



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

Εφαρμογή μεθόδων ενισχυτικής μάθησης σε κατανεμημένα γνωστικά δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Α. Τραγανίτης

Επιβλέπων : Παναγιώτης Κωττής
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2013



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

Εφαρμογή μεθόδων ενισχυτικής μάθησης σε κατανεμημένα γνωστικά δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Α. Τραγανίτης

Επιβλέπων : Παναγιώτης Κωττής

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26^η Ιουνίου 2013.

.....

Π. Κωττής

Καθηγητής Ε.Μ.Π.

.....

Χ. Καψάλης

Καθηγητής Ε.Μ.Π.

.....

Γ. Φικιώρης

Επ. Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2013

.....

Παναγιώτης Α. Τραγανίτης

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

Copyright © Παναγιώτης Τραγανίτης, 2013.

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.

Περίληψη

Η ραγδαία αύξηση της ζήτησης για φάσμα ραδιοσυχνοτήτων και ο τρόπος διάθεσης του έχει οδηγήσει σε υψηλή συμφόρηση στο μη αδειοδοτημένο φάσμα, αλλά και σε υποχρησιμοποίηση του αδειοδοτημένου. Αυτή η κατάσταση, με τη σειρά της, οδηγεί στην ανάγκη ανάπτυξης τεχνολογιών γνωστικών ραδιοεπικοινωνιών και γνωστικών δικτύων, οι οποίες εκμεταλλεύονται αποτελεσματικότερα το διαθέσιμο ηλεκτρομαγνητικό φάσμα.

Σκοπός της εργασίας είναι η μελέτη της εφαρμογής ενισχυτικής μάθησης στην διαδικασία επιλογής καναλιού των γνωστικών ραδιοσυσκευών. Αρχικά παρουσιάζεται το αντικείμενο των γνωστικών ραδιοεπικοινωνιών και των γνωστικών δικτύων, καθώς και της ενισχυτικής μάθησης. Για τη μελέτη των εφαρμογών της ενισχυτικής μάθησης στα γνωστικά δίκτυα δημιουργήθηκε προσομοιωτής σε γλώσσα C#, με τη βοήθεια του οποίου εξάγονται αποτελέσματα και συμπεράσματα για ήδη υπάρχουσες εφαρμογές ενισχυτικής μάθησης σε γνωστικά δίκτυα. Επίσης προτείνονται βελτιώσεις σε αυτές τις εφαρμογές και αξιολογείται η αποτελεσματικότητά τους.

Λέξεις Κλειδιά

Γνωστικές ραδιοεπικοινωνίες, γνωστικά δίκτυα, γνωστικές ραδιοσυσκευές, ενισχυτική μάθηση, δυναμική επιλογή καναλιού.

Abstract

The high demand of RF spectrum and the way it is assigned to various users has led to high congestion in the unlicensed spectrum as well as underutilization of the licensed spectrum. In turn, this situation leads to the need of developing cognitive radio and cognitive network technologies, which are able to use the entire available spectrum, licensed and unlicensed, in a more efficient fashion.

The purpose of this thesis is the study of reinforcement learning application in the channel decision process of cognitive radios. Firstly, the principles of cognitive radios, cognitive networks and reinforcement learning are presented. For the study of the application of reinforcement learning on cognitive networks a simulating environment was developed, in C#, which is used to obtain results and reach conclusions for existing applications. Furthermore, improvements in these applications are proposed and evaluated.

Key words

Cognitive radio, cognitive networks, reinforcement learning, dynamic channel selection.

Θα ήθελα να εκφράσω τις θερμότερες ευχαριστίες μου στον καθηγητή και επιβλέποντα της συγκεκριμένης εργασίας κ. Παναγιώτη Κωττή για την άριστη συνεργασία μας, την πολύτιμη βοήθεια του και τις γνώσεις που μου πρόσφερε κατά τη διάρκεια των σπουδών μου στη σχολή. Επίσης, θα ήθελα να ευχαριστήσω την Υποψήφια Διδάκτορα Αγγελική Κορδαλή για τις πολύτιμες υποδείξεις και την ουσιαστική βοήθεια της στην εκπόνηση της παρούσας εργασίας. Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου για τη αμέριστη συμπαράσταση που μου προσφέρει.

