδευση του πράκτορα.
- Μέθοδοι χωρίς τη χρήση μοντέλου: όπου ο πράκτορας εκπαιδεύεται χωρίς να μάθει κάποιο μοντέλο.

3.10.1 Εκμάθηση Q

Η εκμάθηση Q (Q learning) είναι μία τεχνική ενισχυτικής μάθησης που στηρίζεται στην εκμάθηση μιας συνάρτησης ενέργειας-αξίας (action-value function) που δίνει την αναμενόμενη χρησιμότητα που λαμβάνει ο πράκτορας από την επιλογή μιας δεδομένης ενέργειας σε μια δεδομένη κατάσταση και εν συνεχεία

ακολουθώντας μια σταθερή πολιτική π . Πλεονέκτημα την εκμάθησης Q είναι ότι έχει τη δυνατότητα να συγκρίνει την αναμενόμενη χρησιμότητα από τις διαθέσιμες ενέργειες χωρίς την απαίτηση ύπαρξης μοντέλου για το περιβάλλον.

Το πρόβλημα της εκμάθησης Q αποτελείται από

- ένα πράκτορα
- ένα πλήθος καταστάσεων S
- και ένα πλήθος ενεργειών A για κάθε κατάσταση

Ο πράκτορας μεταβαίνει από μία κατάσταση σε μία άλλη εκτελώντας μια ενέργεια $a \in A$. Κάθε κατάσταση προσφέρει στον πράκτορα μία ανταμοιβή (θετικός πραγματικός ή φυσικός αριθμός) ή μια ποινή (αρνητικός αριθμός). Σκοπός του πράκτορα είναι η μεγιστοποίηση της συνολικής ανταμοιβής. Ο αλγόριθμος διατηρεί εκτιμήσεις Q -τιμών για κάθε ζεύγος κατάστασης-ενέργειας (s, a) όπου $Q : S \times A \rightarrow \mathbb{R}$.

Για κάθε χρονική στιγμή $t(=1,2,\dots)$, ορίζεται ως $Q^*(s, a)$ η αναμενόμενη προεξοφλημένη ανταμοιβή μιας ενέργειας a σε μια κατάσταση s και $V^*(s)$ την χρησιμότητα της κατάστασης s . Θεωρώντας ότι η βέλτιστη ενέργεια έχει πραγματοποιηθεί αρχικά προκύπτει ότι $V^*(s) = \max_a Q^*(s, a)$. Συνεπώς, η $Q^*(s, a)$ γράφεται αναδρομικά

$$Q^*(s, a) = R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') \max_{a'} Q^*(s', a') \quad (3.12)$$

Ο πράκτορας ακολουθεί άπληστη πολιτική με βάση τις εκτιμήσεις του. Αυτό σημαίνει ότι αν την χρονική στιγμή t ο πράκτορας βρεθεί στην κατάσταση s η επόμενη ενέργεια που θα επιλέξει θα είναι η

$$a' := \operatorname{argmax}_{a \in A} Q^*(s, a) \quad (3.13)$$

Επίσης, σημειώνεται ότι εφόσον ισχύει $V^*(s) = \max_a Q^*(s, a)$ η βέλτιστη πολιτική προκύπτει από τη σχέση

$$\pi^*(s) = \operatorname{argmax}_a Q^*(s, a) \quad (3.14)$$

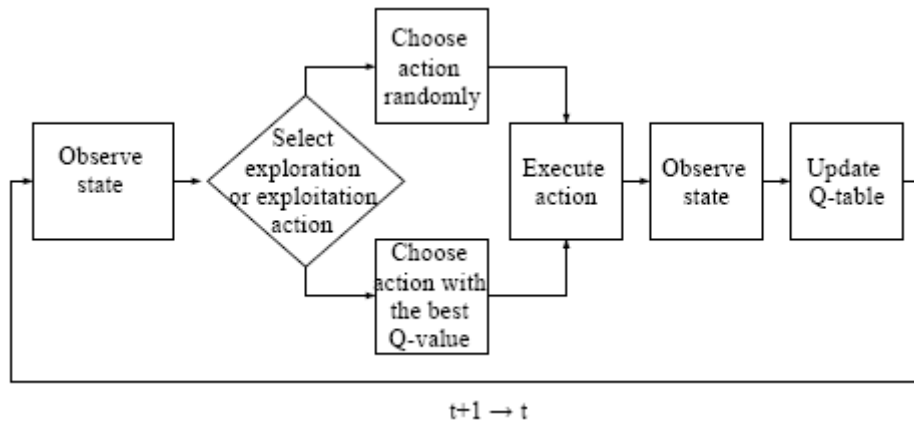
Πριν ξεκινήσει η εκμάθηση, η συνάρτηση Q έχει μια προεπιλεγμένη τιμή, η οποία καθορίζεται από το σχεδιαστή του συστήματος. Στη συνέχεια, κάθε φορά που ο πράκτορας λαμβάνει μια ανταμοιβή υπολογίζονται νέες τιμές Q για κάθε ζεύγος κατάστασης-ενέργειας. Ουσιαστικά, δηλαδή, ο πυρήνας του αλγορίθμου είναι η ενημέρωση μέσω επανάληψης αξιών. Αρχικά ο πράκτορας θεωρεί δεδομένες τις προηγούμενες τιμές και κάνει διορθώσεις βασισμένος στις νέες πληροφορίες που λαμβάνει. Το διάγραμμα λειτουργίας της εκμάθησης Q φαίνεται στο σχήμα 3.3. Ο βασικός κανόνας της εκμάθησης Q δίνεται από τη σχέση

$$Q(s_t, a_t) := Q(s_t, a_t) + \alpha_t(s_t, a_t)(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a) - Q(s_t, a_t)) \quad (3.15)$$

όπου $Q(s_t, a_t)$ η παλιά τιμή της συνάρτησης Q , r_{t+1} η ανταμοιβή κατά τη χρονική στιγμή $t+1$, $\max_{a'} Q(s_{t+1}, a)$ η μέγιστη μελλοντική τιμή, $\alpha_t(s_t, a_t)$ ο ρυθμός μάθησης ($0 < \alpha_t(s_t, a_t) \leq 1$), και 0γ ο τελεστής ελάττωσης ($0 \leq \gamma < 1$). Η (3.15) μπορεί να γραφεί και υπό τη μορφή:

$$Q(s_t, a_t) := Q(s_t, a_t)(1 - \alpha_t(s_t, a_t)) + \alpha_t(s_t, a_t)[r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a)] \quad (3.16)$$

Ο ρυθμός μάθησης καθορίζει σε ποιο βαθμό η νεοαποκτηθείσα πληροφορία θα υπερिशύσει της παλιάς. Ρυθμός που είναι ίσος με μηδέν έχει ως αποτέλεσμα ο πράκτορας να μη μαθαίνει τίποτα, ενώ ρυθμός που είναι ίσος με τη μονάδα έχει σαν αποτέλεσμα ο πράκτορας να λαμβάνει υπόψη του μόνο τη νέα πληροφορία. Ο συντελεστής μείωσης καθορίζει τη σημασία των μελλοντικών ανταμοιβών. Αν είναι ίσος με μηδέν ο πράκτορας λαμβάνει υπόψη του μόνο τις τρέχουσες ανταμοιβές ενώ, αν η τιμή του πλησιάζει τη μονάδα, ο πράκτορας ενδιαφέρεται για υψηλή μακροπρόθεσμη ανταμοιβή. Για τιμές ίσες ή μεγαλύτερες της μονάδας ο αλγόριθμος δεν συγκλίνει. Η ποσότητα $r + \gamma \max_{a'} Q(s', a')$ αντιστοιχεί στην αναμενόμενη προεξοφλημένη ανταμοιβή (expected discounted award).



Σχήμα 3.3: διάγραμμα λειτουργίας εκμάθησης Q

Η τεχνική εκμάθησης Q συγκλίνει στη βέλτιστη πολιτική με πιθανότητα 1 εφόσον όλα τα ζεύγη κατάστασης-ενέργειας ενημερώνονται συνεχώς⁴. Όταν οι Q τιμές βρίσκονται κοντά στη σύγκλιση με τις βέλτιστες τιμές τους, ο πράκτορας μπορεί να ενεργεί άπληστα, επιλέγοντας σε κάθε κατάσταση την ενέργεια με την μεγαλύτερη Q τιμή. Κατά τη διάρκεια της εκμάθησης, ωστόσο, πρέπει να υπάρχει συμβιβασμός μεταξύ εξερεύνησης και εκμετάλλευσης. Επιπλέον, η τεχνική εκμάθησης Q δεν παρουσιάζει ευαισθησία εξερεύνησης (exploration insensitive), γεγονός που σημαίνει ότι οι τιμές Q θα συγκλίνουν στη βέλτιστη λύση ανεξάρτητα του τρόπου συμπεριφοράς του πράκτορα κατά τη διάρκεια της εκμάθησης. Αυτό σημαίνει ότι παρά το γεγονός ότι το θέμα εξερεύνησης-εκμετάλλευσης αφορά την εκμάθηση Q, οι λεπτομέρειες της στρατηγικής εξερεύνησης δεν επηρεάζουν τη σύγκλιση του αλγορίθμου στη βέλτιστη λύση. Για το λόγο αυτό, η τεχνική εκμάθησης Q αποτελεί τον πλέον διαδεδομένο και αποτελεσματικότερο αλγόριθμο εκμάθησης με καθυστερημένη ενίσχυση που δεν στηρίζεται στη χρήση μοντέλου. Μειονέκτημα της αποτελεί το ότι μπορεί να συγκλίνει με μεγάλη καθυστέρηση στη βέλτιστη πολιτική.

Η περιγραφή του αλγορίθμου της εκμάθησης Q με ψευδοκώδικα δίνεται ακολούθως

Inputs: $\gamma, S, A, m, \epsilon_1$

⁴ Watkins 1989, Tsitsikis 1994, Jaakkola, Jordan & Singh 1994

```

For all  $(s, a)$  do
     $Q(s, a) \leftarrow 1/(1 - \gamma)$ 
     $U(s, a) \leftarrow 0$ 
     $l(s, a) \leftarrow 0$ 
     $t(s, a) \leftarrow 0$ 
     $Learn(s, a) \leftarrow true$  //fdfd
End for
 $t^* \leftarrow 0$  // time of most recent
For  $t = 1, 2, 3, \dots$  do
    Let  $s$  denote the state at time  $t$ 
    Choose action  $a := \operatorname{argmax}_{a \in A} Q(s, a')$ 
    Let  $r$  be the immediate award and  $s'$  the next
    state after executing action  $a$  from state  $s$ 
    If  $Learn(s, a) = true$  then
         $U(s, a) \leftarrow U(s, a) + r + \gamma \max_{a'} Q(s', a')$ 
         $l(s, a) \leftarrow l(s, a) + 1$ 
        If  $l(s, a) = m$  then
            If  $Q(s, a) - U(s, a)/m \geq 2\epsilon_1$  then
                 $Q(s, a) \leftarrow U(s, a)/m + \epsilon_1$ 
                 $t^* \leftarrow t$ 
            Else if  $t(s, a) \geq t^*$ 
                 $Learn(s, a) \leftarrow false$ 
            End if
             $t(s, a) \leftarrow t, U(s, a) \leftarrow 0, l(s, a) \leftarrow 0$ 
        End if
    Else if  $t(s, a) < t^*$  then
         $Learn(s, a) \leftarrow false$ 
    End if
End for

```

Ο αλγόριθμος πέρα από τις Q τιμές διατηρεί και μια λογική σημαία $Learn(s, a)$ για κάθε ζεύγος (s, a) η οποία υποδεικνύει αν ο πράκτορας πραγματοποιεί κάποια αλλαγή στην εκτίμηση του για το $Q(s, a)$. Επίσης, ο αλγόριθμος στηρίζεται σε δύο ελεύθερες παραμέτρους, την $\epsilon_1 \in (0, 1)$ και την m που είναι θετικός ακέραιος. Τέλος, ο αλγόριθμος διατηρεί και ένα μετρητή $l(s, a)$ του

οποίου η τιμή δηλώνει το πλήθος των δεδομένων (πλήθος δειγμάτων) που απαιτούνται για να ανανεωθεί η τιμή της $Q(s, a)$. Όταν συγκεντρωθούν m δείγματα και η τιμή της λογική σημαίας $Learn(s, a)$ είναι αληθής (true), ο πράκτορας προχωρά σε ανανέωση της $Q(s, a)$.

3.10.2 Εφαρμογή εκμάθησης Q σε συστήματα γνωστικών ραδιοεπικοινωνιών

Οι γνωστικές ραδιοεπικοινωνίες αποτελούν ένα νέο μέσο ασύρματης επικοινωνίας που δίνει τη δυνατότητα σε μη αδειοδοτημένους χρήστες να εντοπίζουν και να χρησιμοποιούν το υποχρησιμοποιούμενο αδειοδοτημένο φάσμα ευκαιριακά όταν αυτό δεν χρησιμοποιείται από τους αδειοδοτημένους χρήστες. Με βάση τα προηγούμενα δημιουργείται το πρόβλημα της εύρεσης της βέλτιστης στρατηγικής για την επιλογή ενός διαθέσιμου αδειοδοτημένου καναλιού για την επικοινωνία μεταξύ των δευτερογενών χρηστών, δεδομένου ότι ο στόχος είναι η μεγιστοποίηση της συνολικής διέλευσης (throughput) και η ελαχιστοποίηση της καθυστέρησης μετάδοσης. Τα ανωτέρω μπορούν να επιτευχθούν χρησιμοποιώντας μια προσέγγιση ενισχυτικής μάθησης για τη δημιουργία μοντέλου για το σύστημα δυναμικής πρόσβασης φάσματος. Η εφαρμογή ενισχυτικής μάθησης στη δυναμική πρόσβαση φάσματος προσφέρει πολλά πλεονεκτήματα καθώς μπορεί να βοηθήσει ένα δευτερογενή χρήστη να προσαρμοστεί ταχέως στο περιβάλλον λειτουργίας του. Επίσης, μέσω της ενισχυτικής μάθησης μειώνεται η πολυπλοκότητα της περιγραφής του περιβάλλοντος και της ανομοιογένειας των καναλιών. Για παράδειγμα, ένας δευτερογενής χρήστης που επιλέγει ένα κανάλι για εκπομπή δεν χρειάζεται να μοντελοποιεί τη συμπεριφορά του καναλιού, η οποία μπορεί να χαρακτηρίζεται από επιλεκτική σκίαση, απώλειες πολλαπλών διαδρομών, παρεμβολή των πρωτογενών χρηστών ή άλλους παράγοντες που επηρεάζουν την απόδοση ενός δευτερογενούς χρήστη. Αντί να εντοπίζει ένα παράγοντα κάθε φορά, ένας πράκτορας ενισχυτικής μάθησης βλέπει τους σχετικούς παράγοντες κατά τη διαδικασία απόφασης ως κατάσταση και βελτιστοποιεί ένα γενικό στόχο, όπως για παράδειγμα τη ρυθμαπόδοση (throughput), μέσω της μεγιστοποίησης της συνάρτησης χρησιμότητας $V^\pi(s_t)$. Στη συνέχεια, ακολουθεί παράδειγμα

εφαρμογής ενισχυτικής μάθησης Q σε σύστημα γνωστικών ραδιοεπικοινωνιών που προτάθηκε από τους Kok-Lim Alvin Yau, Peter Komisarczuk και Paul D. Teal[[44].

Το μοντέλο ενισχυτικής μάθησης που χρησιμοποιήθηκε φαίνεται στον πίνακα 3.1. Σύμφωνα με το μοντέλο αυτό, ο μηχανισμός ενισχυτικής μάθησης ενσωματώνεται στο δευτερογενή σταθμό βάσης, ο οποίος είναι υπεύθυνος για την επιλογή του καναλιού που θα χρησιμοποιηθεί για τη μετάδοση προς τους δευτερογενείς χρήστες. Η κατάσταση S_c ενός κόμβου i έχει μια μόνο συνιστώσα, το πλήθος των γειτονικών κόμβων j που συμβολίζεται με N . Καθώς ο μηχανισμός ενισχυτικής μάθησης είναι ενσωματωμένος στο σταθμό βάσης, ο κόμβος i αντιστοιχεί στο σταθμό βάσης και ο j στο δευτερογενή χρήστη. Η κατάσταση του περιβάλλοντος μεταβάλλεται με το χρόνο, ενδεχομένως λόγω της αλλαγής απόστασης μεταξύ του σταθμού βάσης και του χρήστη ή της μεταβολής του ρυθμού λανθασμένων πακέτων (packet error rate-PER). Η πιθανότητα επιτυχημένης εκπομπής εξαρτάται από τη χρησιμοποίηση του καναλιού από τους πρωτογενείς χρήστες, το PER και τις συγκρούσεις πακέτων. Ως ενέργεια A_c θεωρείται η επιλογή ενός καναλιού για μετάδοση από ένα σύνολο διαθέσιμων καναλιών K , δηλαδή $a_c = \{c_1, c_2, \dots, c_k\}$. Για κάθε επιτυχημένη μετάδοση πακέτου υπάρχει μια ανταμοιβή με σταθερή θετική τιμή RW ενώ για κάθε αποτυχία μετάδοσης υπάρχει ένα κόστος με σταθερή αρνητική τιμή CT . Η μετάδοση ενός πακέτου θεωρείται επιτυχημένη όταν ληφθεί το αντίστοιχο πακέτο επιβεβαίωσης. Επιπλέον, εφόσον ένα κανάλι καταληφθεί ξανά από ένα πρωτεύοντα χρήστη αμέσως πριν τη μετάδοση από ένα δευτερογενή χρήστη θεωρείται ανεπιτυχής μετάδοση.

Σε κάθε προσπάθεια μετάδοσης ενός πακέτου ο δευτερογενής σταθμός βάσης i επιλέγει ένα κανάλι για μετάδοση. Ο μηχανισμός ενισχυτικής μάθησης διατηρεί μια τιμή $Q(s,a)$ για όλες τις πιθανές ενέργειες σε ένα πίνακα Q με $|A|$ εισόδους.

Στοιχείο ενισχυτικής μάθησης	Περιγραφή	Αναπαράσταση
Κατάσταση	Σύνολο γειτονικών κόμβων j του κόμβου i	$S_c = (s_c = j)$ $j = (n_1, n_2, \dots, n_N)$
Ενέργεια	Διαθέσιμα κανάλια για μετάδοση	$A_c = a_c = \{c_1, c_2, \dots, c_k\}$
Ανταμοιβή	Σταθερή τιμή κέρδους/κόστους για επιτυχημένη/αποτυχημένη μετάδοση αντίστοιχα	$R_c = r(s, a)$ $= \begin{cases} +RW, & \text{επιτυχημένη} \\ -CT, & \text{αποτυχημένη} \end{cases}$

Πίνακας 3.1

Η (3.16) ξαναγράφεται υπό τη μορφή

$$Q(s_t, a_t) := Q(s_t, a_t)(1 - \alpha_t(s_t, a_t)) + \alpha_t(s_t, a_t)r_{t+1} \quad (3.17)$$

Στην (3.17) ο όρος $\gamma \max_{a'} Q(s_{t+1}, a)$ έχει παραλειφθεί καθώς δεν υπάρχει εξάρτηση από μελλοντικές ανταμοιβές. Ακλουθώντας την άπληστη στρατηγική η επόμενη ενέργεια επιλέγεται ως εξής :

$$\alpha' := \operatorname{argmax}_{a \in A_c} Q(s_t, a) \quad (3.18)$$

Στην αρχή κάθε προσπάθειας εκπομπής ο δευτερογενής σταθμός βάσης επιλέγει είτε να συνεχίσει είτε να αλλάξει ενέργεια ή κανάλι. Προκειμένου να μειωθεί ο αριθμός εναλλαγών καναλιών ο σταθμός βάσης δεν αλλάζει κανάλι παρά μόνο όταν η τιμή Q μιας άλλης ενέργειας είναι μεγαλύτερη ή βρίσκεται ακόμα στο στάδιο εξερεύνησης.

ΚΕΦΑΛΑΙΟ 4: ΕΦΑΡΜΟΓΗ ΕΝΙΣΧΥΤΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΑΝΙΧΝΕΥΣΗ ΦΑΣΜΑΤΟΣ ΣΕ ΣΥΣΤΗΜΑΤΑ ΓΝΩΣΤΙΚΩΝ ΡΑΔΙΟΕΠΙΚΟΙΝΩΝΙΩΝ

Οι γνωστικές ραδιοεπικοινωνίες έχουν προταθεί ως πιθανή λύση για την αποτελεσματικότερη αξιοποίηση του αδειοδοτημένου φάσματος, το οποίο παρουσιάζει υποχρησιμοποίηση σε ορισμένες περιοχές και σε διάφορες χρονικές περιόδους. Τα συστήματα γνωστικών επικοινωνιών είναι δευτερογενή συστήματα χαμηλότερης προτεραιότητας που αποκτούν πρόσβαση στο φάσμα όταν αυτό δεν χρησιμοποιείται από τους αδειοδοτημένους χρήστες. Παράλληλα, πρέπει να εξασφαλίζεται η προστασία των αδειοδοτημένων χρηστών από παρεμβολές. Για το λόγο αυτό, η ανίχνευση φάσματος αποτελεί θεμελιώδη λειτουργία των γνωστικών ραδιοεπικοινωνιών. Επιπλέον, η αξιόπιστη ανίχνευση δημιουργεί περισσότερες ευκαιρίες για αύξηση της χωρητικότητας των συστημάτων γνωστικών ραδιοεπικοινωνιών και οδηγεί σε υψηλούς ρυθμούς μετάδοσης και κάλυψη υψηλότερων απαιτήσεων QoS. Για την επίτευξη αποτελεσματικής ανίχνευσης έχουν προταθεί στη βιβλιογραφία διάφορες τεχνικές. Στο κεφάλαιο αυτό ερευνάται η εφαρμογή τεχνικών ενισχυτικής μάθησης.

Το ότι η ενισχυτική μάθηση μπορεί να χρησιμοποιηθεί χωρίς την ύπαρξη προηγούμενης γνώσης καθώς και το ότι αποσκοπεί στη βελτιστοποίηση της μακροπρόθεσμης συμπεριφοράς ενός πράκτορα, την καθιστούν κατάλληλη για χρήση σε συστήματα γνωστικών ραδιοεπικοινωνιών. Συγκεκριμένα, ένας μη αδειοδοτημένος χρήστης μπορεί να χρησιμοποιήσει κάποιο αλγόριθμο ενισχυτικής μάθησης προκειμένου να διερευνήσει τις πιθανές στρατηγικές ανίχνευσης και πρόσβασης στο φάσμα ώστε τελικά να επιλέξει αυτή που αυξάνει την πιθανότητα εντοπισμού διαθέσιμων καναλιών και οδηγεί σε υψηλούς ρυθμούς μετάδοσης. Η εφαρμογή της ενισχυτικής μάθησης κρίνεται απαραίτητη ειδικά σε συστήματα μη συνεργατικής ανίχνευσης.