Περιεχόμενα

1	Κεφάλαιο 1 – Γνωστικά Δίκτυα	13
1.1	Ορισμός	13
1.2	Αρχιτεκτονική ενός cognitive radio	18
1.3	Τεχνικές Ανίχνευσης Φάσματος (Spectrum Sensing).....	19
1.3.1	Ανιχνευτής Ενέργειας (Energy Detector)	20
1.3.2	Ανίχνευση κυματομορφής (Waveform-Based Sensing).....	22
1.3.3	Ανίχνευση Κυκλοστατικότητας (Cyclostationarity-Based Sensing).....	22
1.3.4	Ανίχνευση Ραδιοαναγνώρισης (Radio Identification Based Sensing).....	24
1.3.5	Ανίχνευση προσαρμοσμένου φίλτρου (Matched Filtering).....	24
1.3.6	Άλλες μέθοδοι ανίχνευσης.....	24
1.3.7	Σύγκριση μεθόδων ανίχνευσης.....	25
1.4	Συνεργατική Ανίχνευση (Cooperative Sensing)	26
1.5	Εξωτερική ανίχνευση (External Sensing)	26
1.6	Χρησιμοποίηση του ιστορικού για προβλέψεις	27
1.7	Διαχείριση Φάσματος (Spectrum Management).....	27
1.7.1	Ανάλυση φάσματος (Spectrum Analysis).....	27
1.7.2	Επιλογή φάσματος	29
1.8	Κινητικότητα στο φάσμα (Spectrum Mobility)	29
1.9	Διαμοιρασμός φάσματος (Spectrum sharing)	29
1.10	Σύγχρονες τεχνολογίες.....	31
1.10.1	IEEE 802.11k	31
1.10.2	Bluetooth.....	32
1.10.3	IEEE 802.22	32
2	Κεφάλαιο 2 – Ενισχυτική Μάθηση.....	35
2.1	Το μοντέλο της ενισχυτικής μάθησης	35
2.2	Μοντέλα Βέλτιστης Συμπεριφοράς	37
2.2.1	Το μοντέλο πεπερασμένου ορίζοντα (finite-horizon model).	37
2.2.2	Το απομειούμενο μοντέλο απείρου ορίζοντα (infinite-horizon discounted model).	38
2.2.3	Το μοντέλο μέσης ανταμοιβής (average-reward model)	38
2.3	Μέτρηση της ποιότητας της μάθησης	39
2.3.1	Τελική σύγκλιση στο βέλτιστο.(Eventual Convergence to optimality)	39
2.3.2	Ταχύτητα σύγκλισης στο βέλτιστο. (Speed of convergence to optimality)	39

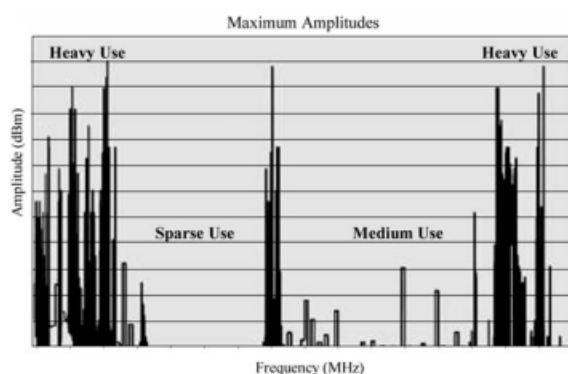
2.3.3	Μεταμέλεια - Regret	39
2.3.4	Ο συμβιβασμός μεταξύ εξερεύνησης και εκμετάλλευσης	40
2.3.5	Καθυστερημένη Ανταμοιβή	41
2.4	Διαδικασίες αποφάσεων Markov (Markov Decision Processes, MDPs).....	41
2.5	Εύρεση πολιτικής για δεδομένο Μοντέλο μάθησης.....	41
2.5.1	Επανάληψη τιμών(Value iteration).....	42
2.5.2	Επανάληψη πολιτικών (Policy iteration).....	43
2.6	Μαθαίνοντας μια βέλτιστη πολιτική: Μέθοδοι χωρίς μοντέλα	44
2.6.1	Q-μάθηση (Q-learning).....	44
2.7	Εφαρμογές της ενισχυτικής μάθησης	46
3	Κεφάλαιο 3 – Περιγραφή Μοντέλου	49
3.1	Το μοντέλο προσομοίωσης	49
3.1.1	Το κανάλι.....	49
3.1.2	Το κανάλι ελέγχου.....	50
3.1.3	Πρωτεύοντες Χρήστες (Primary Users)	50
3.1.4	Δευτερεύοντες Χρήστες (Secondary Users)	50
3.1.5	Μηχανισμοί εξερεύνησης	53
3.1.6	Μετρήσιμα μεγέθη.....	54
3.2	Πρωτόκολλα Πρόσβασης στο Μέσο (MAC)	54
3.2.1	SMAC (Single agent Medium Access Control)	56
3.2.2	MMAC (Multi agent Medium Access Control)	57
3.2.3	SMAC με έλεγχο υπαναχωρήσεων (Backoff)	57
3.3	Ο προσομοιωτής	58
3.4	Συνεισφορές της παρούσας εργασίας	61
4	Κεφάλαιο 4 – Αποτελέσματα	63
4.1	Παράμετροι προσομοίωσης.....	63
4.2	Πειράματα και αποτελέσματα	64
4.2.1	Επίπεδα χρήσης καναλιού	64
4.2.2	Σύγκριση SARL και MARL.....	66
4.2.3	Σύγκριση διαφορετικών μεθόδων εξερεύνησης(exploration)	71
4.2.4	Σύγκριση SMAC και SMAC με έλεγχο υπαναχωρήσεων	75
4.3	Συμπεράσματα – Παρατηρήσεις.....	82
Παράρτημα Α – Μέση τιμή του ελαχίστου n ανεξάρτητων ομοιόμορφα κατανεμημένων μεταβλητών.....		85
		10

Βιβλιογραφία 87

1 Κεφάλαιο 1 – Γνωστικά Δίκτυα

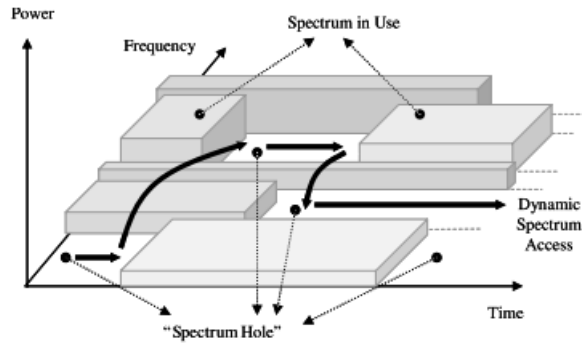
1.1 Ορισμός

Το ηλεκτρομαγνητικό φάσμα είναι χωρισμένο σε αδειοδοτημένες (licensed) και μη-αδειοδοτημένες (unlicensed) συχνότητες. Το αδειοδοτούμενο φάσμα χρησιμοποιείται αποκλειστικά από τους αδειοδοτημένους χρήστες (πρωτεύοντες χρήστες) , ενώ το μη-αδειοδοτούμενο φάσμα μπορεί να χρησιμοποιηθεί από οποιονδήποτε (δευτερεύοντα χρήστη), αρκεί να ακολουθούνται συγκεκριμένοι κανόνες (π.χ. η ισχύς εκπομπής να μην υπερβαίνει κάποιο όριο ή να μην γίνεται ταυτόχρονη εκπομπή). Στο μη-αδειοδοτημένο φάσμα λειτουργούν συσκευές ISM (Industrial, Scientific and Medical) όπως συσκευές Wi-Fi για ασύρματα τοπικά δίκτυα (WLAN, IEEE 802.11), Bluetooth (IEEE 802.15.1), ZigBee(IEEE 802.15.4), ασύρματα τηλέφωνα και άλλες. Αυτές οι συσκευές, που λειτουργούν στις ίδιες ζώνες συχνοτήτων, ανταγωνίζονται μεταξύ τους για το ίδιο περιορισμένο φάσμα. Ενώ στις μη-αδειοδοτημένες συχνότητες αυξάνεται ο συνωστισμός όλο και περισσότερο, ιδιαίτερα σε πυκνοκατοικημένες περιοχές, οι αδειοδοτημένες συχνότητες συνήθως υποχρησιμοποιούνται, δημιουργώντας προσωρινές ευκαιρίες επικοινωνίας στο φάσμα, οι οποίες μεταβάλλονται στο χρόνο και το χώρο.