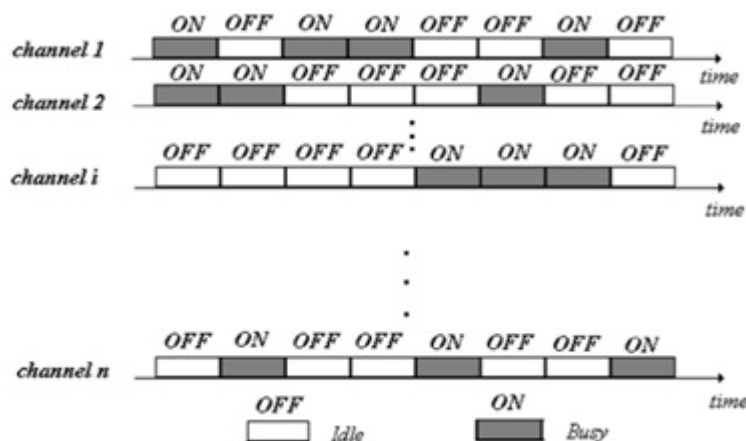
Οι τεχνικές ενισχυτικής μάθησης που θα μελετηθούν είναι ο αλγόριθμος επανάληψης πολιτικών και η εκμάθηση Q. Και οι δύο τεχνικές στηρίζονται σε

διαδικασίες απόφασης Markov (MDP). Η βασική διαφορά τους είναι ότι η πρώτη απαιτεί την ύπαρξη μοντέλου για το περιβάλλον ενώ η δεύτερη όχι. Κριτήριο επίδοσης των αλγορίθμων αποτελεί η πιθανότητα επιτυχημένης ανίχνευσης που προκύπτει ύστερα από την εφαρμογή τους. Τα αποτελέσματα συγκρίνονται με αυτά της τυχαίας ανίχνευσης προκειμένου να διαπιστωθεί η συνεισφορά της ενισχυτικής μάθησης στη διαδικασία ανίχνευσης.

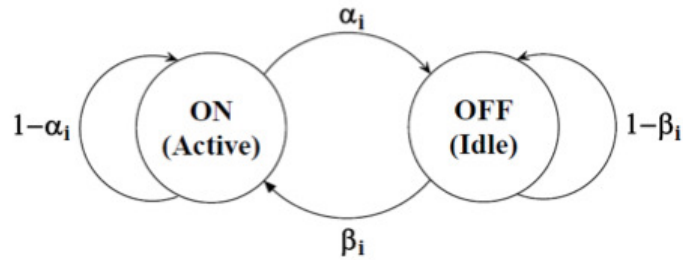
4.1 Διατύπωση μοντέλου για τη χρήση του φάσματος

Θεωρείται ότι το διαθέσιμο αδειοδοτημένο φάσμα αποτελείται από n κανάλια έκαστο των οποίων είναι χωρισμένο σε χρονοσχισμές (time slots) έτσι ώστε η επικοινωνία των πρωτογενών χρηστών να είναι σύγχρονη. Ως μοντέλο χρήσης ενός καναλιού χρησιμοποιήθηκε το μοντέλο κατάστασης ON/OFF, όπως φαίνεται στο σχήμα 4.1. Σύμφωνα με αυτό, μια χρονοσχισμή βρίσκεται σε κατάσταση ON όταν το κανάλι χρησιμοποιείται από κάποιο πρωτογενή χρήστη ή σε κατάσταση OFF όταν είναι ελεύθερο. Επειδή κάθε κανάλι μεταβάλλει την κατάσταση του ανεξάρτητα και η επόμενη κατάσταση ενός καναλιού δεν εξαρτάται από τις προηγούμενες καταστάσεις, το μοντέλο κατάστασης ON/OFF αποτελεί για κάθε κανάλι μια αλυσίδα Markov δύο καταστάσεων όπως αυτή του σχήματος 4.2.

Οι δευτερογενείς χρήστες προσπαθούν να αποκτήσουν πρόσβαση στο αδειοδοτημένο φάσμα όταν αυτό δεν χρησιμοποιείται από τους πρωτογενείς χρήστες. Συνεπώς, επιτρέπεται να μεταδώσουν σε κανάλια που βρίσκονται σε κατάσταση OFF.



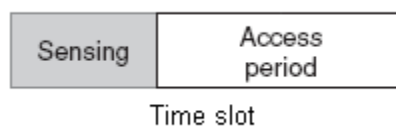
Σχήμα 4.1: Μοντέλο χρήσης φάσματος



Σχήμα 4.2: Αλυσίδα Markov δύο καταστάσεων

4.2 Ανίχνευση και πρόσβαση στο φάσμα

Το μοντέλο δυναμικής πρόσβασης που χρησιμοποιείται είναι μη συνεργατικό και κατανεμημένο, δηλαδή οι δευτερογενείς χρήστες δεν συνεργάζονται για πρόσβαση στο φάσμα αλλά δρουν εγωιστικά. Κάθε δευτερογενής χρήστης πραγματοποιεί ανίχνευση και όταν ανιχνεύσει κάποιο φασματικό κενό (επιτυχημένη ανίχνευση) εκπέμπει με πιθανότητα 1. Κάθε χρήστης μπορεί να ανιχνεύει ένα κανάλι σε κάθε χρονοσχισμή. Αν το κανάλι είναι ελεύθερο, εκπέμπει. Άλλως, περιμένει μέχρι την επόμενη χρονοσχισμή για να ανιχνεύσει κάποιο κανάλι. Σημειώνεται ότι η ανίχνευση και η πρόσβαση στο φάσμα (εφόσον ανιχνευτεί φασματικό κενό) πραγματοποιούνται στην ίδια χρονοσχισμή. Συγκεκριμένα θεωρείται ότι κάθε χρονοσχισμή αποτελείται από μία περίοδο ανίχνευσης και μια περίοδο εκπομπής, όπως φαίνεται στο σχήμα 4.3.



Σχήμα 4.3: Περίοδοι ανίχνευσης και πρόσβασης σε μια χρονοσχισμή

4.3 Συνάρτηση ανταμοιβής

Η συνάρτηση ανταμοιβής αποτελεί βασικό στοιχείο της ενισχυτικής μάθησης καθώς ορίζει το στόχο του προβλήματος. Με βάση τα αποτελέσματα της συνάρτησης ανταμοιβής διαμορφώνεται και η στρατηγική του πράκτορα. Στις

γνωστικές ραδιοεπικοινωνίες, ένας δευτερογενής χρήστης ανανεώνει την πολιτική του για πρόσβαση στο φάσμα ανάλογα με τις τιμές που επιστρέφει η συνάρτηση ανταμοιβής για κάθε ενέργεια του. Στο υπό μελέτη πρόβλημα ως συνάρτηση ανταμοιβής ορίζεται το κέρδος/κόστος που έχει ένας δευτερογενής χρήστης από την ανίχνευση ενός ελεύθερου/κατειλημμένου καναλιού ή το κόστος αν αποφασίσει να μην ανιχνεύσει κάποιο κανάλι.

Η συνάρτηση ανταμοιβής που χρησιμοποιήθηκε είναι η εξής

$$R(s, a) = \begin{cases} wb - c & , \text{εκπομπή} \\ -c_{no} & , \text{μη εκπομπή} \end{cases} \quad (4.1)$$

όπου b είναι το όφελος επιτυχημένη εκπομπής, c το κόστος ανίχνευσης και c_{no} το κόστος απόφασης μη εκπομπής. Εδώ σημειώνεται ότι, όταν ένας χρήστης αποφασίσει να μην εκπέμψει, δεν πραγματοποιεί ανίχνευση φάσματος και επομένως δεν επωμίζεται το κόστος ανίχνευσης.

Στον κατωτέρω πίνακα φαίνονται η τιμές των συντελεστών που χρησιμοποιήθηκαν:

b	c	c_{no}
2	0.5	0.1

Πίνακας 4.1

Η μεταβλητή w αντιστοιχεί στη χρήση του καναλιού κατά την χρονοσχισμή που πραγματοποιείται η ανίχνευση και η τιμή της δίνεται από τη σχέση:

$$w = \begin{cases} 1 & , \text{channel OFF} \\ 0 & , \text{channel ON} \end{cases} \quad (4.2)$$

Όταν το κανάλι είναι ελεύθερο, ο χρήστης εκπέμπει και λαμβάνει το κέρδος εκπομπής, ενώ όταν το κανάλι είναι κατειλημμένο επωμίζεται το κόστος ανίχνευσης. Όταν ο χρήστης αποφασίζει να μην εκπέμψει, τότε, μολονότι δεν επιβαρύνεται με το κόστος ανίχνευσης εξακολουθεί να έχει ένα κόστος που αντιστοιχεί στη μείωση το ρυθμού αποστολής δεδομένων.

4.4 Μοντέλο ενισχυτικής μάθησης

Σύμφωνα με το μοντέλο του συστήματος, ως πράκτορας προς εκπαίδευση ορίζεται ένας δευτερογενής χρήστης. Στόχος της εφαρμογής της ενισχυτικής μάθησης είναι η εύρεση εκείνης της πολιτικής ανίχνευσης φάσματος που οδηγεί σε υψηλότερα ποσοστά ανίχνευσης διαθέσιμων καναλιών (επιτυχημένη ανίχνευση). Στο υπό μελέτη πρόβλημα το μοντέλο της ενισχυτικής μάθησης διατυπώνεται με χρήση διαδικασιών απόφασης Markov, θεωρώντας το περιβάλλον πλήρως παρατηρήσιμο. Οι τεχνικές ενισχυτικής μάθησης που χρησιμοποιούνται είναι ο αλγόριθμος επανάληψης πολιτικών και η εκμάθηση Q.

4.4.1 Επανάληψη πολιτικών

Μία διαδικασία απόφασης Markov που κάνει χρήση μοντέλου για το περιβάλλον (model based) αποτελείται από τις εξής συνιστώσες:

- ένα σύνολο καταστάσεων S
- ένα σύνολο ενεργειών A
- μία συνάρτηση ανταμοιβής $R : S \times A \rightarrow \mathbb{R}$, και
- και μία συνάρτηση μετάβασης καταστάσεων $T : S \times A \rightarrow \prod(S)$

Το σύνολο καταστάσεων S του χρήστη αποτελείται από τα n διαθέσιμα κανάλια και την κατάσταση μη εκπομπής, δηλαδή $|S| = n + 1$. Ο χρήστης βρίσκεται στην κατάσταση i αν το τελευταίο κανάλι το οποίο ανίχνευσε (ανεξάρτητα από το αν εξέπεμψε ή όχι) ήταν το κανάλι i . Το σύνολο των ενεργειών είναι ίδιο με αυτό των καταστάσεων ($|A| = n + 1$), καθώς ως ενέργεια θεωρείται το επόμενο κανάλι προς ανίχνευση ή η απόφαση για μη εκπομπή και συνεπώς μη ανίχνευση. Η συνάρτηση μετάβασης καταστάσεων είναι ένας τρισδιάστατος πίνακας διαστάσεων $S \times S \times A$ που περιέχει τις πιθανότητες μετάβασης από μια κατάσταση s_i σε μια κατάσταση s_j πραγματοποιώντας την ενέργεια a . Επειδή η ενέργεια που θα επιλεγεί ταυτίζεται με την επόμενη κατάσταση, η πιθανότητα μετάβασης από οποιαδήποτε κατάσταση στην κατάσταση που ορίζει η επόμενη ενέργεια είναι ίση με 1. Η συνάρτηση ανταμοιβής είναι επίσης ένας πίνακας διαστάσεων $S \times S \times A$, ο οποίος περιέχει το κέρδος/κόστος από την μετάβαση του πράκτορα από την κατάσταση s_i στην κατάσταση s_j πραγματοποιώντας την ενέργεια a . Ο πίνακας της

συνάρτησης ανταμοιβής συμπληρώνεται με χρήση των τιμών της συνάρτησης ανταμοιβής που περιγράφηκε προηγουμένως ύστερα από την προσομοίωση του συστήματος για κάποιο αριθμό χρονοσχισμών.

4.4.2 Εκμάθηση Q

Εκτός από την τεχνική επανάληψης πολιτικών μελετάται και η εφαρμογή της εκμάθησης Q, η οποία δεν απαιτεί την ύπαρξη μοντέλου για το περιβάλλον. Αυτό σημαίνει ότι δεν υπάρχει ανάγκη για τη δημιουργία των πινάκων συνάρτησης ανταμοιβής και μετάβασης καταστάσεων.

Το πρόβλημα της εκμάθησης Q αποτελείται από:

- ένα πράκτορα
- ένα σύνολο καταστάσεων
- και ένα σύνολο ενεργειών για κάθε κατάσταση

Ο πράκτορας (δευτερογενής χρήστης) μεταβαίνει από μία κατάσταση σε μία άλλη εκτελώντας μια ενέργεια. Κάθε κατάσταση προσφέρει στον πράκτορα μία ανταμοιβή ή κόστος. Το σύνολο των καταστάσεων και των ενεργειών είναι το ίδιο με αυτό που περιγράφηκε προηγουμένως, ενώ η ανταμοιβή(ή κόστος) κάθε κατάστασης προέρχεται από τη συνάρτηση κέρδους (4.1). Τα στοιχεία ενισχυτικής μάθησης συνοψίζονται στον πίνακα 4.2.

Στοιχείο ενισχυτικής μάθησης	Περιγραφή	Αναπαράσταση
Κατάσταση	Κανάλι που ανιχνεύτηκε την χρονοσχισμή t ή κατάσταση μη ανίχνευσης	$S = \{no_tr, i\}, i=1, \dots, n$
Ενέργεια	Κανάλι που θα ανιχνευτεί την χρονοσχισμή $t+1$ ή απόφαση για μη ανίχνευση	$A_c = \{no_tr, i\}, i=1, \dots, n$
Συνάρτηση κέρδους	Τιμή κέρδους/κόστους για επιτυχημένη/αποτυχημένη ανίχνευση ή μη εκπομπή	$R(s, a)$ $= \begin{cases} wb - c & , \text{εκπομπή} \\ -c_{no} & , \text{μη εκπομπή} \end{cases}$

Πίνακας 4.2

Το σύνολο των καταστάσεων και των ενεργειών είναι το ίδιο με αυτό που περιγράφηκε παραπάνω, ενώ η ανταμοιβή(ή κόστος) κάθε κατάστασης προέρχεται από τη συνάρτηση ανταμοιβής. Τα στοιχεία ενισχυτικής μάθησης συνοψίζονται στον πίνακα 4.2.

4.5 Προσομοίωση

Θεωρείται ένα βασικό σενάριο ασύρματης επικοινωνίας μεταξύ ενός δευτερογενούς πομπού και ενός δευτερογενούς δέκτη το οποίο μπορεί να αποτελέσει βάση για διάφορα συστήματα επικοινωνιών. Το διαθέσιμο φάσμα αποτελείται από $n=6$ κανάλια. Τα κανάλια χωρίζονται σε χρονοσχισμές και η χρήση τους ακολουθεί το μοντέλο ON/OFF. Η πρόσβαση στο φάσμα από τους πρωτογενείς χρήστες είναι τυχαία. Ωστόσο, κάποια κανάλια παρουσιάζουν χαμηλότερη χρήση σε σχέση με κάποια άλλα. Στόχος της ενισχυτικής μάθησης είναι ο εντοπισμός αυτών των καναλιών. Οι προσομοιώσεις πραγματοποιήθηκαν για επίπεδα φασματικής χρήσης 20%, 60% και 80%, με στόχο να μελετηθούν τα οφέλη από την εφαρμογή ενισχυτικής μάθησης σε περιβάλλοντα χαμηλής, μέσης και υψηλής φασματικής χρήσης, αντίστοιχα.

Για τη δημιουργία του πίνακα ανταμοιβής της policy iteration προσομοιώθηκε η ανίχνευση φάσματος για 100 χρονοσχισμές, ενώ η προσομοίωση της λειτουργίας του συστήματος διήρκεσε 1000 χρονοσχισμές. Επιπλέον, λόγω της τυχαιότητας που παρουσιάζει η χρήση του φάσματος, τα αποτελέσματα ανίχνευσης παρουσίαζαν μικρές διαφορές σε διαφορετικές προσομοιώσεις. Για το λόγο αυτό, πραγματοποιήθηκαν 100 επαναλήψεις της διαδικασίας και στο τέλος υπολογίστηκαν οι μέσοι όροι.

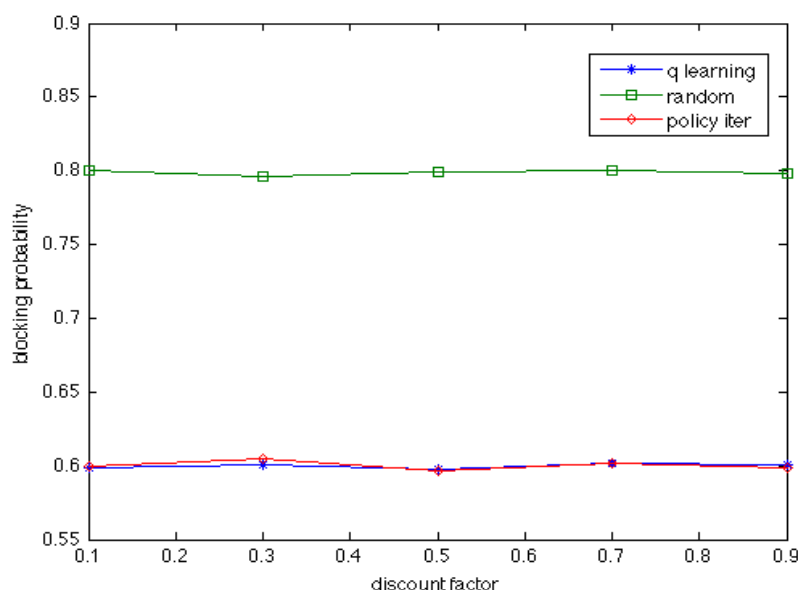
Πέραν του υπολογισμού του ποσοστού επιτυχημένης ανίχνευσης πραγματοποιήθηκε και μελέτη της ταχύτητας σύγκλισης της Q learning στη βέλτιστη πολιτική για διαφορετικές τιμές του συντελεστή προεξόφλησης. Η ταχύτητα σύγκλισης αποτελεί βασικό μέτρο επίδοσης των αλγορίθμων και, συνεπώς, κριτήριο για την επιλογή του κατάλληλου συντελεστή ελάττωσης. Ως μέτρο σύγκλισης χρησιμοποιήθηκε η μέση διαφορά (mean discrepancy) της συνάρτησης

χρησιμότητας που προέκυπτε κατά την εκάστοτε επανάληψη του αλγορίθμου, με τη συνάρτηση χρησιμότητας της πολιτικής που θεωρείτο βέλτιστη ως εκείνη τη στιγμή.

4.6 Αποτελέσματα προσομοίωσης

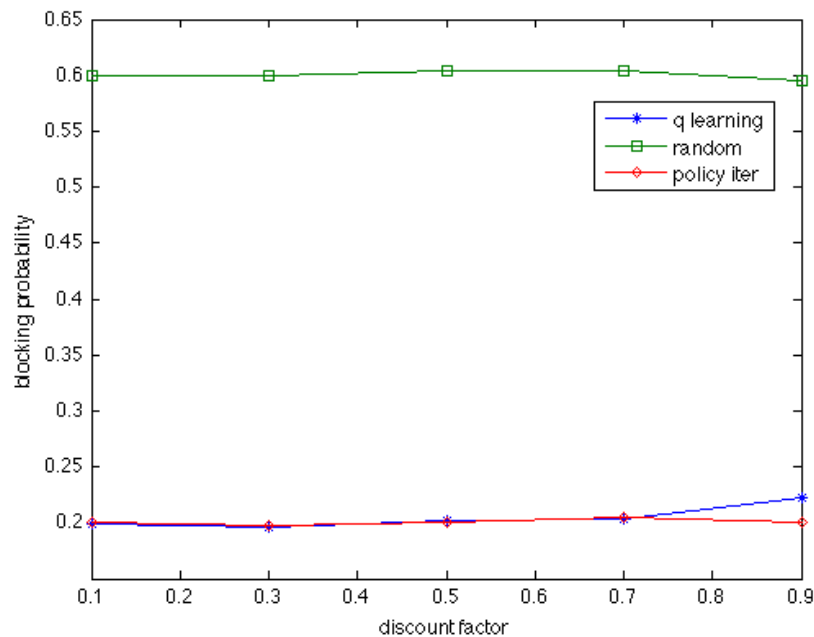
4.6.1 Πιθανότητα μη εκπομπής λόγω ανίχνευσης κατειλημμένου καναλιού

Οι γραφικές παραστάσεις που ακολουθούν παρουσιάζουν την πιθανότητα μη εκπομπής λόγω αποτυχημένης ανίχνευσης (blocking probability) που προκύπτει ύστερα από την εφαρμογή των τεχνικών ενισχυτικής μάθησης, Q learning και policy iteration, κατά τη διαδικασία ανίχνευσης φάσματος. Τα αποτελέσματα συγκρίνονται με αυτά της τυχαίας ανίχνευσης καναλιού προκειμένου να διαπιστωθεί η θετική επίδραση της ενισχυτικής μάθησης. Επίσης, προκειμένου να διερευνηθεί και η επίδραση του παράγοντα προεξόφλησης (discount factor) στις τεχνικές ανίχνευσης φάσματος, η πιθανότητα αποτυχημένης ανίχνευσης υπολογίστηκε για τις τιμές του γ 0.1, 0.3, 0.5, 0.7 και 0.9. Ο παράγοντας προεξόφλησης δεν έχει κάποια επίδραση στην τυχαία ανίχνευση. Οι πιθανότητες blocking που αντιστοιχίζονται σε κάθε γ είναι απλά οι τιμές που προέκυψαν από τις διαφορετικές προσομοιώσεις που πραγματοποιήθηκαν. Ακολουθούν οι γραφικές παραστάσεις για φασματική χρήση 80%, 60% και 20% αντίστοιχα.



Σχήμα 4.4: Πιθανότητα μη εκπομπής για φασματική χρήση 80%

Όπως φαίνεται από τη γραφική παράσταση αυτή, όταν η φασματική χρήση είναι 80%, με την εφαρμογή κάποιας από τις δύο υπό μελέτη τεχνικές ενισχυτικής μάθησης, η πιθανότητα μη εκπομπής είναι περίπου 60%. Αν πραγματοποιηθεί ανίχνευση με τυχαίο τρόπο, η πιθανότητα μη εκπομπής είναι περίπου 80% όσο δηλαδή και το ποσοστό κατάληψης του φάσματος. Φαίνεται δηλαδή, ότι με την εφαρμογή ενισχυτικής μάθησης, η πιθανότητα επιτυχημένης ανίχνευσης γίνεται περίπου 40% έναντι 20% που επιτυγχάνεται με τυχαία ανίχνευση. Δηλαδή υπάρχει 100% βελτίωση της απόδοσης. Επίσης, δεν παρατηρείται ουσιαστική διαφοροποίηση της επίδοσης των δύο τεχνικών καθώς μεταβάλλεται το γ .