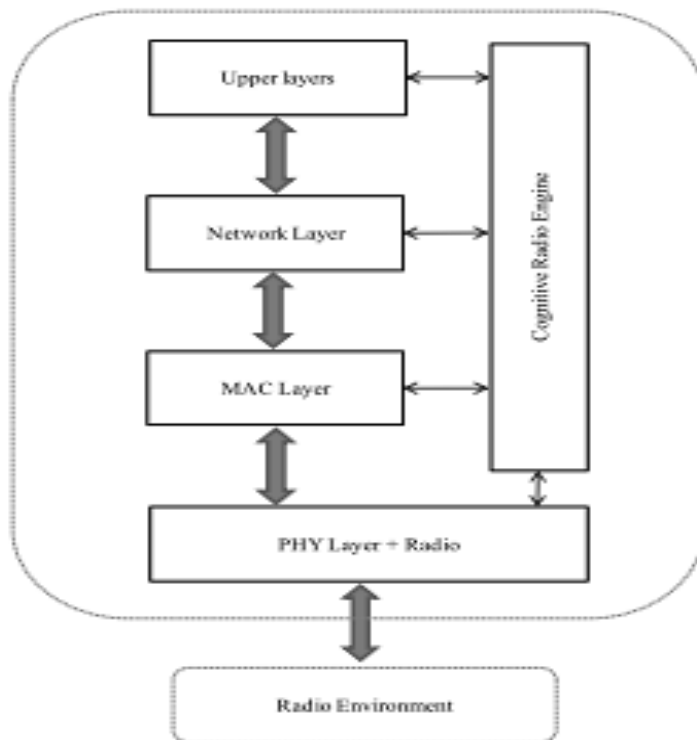
Σχήμα 1-1 – Χρήση φάσματος ραδιοσυχνοτήτων [15].

Τα Γνωστικά Δίκτυα (Cognitive Networks) και οι Γνωστικές Ραδιοσυσκευές (Cognitive Radio) προτείνονται ως λύσεις στο πρόβλημα της στενότητας και της μη αποδοτικής χρήσης του ηλεκτρομαγνητικού φάσματος, αποσκοπώντας στην εκμετάλλευση των αχρησιμοποίητων αδειοδοτημένων συχνοτήτων, που καλούνται «τρύπες φάσματος» (spectrum holes) ή «λευκά διαστήματα» (white spaces). Σκοπός των γνωστικών δικτύων είναι η χρησιμοποίηση των αδειοδοτημένων αλλά μη χρησιμοποιούμενων συχνοτήτων όταν την ίδια στιγμή σε άλλες συχνότητες, αδειοδοτημένες ή μη, υπάρχει συνωστισμός χρηστών. Οι δευτερεύοντες χρήστες (secondary users) πρέπει να μπορούν αυτόνομα να χρησιμοποιούν τις «τρύπες φάσματος» για να αυξήσουν την απόδοσή τους, χωρίς όμως να προκαλούν παρεμβολές στους πρωτεύοντες χρήστες (primary users).



Σχήμα 1-2 - Κενά φάσματος.

Τεχνικές δυναμικής πρόσβασης στο φάσμα (Dynamic Spectrum Access) επιτρέπουν στις γνωστικές ραδιοσυσκευές να λειτουργούν στο καλύτερο κανάλι που είναι εκάστοτε διαθέσιμο. Η γνωστική ραδιοσυσκευή είναι ένας προγραμματιζόμενος πομποδέκτης (SDR – Software Defined Radio) ο οποίος μπορεί να μεταβάλλει έξυπνα τη λειτουργία του και τη χρήση του φάσματος, με κριτήριο τις συνθήκες του ηλεκτρομαγνητικού περιβάλλοντος και σύμφωνα με προκαθορισμένους στόχους, επιλέγοντας ανά πάσα στιγμή και με βέλτιστο τρόπο τα κανάλια, τις παραμέτρους και τα πρωτόκολλα μετάδοσης. Ο όρος Cognitive Radio χρησιμοποιήθηκε για πρώτη φορά το 1999 από τον Mitola [1]. Η τεχνολογία των γνωστικών ραδιοσυσκευών και δικτύων αφορά όλα τα στρώματα του δικτύου αλλά εστιάζεται κυρίως στο φυσικό στρώμα (PHY) και στο στρώμα ζεύξης (MAC).