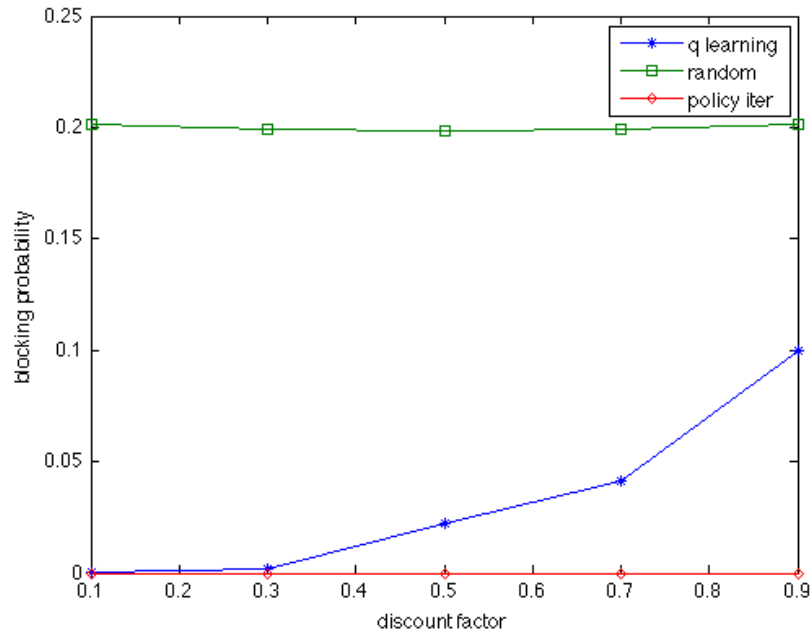


Σχήμα 4.5: Πιθανότητα μη εκπομπής για φασματική χρήση 60%

Όπως φαίνεται από τη γραφική παράσταση του σχήματος 4.5, όταν η χρησιμοποίηση του φάσματος από τους πρωτογενείς χρήστες είναι της τάξης του 60% η πιθανότητα μη εκπομπής με την εφαρμογή του αλγορίθμου MDP με επανάληψη πολιτικών κυμαίνεται γύρω στο 20% ανεξάρτητα από το γ . Ανάλογα ποσοστά μη εκπομπής προκύπτουν με την εφαρμογή του αλγορίθμου Q learning, ο

οποίος όμως, για συντελεστή προεξόφλησης 0.9 παρουσιάζει ελαφρώς αυξημένη πιθανότητα μη εκπομπής (~ 23%). Όπως είναι αναμενόμενο, η πιθανότητα αποτυχημένης ανίχνευσης όταν ακολουθείται η στρατηγική τυχαίας επιλογής καναλιού είναι περίπου ίση με το ποσοστό χρήσης του καναλιού. Παρατηρείται ότι, λόγω της εφαρμογής ενισχυτικής μάθησης, η πιθανότητα επιτυχημένης ανίχνευσης είναι περίπου 80% έναντι 40%, δηλαδή υπάρχει 100% βελτίωση της απόδοσης.

Το όφελος από την εφαρμογή ενισχυτικής μάθησης είναι περισσότερο εμφανές όταν η φασματική χρήση είναι χαμηλή. Στο σχήμα που ακολουθεί φαίνεται ότι, για ποσοστό χρήσης 20%, η πιθανότητα μη εκπομπής είναι μηδενική, ανεξαρτήτως συντελεστή προεξόφλησης, όταν εφαρμόζεται MDP επανάληψης πολιτικών. Πρακτικά αυτό σημαίνει ότι ο δευτερογενής χρήσης συμπεριφέρεται ως πρωτογενής καθώς αποκτά πρόσβαση στο φάσμα όποτε επιθυμεί. Όσον αφορά την εκμάθηση Q, η πιθανότητα μη εκπομπής είναι μηδενική για $\gamma=0.1$ και σχεδόν μηδενική για $\gamma=0.3$ (~0.5%). Ωστόσο, για μεγαλύτερες τιμές του συντελεστή προεξόφλησης παρουσιάζονται σημαντικές συγκριτικά μεταβολές στις τιμές πιθανότητας μη εκπομπής. Συγκεκριμένα, η πιθανότητα αποτυχημένης ανίχνευσης είναι $P_b=2.5\%$ για $\gamma=0.5$, $P_b=5\%$ για $\gamma=0.7$ και $P_b=10\%$ για $\gamma=0.9$. Όταν χρησιμοποιείται τυχαία ανίχνευση, η πιθανότητα αποτυχημένης ανίχνευσης κυμαίνεται στα επίπεδα χρήσης του φάσματος, δηλαδή 20%. Από τα προηγούμενα, προκύπτει ότι για χρήση φάσματος 20%, η πιθανότητα επιτυχημένης ανίχνευσης αυξάνεται από 80% σε 100% όταν εφαρμόζεται ενισχυτική μάθηση επανάληψης πολιτικών, δηλαδή υπάρχει βελτίωση της πιθανότητας επιτυχημένης ανίχνευσης κατά 25% σε σχέση με την τυχαία ανίχνευση. Αντίστοιχα, όταν εφαρμόζεται Q learning, η βελτίωση στην πιθανότητα επιτυχημένης ανίχνευσης κυμαίνεται από 12.5% μέχρι 25% ανάλογα με την τιμή του συντελεστή ελάττωσης γ . Σημαντικότερο συμπέρασμα, όμως, αποτελεί το ότι καθίσταται δυνατή η επίτευξη ανίχνευσης διαθέσιμου καναλιού με πιθανότητα 1.



Σχήμα 4.6: Πιθανότητα μη εκπομπής για φασματική χρήση 20%

Από τα προηγούμενα προκύπτει ότι ο αλγόριθμος επανάληψης πολιτικών δεν φαίνεται να επηρεάζεται από το παράγοντα προεξόφλησης. Αυτό οφείλεται στη χρήση μοντέλου όπου στηρίζεται η διαδικασία μάθησης. Αντίθετα, η εκμάθηση Q που δεν βασίζεται σε χρήση μοντέλου, οπότε ο αλγόριθμος πρέπει να εξερευνήσει το περιβάλλον, επηρεάζεται από την τιμή του γ και φαίνεται να αποδίδει καλύτερα για μικρές τιμές του συντελεστή, ($\gamma < 0.3$).

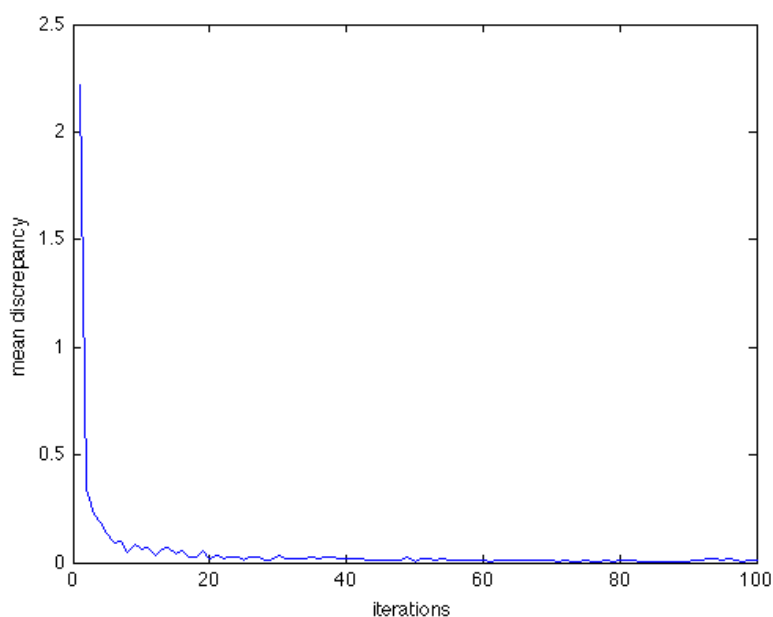
Είναι, επομένως, φανερό ότι η τεχνική επανάληψης πολιτικών αποδίδει εν γένει καλύτερα ή τουλάχιστον το ίδιο καλά με την εκμάθηση Q. Συγκεκριμένα, μόνο για $\gamma < 0.3$, οι δύο τεχνικές έχουν στις τρεις περιπτώσεις παρόμοια απόδοση.

4.6.2 Σύγκλιση της εκμάθησης Q προς τη βέλτιστη πολιτική

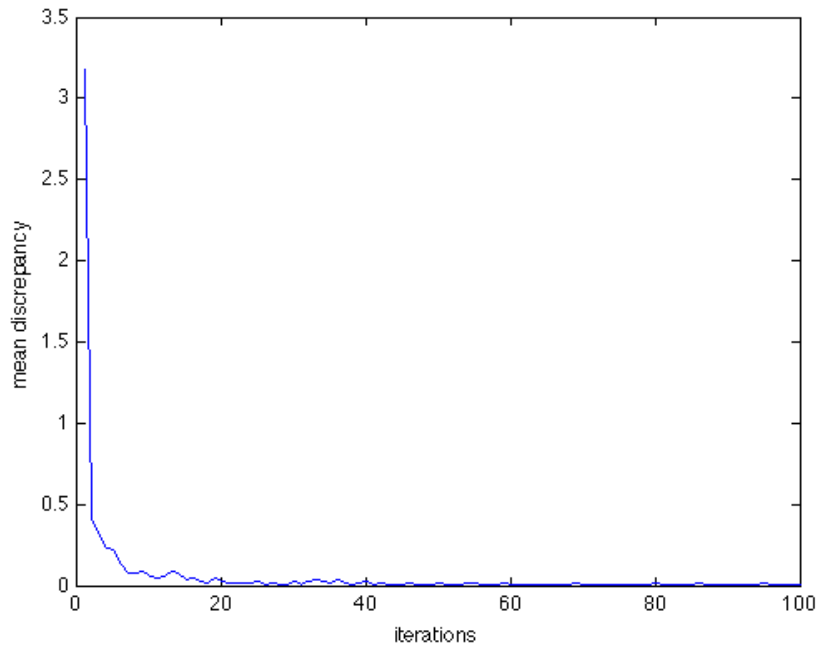
Όπως φάνηκε από τις προηγούμενες γραφικές παραστάσεις, ο συντελεστής προεξόφλησης δεν επηρεάζει σημαντικά την επίδοση του αλγορίθμου επανάληψης πολιτικών. Αντίθετα η επίδοση της εκμάθησης Q μειώνεται όταν ο συντελεστής υπερβαίνει κάποια τιμή, με τα αποτελέσματα να είναι περισσότερο αισθητά όταν η φασματική χρήση από τους πρωτογενείς χρήστες είναι μικρότερη.

Αυτές οι διαφορές στην επίδοση οφείλονται στο ότι ο συντελεστής ελάττωσης επηρεάζει τον τρόπο μάθησης. Συγκεκριμένα και όπως φαίνεται από την (3.15), για μεγαλύτερες τιμές του γ η συνάρτηση Q λαμβάνει περισσότερο υπόψη τις μελλοντικές τιμές με αποτέλεσμα να επηρεάζεται ο τρόπος σύγκλισης προς τη βέλτιστη πολιτική. Συγκεκριμένα, επειδή κατά την εκμάθηση Q δεν γίνεται χρήση μοντέλου, ο πράκτορας πρέπει να “εξερευνηί” το περιβάλλον του. Στο πλαίσιο αυτό επιλέγει και ενέργειες που δεν είναι βέλτιστες. Όταν λοιπόν το γ είναι μεγαλύτερο, λαμβάνεται περισσότερο υπόψη η χαμηλή ανταμοιβή των ενεργειών αυτών, με αποτέλεσμα να καθυστερεί η σύγκλιση προς τη βέλτιστη πολιτική. Σημειώνεται ότι, για να είναι δυνατή η σύγκλιση, πρέπει να ισχύει $0 < \gamma < 1$.

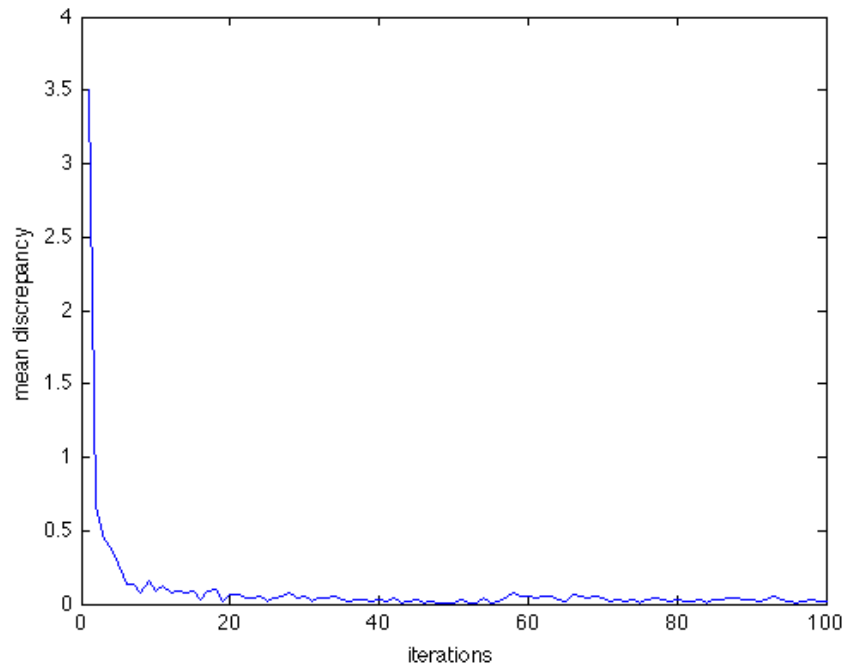
Στη συνέχεια εξετάζεται η επίδραση του συντελεστή ελάττωσης στη ταχύτητα σύγκλισης προς τη βέλτιστη πολιτική. Οι γραφικές παραστάσεις που ακολουθούν δείχνουν τη σύγκλιση προς τη βέλτιστη πολιτική για τις τιμές συντελεστή ελάττωσης $\gamma=0.1, 0.3, 0.5, 0.7, 0.9, 0.999$. Ως μέτρο σύγκλισης χρησιμοποιήθηκε η μέση διαφορά (mean discrepancy) της συνάρτησης χρησιμότητας που προκύπτει κατά την εκάστοτε επανάληψη (iteration) του αλγορίθμου, με τη συνάρτηση χρησιμότητας της πολιτικής που θεωρείται βέλτιστη ως εκείνη τη στιγμή.



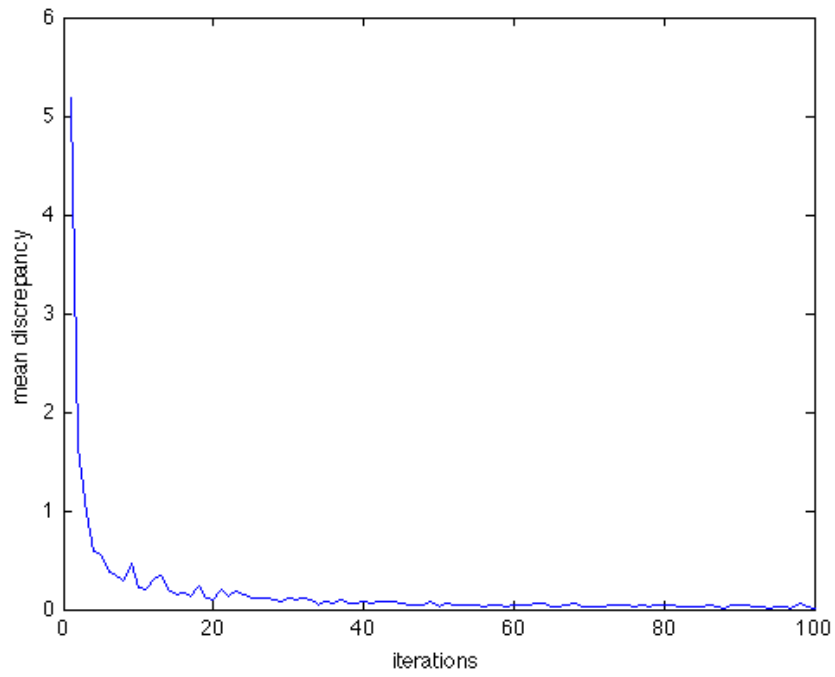
Σχήμα 4.7: Σύγκλιση προς τη βέλτιστη πολιτική για $\gamma=0.1$



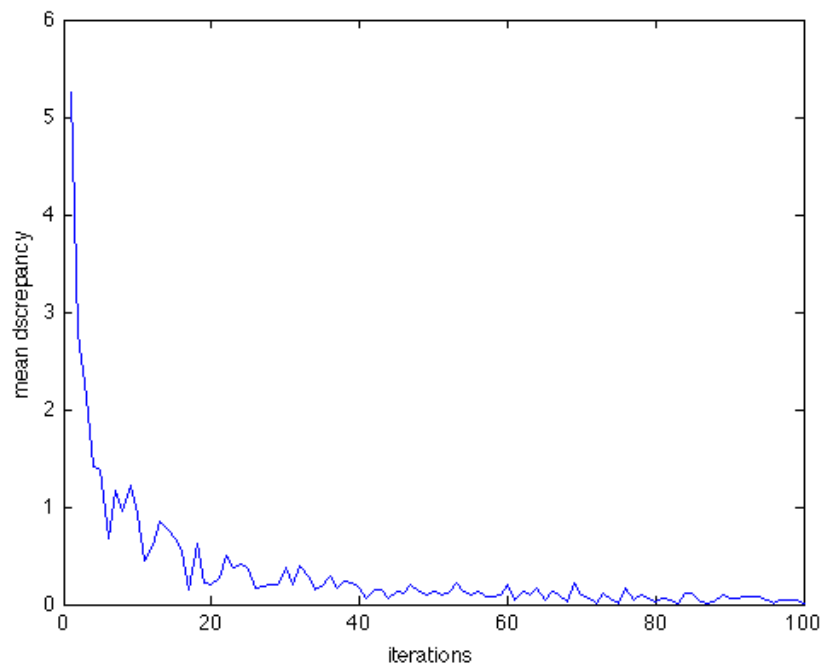
Σχήμα 4.8: Σύγκλιση προς τη βέλτιστη πολιτική για $\gamma=0.3$



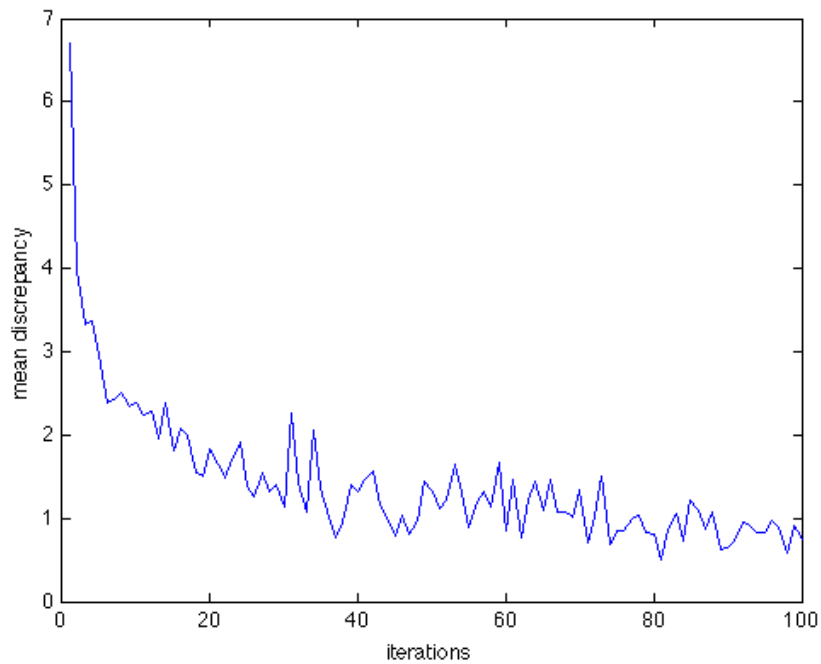
Σχήμα 4.9: Σύγκλιση προς τη βέλτιστη πολιτική για $\gamma=0.5$



Σχήμα 4.10: Σύγκλιση προς τη βέλτιστη πολιτική για $\gamma=0.7$



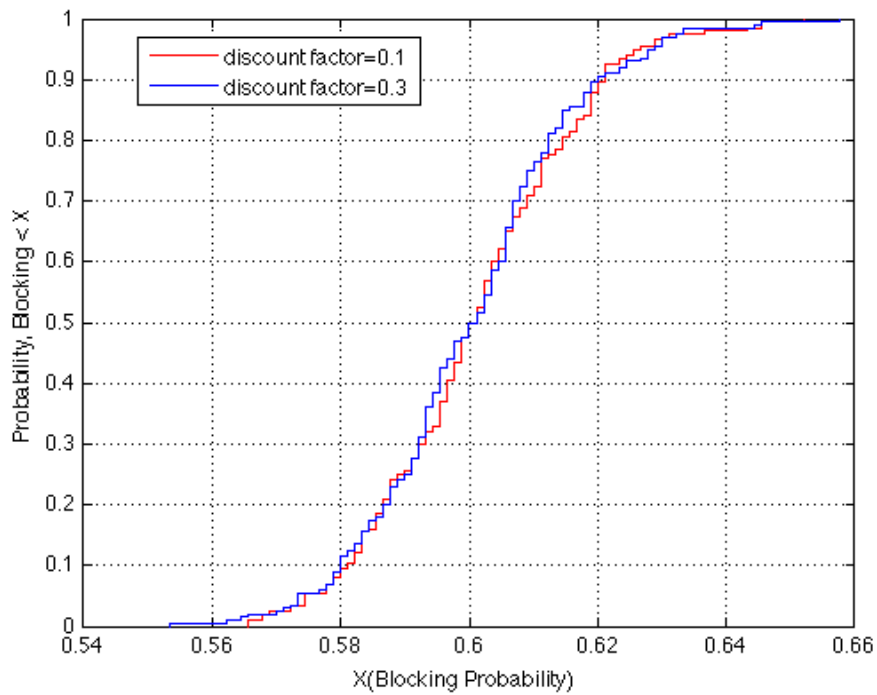
Σχήμα 4.11: Σύγκλιση προς τη βέλτιστη πολιτική για $\gamma=0.9$



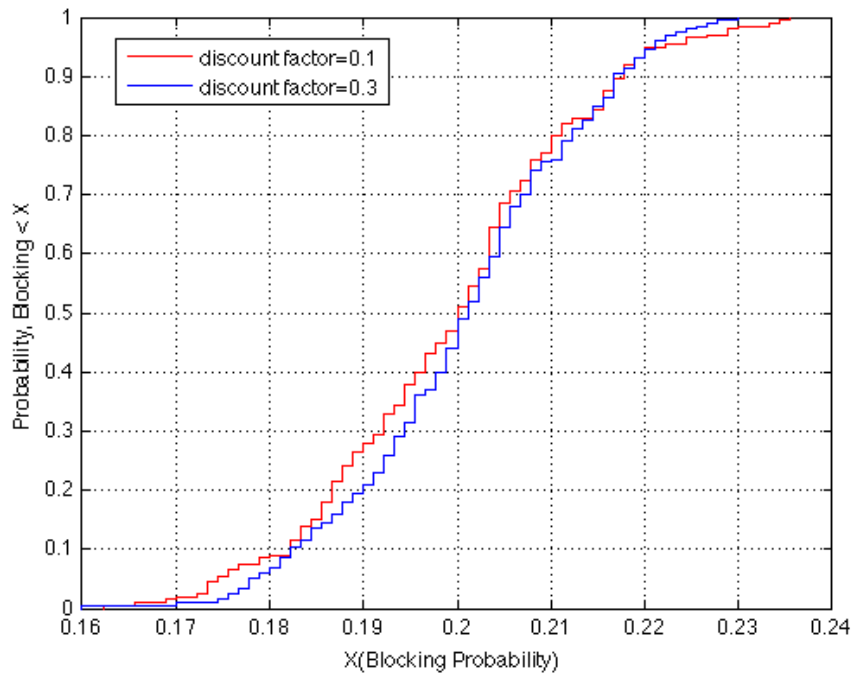
Σχήμα 4.12: Σύγκλιση προς τη βέλτιστη πολιτική για $\gamma=0.999$

Όπως προκύπτει από τα ανωτέρω σχήματα, ο αλγόριθμος συγκλίνει προς τη βέλτιστη πολιτική καθυστερημένα για συντελεστές προεξόφλησης μεγαλύτερους από 0.3. Για τιμές συντελεστή προεξόφλησης $\gamma=0.1$ και $\gamma=0.3$ η ασυμπτωτική σύγκλιση προς τη βέλτιστη πολιτική επιτυγχάνεται ύστερα από περίπου 20 επαναλήψεις του αλγορίθμου. Για $\gamma=0.5$ απαιτούνται περίπου 40 επαναλήψεις, ενώ για $\gamma=0.7$ περίπου 60. Για $\gamma=0.9$ η σύγκλιση εμφανίζεται μετά τις 90 επαναλήψεις παρουσιάζοντας αρκετές ανωμαλίες. Τέλος για $\gamma=0.999$ δεν υπάρχει σύγκλιση καθώς η τιμή του πλησιάζει πολύ την μονάδα, επιβεβαιώνοντας ότι ο αλγόριθμος εκμάθησης Q συγκλίνει για $0 < \gamma < 1$.

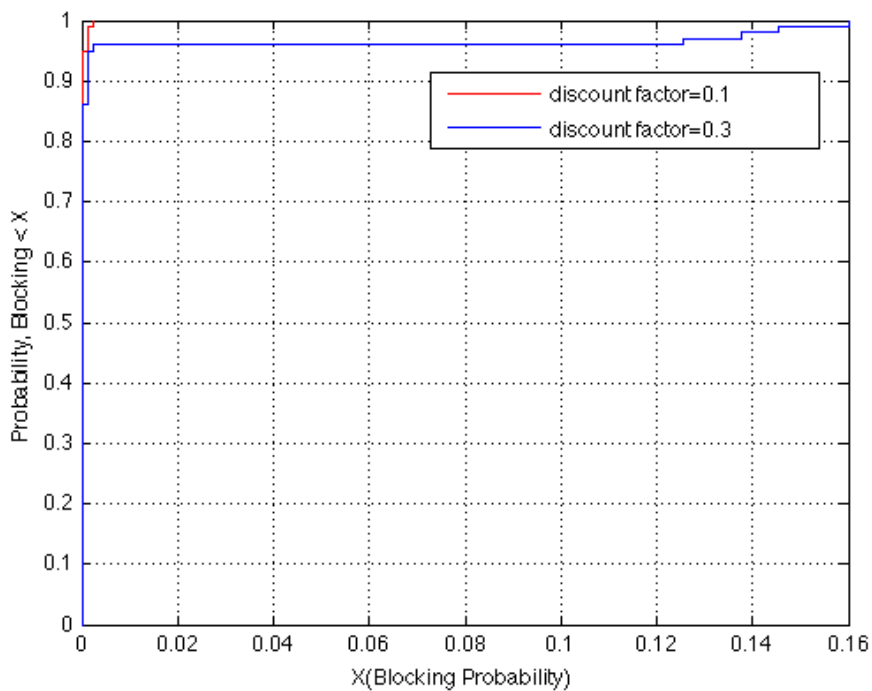
Οι καμπύλες σύγκλισης επιβεβαιώνουν τα αποτελέσματα των καμπυλών πιθανότητας αποτυχημένης ανίχνευσης, όπου η ενισχυτική μάθηση Q αποδίδει καλύτερα για συντελεστές προεξόφλησης μέχρι 0.3. Για περαιτέρω ανάλυση ακολουθούν οι CDF καμπύλες πιθανότητας μη εκπομπής λόγω αποτυχημένης ανίχνευσης (blocking probability) όταν εφαρμόζεται Q learning με discount factor 0.1 και 0.3. Οι καμπύλες αυτές, απεικονίζουν την πιθανότητα το ποσοστό μη εκπομπής να είναι κάτω από μια τιμή X.



Σχήμα 4.13: CDF καμπύλη αποτυχημένης ανίχνευσης για $\gamma=0.1, 0.3$ όταν εφαρμόζεται εκμάθηση Q και η φασματική χρήση είναι 80%



Σχήμα 4.14: CDF καμπύλη αποτυχημένης ανίχνευσης για $\gamma=0.1, 0.3$ όταν εφαρμόζεται εκμάθηση Q και η φασματική χρήση είναι 60%

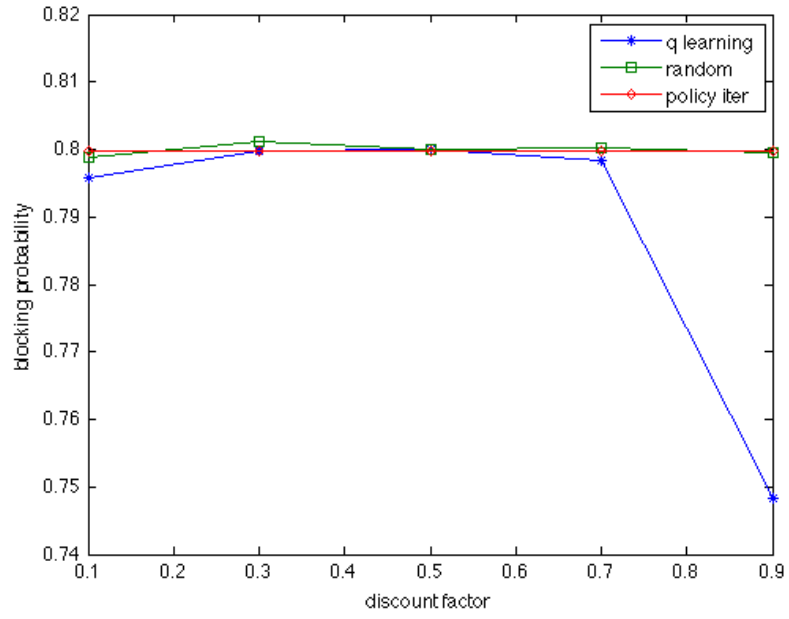


Σχήμα 4.15: CDF καμπύλη αποτυχημένης ανίχνευσης για $\gamma=0.1, 0.3$ όταν εφαρμόζεται εκμάθηση Q και η φασματική χρήση είναι 20%

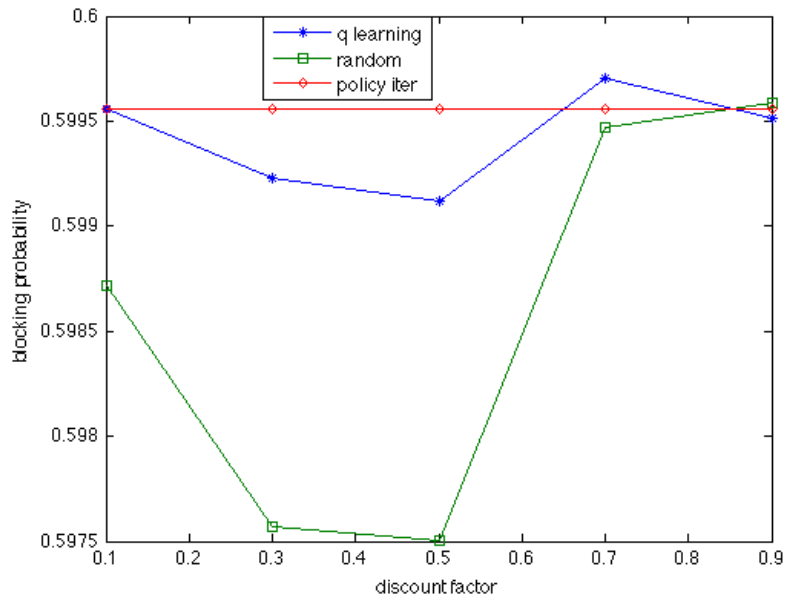
Όπως φαίνεται από τις ανωτέρω καμπύλες η εκμάθηση Q αποδίδει εξίσου καλά για $\gamma=0.1$ και $\gamma=0.3$ σε υψηλά ποσοστά φασματικής χρήσης ενώ για χαμηλότερα ποσοστά αποδίδει ελαφρώς καλύτερα για $\gamma=0.1$.

4.6.3 Εφαρμογή ενισχυτικής μάθησης όταν η φασματική χρήση δεν παρουσιάζει κανονικότητα

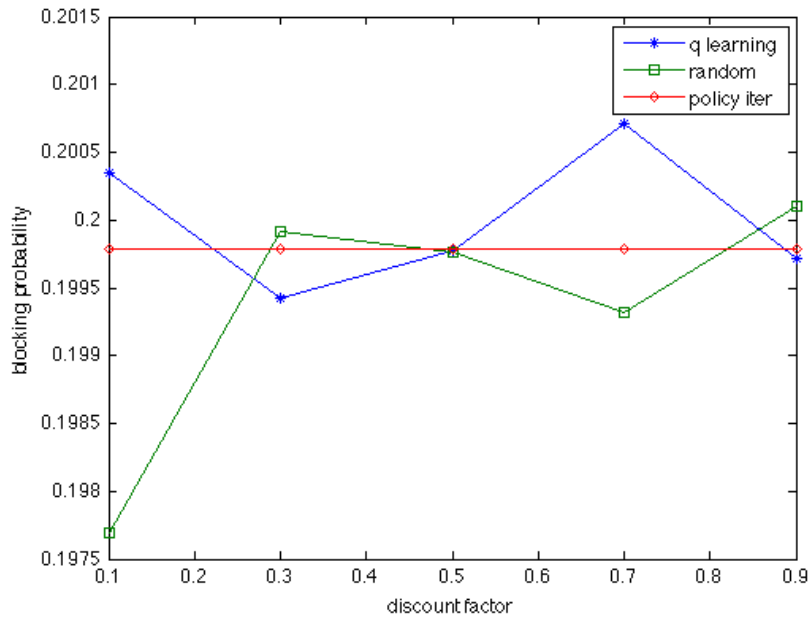
Στα προηγούμενα μελετήθηκε η εφαρμογή της ενισχυτικής μάθησης για την ανίχνευση φάσματος όταν αυτό παρουσιάζει κάποια στοιχεία κανονικότητας. Συγκεκριμένα, σύμφωνα με το μοντέλο φασματικής χρήσης που εφαρμόστηκε, η πρόσβαση στο φάσμα από τους πρωτογενείς χρήστες γίνεται κατά τυχαίο τρόπο, μολονότι κάποια κανάλια παρουσιάζουν υψηλότερα ποσοστά χρήσης σε σχέση με άλλα. Στην παράγραφο αυτή μελετάται η εφαρμογή της ενισχυτικής μάθησης όταν η φασματική χρήση είναι εντελώς τυχαία. Ακολουθούν οι καμπύλες πιθανοτήτων μη εκπομπής λόγω αποτυχημένης ανίχνευσης.



Σχήμα 4.16: Πιθανότητα μη εκπομπής για φασματική χρήση 80%



Σχήμα 4.17: Πιθανότητα μη εκπομπής για φασματική χρήση 60%



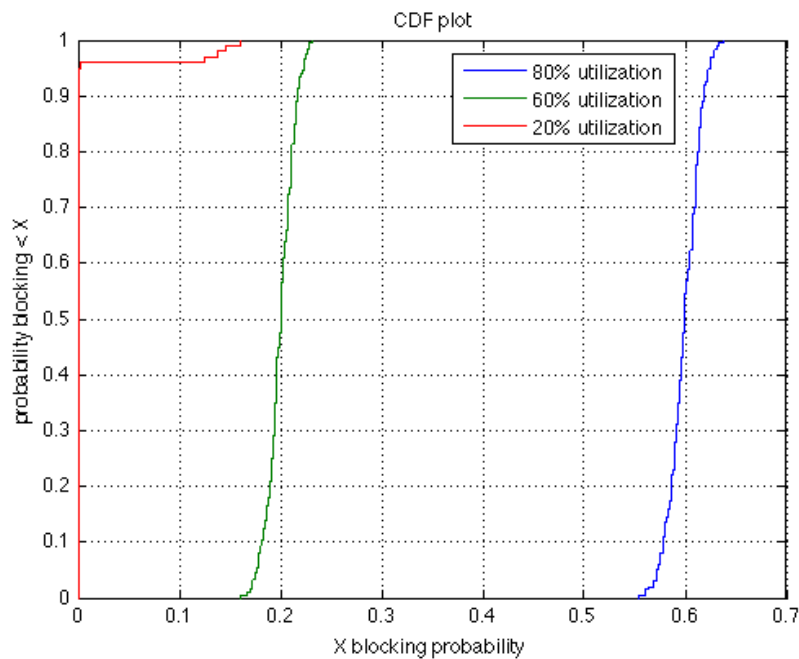
Σχήμα 4.18: Πιθανότητα μη εκπομπής για φασματική χρήση 20%

Από τα προηγούμενα σχήματα φαίνεται ότι η εφαρμογή ενισχυτικής μάθησης κατά τη διαδικασία ανίχνευσης όταν η φασματική χρήση είναι τελείως τυχαία, δεν βελτιώνει την επίδοση του συστήματος παρουσιάζοντας αποτελέσματα αντίστοιχα με αυτά της τυχαίας ανίχνευσης. Το αποτέλεσμα αυτό είναι απόλυτα κατανοητό καθώς όταν το περιβάλλον του πράκτορα(φάσμα συχνοτήτων) είναι τυχαίο χωρίς να παρουσιάζει συγκεκριμένα χαρακτηριστικά χρήσης, ο πράκτορας δεν μπορεί να εκπαιδευτεί σε κάποια βέλτιστη πολιτική αφού δεν υπάρχει τέτοια.

4.7 Συμπεράσματα

Από τα προηγούμενα έχει καταστεί εμφανής η συμβολή της ενισχυτικής μάθησης κατά τη διαδικασία ανίχνευσης φάσματος, υπό την προϋπόθεση, βέβαια, ότι η φασματική χρήση παρουσιάζει ορισμένα χαρακτηριστικά χρήσης. Αυτός, άλλωστε, είναι και ο στόχος της ενισχυτικής μάθησης, να εντοπίζει κάθε φορά ποια κανάλια παρουσιάζουν χαμηλότερη χρησιμοποίηση. Για υψηλά ποσοστά χρήσης, η πιθανότητα αποτυχημένης ανίχνευσης ακόμα και μετά την εφαρμογή ενισχυτικής μάθησης παραμένει υψηλή για να καλύψει τις απαιτήσεις σε QoS ενός απαιτητικού

συστήματος καθώς οι καθυστερήσεις που προκύπτουν από την αδυναμία εκπομπής οδηγούν σε χαμηλούς ρυθμούς μετάδοσης δεδομένων. Ωστόσο, αν ληφθεί υπόψη ότι τα δίκτυα γνωστικών ραδιοεπικοινωνιών προορίζονται για εφαρμογή σε περιοχές με χαμηλή φασματική χρήση (κάτω του 20%) η χρήση ενισχυτικής μάθησης είναι καθοριστική, καθώς μπορεί να οδηγήσει σε ανίχνευση φασματικών κενών με ποσοστό επιτυχίας που πλησιάζει το 100%. Αυτό σημαίνει ότι μετά από κάποιο χρονικό διάστημα δεν είναι απαραίτητη η ανίχνευση φάσματος αλλά ο δευτερογενής χρήστης θα μπορεί να αποκτά άμεση πρόσβαση στο φάσμα, να δρα δηλαδή ως πρωτογενής χρήστης. Η επίδοση της εφαρμογής ενισχυτικής μάθησης κατά τη φασματική ανίχνευση για διαφορετικά ποσοστά χρήσης φάσματος απεικονίζεται στο σχήμα 4.19.



Σχήμα 4.19: CDF καμπύλες μη εκπομπής για ποσοστά χρήσης 20%, 60% και 80%

Όσον αφορά τη σύγκριση μεταξύ των δύο μεθόδων που μελετήθηκαν φαίνεται ότι η επίδοση της τεχνικής επανάληψης πολιτικών είναι γενικά καλύτερη από αυτή της εκμάθησης Q. Αυτό οφείλεται στη χρήση μοντέλου καθώς μειώνει την εξερεύνηση του περιβάλλοντος που οδηγεί σε καθυστερημένη σύγκλιση. Στη χρήση μοντέλου οφείλεται και η χαμηλή εξάρτηση της απόδοσης της policy iteration από

το συντελεστή προεξόφλησης. Ωστόσο, η εξάρτηση της επανάληψης πολιτικών από το μοντέλο του περιβάλλοντος αποτελεί και το αδύναμο σημείο της καθώς τα σφάλματα που μπορεί να προκύψουν κατά τη δημιουργία του μοντέλου μπορούν να οδηγήσουν σε χαμηλή απόδοση. Αντίθετα, η εκμάθηση Q επηρεάζεται από την τιμή του γ . Η επιρροή αυτή είναι περισσότερο εμφανής σε χαμηλά ποσοστά χρήσης φάσματος, όπου για $\gamma \leq 0.3$ παρουσιάζει αισθητά καλύτερα αποτελέσματα, σε σχέση με μεγαλύτερες τιμές του γ .

Κλείνοντας, η ανωτέρω μελέτη θα μπορούσε να αποτελέσει τη βάση για την ανάπτυξη άλλων αλγορίθμων φασματικής ανίχνευσης και πρόσβασης όπως για παράδειγμα αλγορίθμων που σταματούν να ανιχνεύουν το φάσμα όταν επιτευχθεί σύγκλιση προς τη βέλτιστη πολιτική. Επίσης, καθώς αποτελεί υλοποίηση φυσικού επιπέδου, με την εφαρμογή του κατάλληλου στρώματος πρόσβασης μέσου (MAC) μπορεί να εφαρμοστεί σε διάφορα ασύρματα συστήματα που πραγματοποιούν είτε συνεργατική (συγκεντρωτική ή κατανεμημένη) είτε μη συνεργατική ανίχνευση φάσματος. Για παράδειγμα σε ένα σενάριο συγκεντρωτικής, συνεργατικής ανίχνευσης ο σταθμός βάσης μπορεί να χρησιμοποιεί ενισχυτική μάθηση για την εύρεση της βέλτιστης πολιτικής ανάθεσης πόρων στους τερματικούς σταθμούς.

Βιβλιογραφία

- [1] FCC Spectrum Policy Task Force, “FCC Report of the Spectrum Efficiency Working Group”, Nov. 2002.
- [2] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, “Chicago spectrum occupancy measurements & analysis and a long-term studies proposal,” in *Proceedings of First International Workshop on Technology and Policy for Accessing Spectrum (TAPAS)*. ACM, 2006, p. 1.
- [3] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, “Chicago spectrum occupancy measurements & analysis and a long-term studies proposal”.
- [4] Frequency allocation, http://en.wikipedia.org/wiki/Frequency_allocation.
- [5] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, “Chicago spectrum occupancy measurements & analysis and a long-term studies proposal”.
- [6] Qing Zhao and Brian M. Sadler, A Survey of Dynamic Spectrum Access, IEEE signal processing magazine May 2007.
- [7] S. Haykin, “Cognitive Radio: Brain-Empowered Wireless Communications,” IEEE JSAC, vol. 23, no. 2, Feb. 2005, pp. 201–20.
- [8] J. Mitola, “Cognitive radio for flexible mobile multimedia communications,” in Proc. IEEE Int. Workshop Mobile Multimedia Communications, 1999, pp. 3-10.
- [9] 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.
- [10] Γ. Καραγιάννης, Κ. Τζιτζιράχου, “Εισαγωγή στα Σήματα και Συστήματα”, Εκδόσεις Παπασωτηρίου, Αθήνα, 2003.
- [11] J. Mitola III, “Cognitive radio: An integrated agent architecture for software defined radio,” PhD Dissertation, Royal Institute of Technology, Stockholm, Sweden, May 2000.
- [12] FCC ET Docket No. 03-108, March 11, 2005.
- [13] FCC, ET Docket No 03-237, November 2003.

- [14] S. Haykin, *Communication Systems*, 4th Ed. New York 2001
- [15] T. C. Clancy, "Achievable capacity under the interference temperature model," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, May 2007, pp. 794–802.
- [16] B. Wild and K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in *Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, November 2005, pp. 124–130.
- [17] D. Cabric, S. M. Mishra, D. Willkomm, R. Brodersen, and A. Wolisz, "A cognitive radio approach for usage of virtual unlicensed spectrum," in *Proceedings of 14th IST Mobile and Wireless Communications Summit*, June 2005.
- [18] M. M. Buddhikot, P. Kolody, S. Miller, K. Ryan, and J. Evans, "DIMSUNnet: new directions in wireless networking using coordinated dynamic spectrum," in *Proceedings of IEEE International Symposium on a World of Wireless, Mobile, and Multimedia Networks (WoWMoM)*, June 2005, pp. 78–85.
- [19] L. Xu, R. Tonjes, T. Paila, W. Hansmann, M. Frank, and M. Albrecht, "DRiVE-ing to the Internet: dynamic radio for IP services in vehicular environments," in *Proceedings of 25th Annual IEEE Conference on Local Computer Networks (LCN)*, November 2000, pp. 281–289.
- [20] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.
- [21] "IEEE 1900 standards committee, IEEE SCC 41." <http://www.scc41.org>.
- [22] C. Cordeiro, K. Challapali, K. Birru, S. Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios", In proc. of DySPAN'05, November 2005.
- [23] C. Cordeiro, K. Challapali, and D. Birru, "IEEE 802.22: An introduction to the first wireless standard based on cognitive radios," *Journal of communications*, vol. 1, no. 1, Apr. 2006.
- [24] Ekram Hossain, Dusit Niyato, Zhu Han, *Dynamic Spectrum Access and Management in Cognitive Radio Networks*, Cambridge University Press 2009.
- [25] K. Challapali, C. Cordeiro, and D. Birru, "Evolution of spectrum-agile cognitive radios: first wireless Internet standard and beyond," in *Proceedings of 2nd Annual International Workshop on Wireless Internet (WICON)*. ACM, 2006, p. 27.

- [26] V. R. Petty, R. Rajbanshi, D. Datla *et al.*, "Feasibility of dynamic spectrum access in underutilized television bands," in *Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, April 2007, pp. 331–339.
- [27] R.D. Hinman, "Application of cognitive radio technology to legacy military waveforms in a JTRS (joint tactical radio system) radio," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, October 2006.
- [28] Huseyin Arslan, *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*, Springer Netherlands 2007.
- [29] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environment," in *Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, November 2005, pp. 131–136.
- [30] A. Sahai, N. Hoven, R. Tandra, Some fundamental limits in cognitive radio, Allerton Conf. on Commun., Control and Computing 2004, October 2004.
- [31] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," in *Proceedings of International Conference on Communications (ICC)*, vol. 5, May 2003, pp. 3575–3579.
- [32] 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 2004, pp. 772–776.
- [33] G. Ganesan and Y. G. Li, "Cooperative spectrum sensing in cognitive radio networks," in *Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, November 2005, pp. 137–143.
- [34] Q. Zhao, L. Tong, A. Swami, Decentralized cognitive MAC for dynamic spectrum access, in: Proc. IEEE DySPAN 2005, November 2005, pp. 224–232.
- [35] 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.
- [36] T. C. Clancy, "Formalizing the interference temperature model", *Wireless Communications and Mobile Computing*, vol. 7, no. 9, pp. 1077–1086, 2007.
- [37] 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.

[38] E. G. Villegas, R.V. Ferre, and J. P. Aspas, "Load balancing in WLANs through IEEE802.11k mechanisms," in *Proceedings of IEEE Symposium on Computers and Communications (ISCC)*, June 2006, pp. 844–850.

[39] S. Hermann, M. Emmelmann, O. Belaifa, and A. Wolisz, "Investigation of IEEE 802.11k based access point coverage area and neighbor discovery," in *Proceedings of IEEE Conference on Local Computer Networks (LCN)*, October 2007, pp. 949–954.

[40] N. Golmie, O. Rebala, and N. Chevrollier, "Bluetooth adaptive frequency hopping and scheduling," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, vol. 2, October 2003, pp. 1138–1142.

[41] Kaelbling, Leslie P., Michael L. Littman, Andrew W. Moore, "Reinforcement Learning: A Survey" (1996), *Journal of Artificial Intelligence Research* 4: 237–285.

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

[43] K. Murphy, "A survey of POMDP solution techniques", tech. rep., 2000, Technical Report, UC Berkeley.

[44] Kok-Lim Alvin Yau, Peter Komisarczuk, 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.

[45] C. Watkins, *Learning From Delayed Rewards*, PhD Thesis, The University of Cambridge, UK, 1989.

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

[47] Singh, S., and Bertsekas, D. P. (1996), "Reinforcement Learning for Dynamic Channel Allocation in Cellular Telephone Systems", MIT Press, Cambridge, Massachusetts 1996.

[48] Hang Su and Xi Zhang, "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.