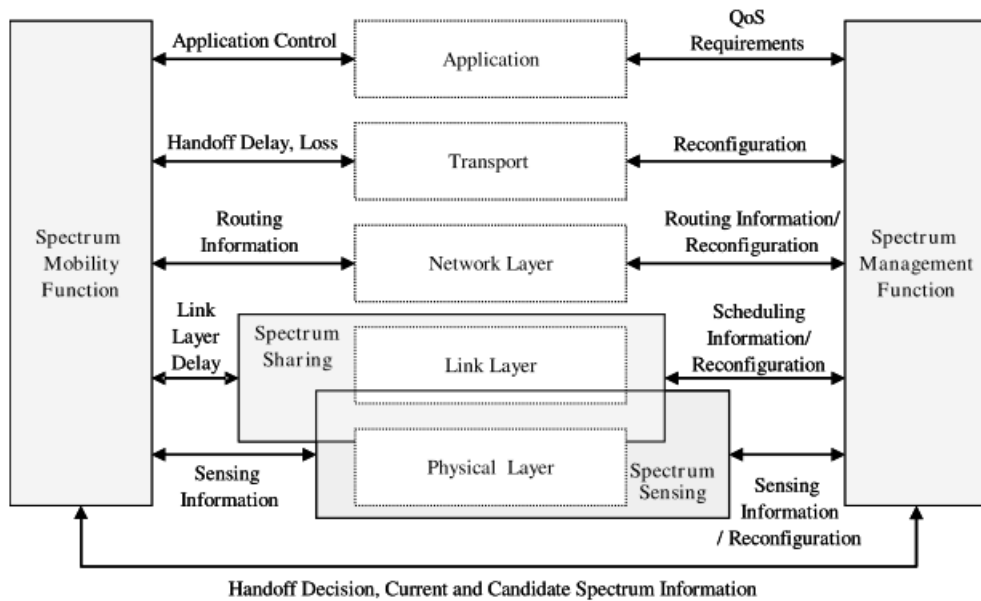
Σχήμα 1-3 - Η αλληλεπίδραση της τεχνολογίας γνωστικών ραδιοσυσκευών με τα διάφορα στρώματα του δικτύου.

Οι κόμβοι των γνωστικών δικτύων (οι γνωστικές ραδιοσυσκευές) πρέπει να επιτελούν τις ακόλουθες λειτουργίες:

- Ανίχνευση φάσματος (spectrum sensing)
- Διαχείριση φάσματος (spectrum management)
- Κινητικότητα στο φάσμα (spectrum mobility)
- Διαμοιρασμός φάσματος (spectrum sharing)

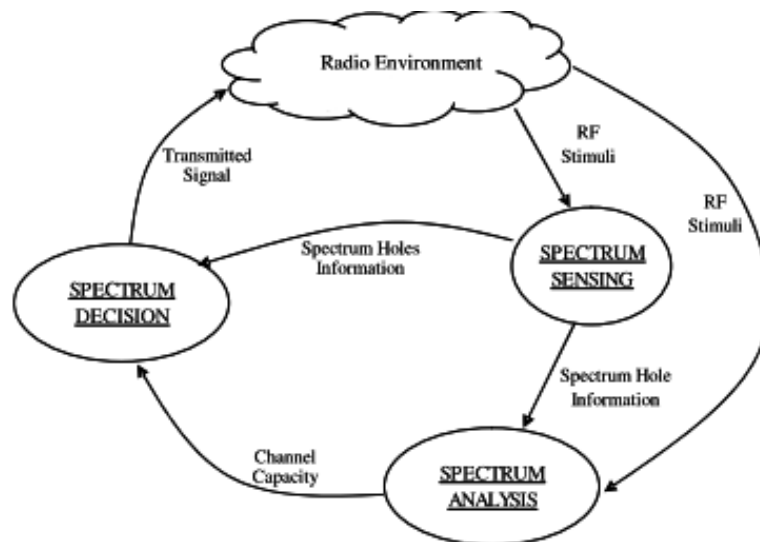
Οι κόμβοι των γνωστικών δικτύων πρέπει να μπορούν να ανακαλύπτουν κενά στο φάσμα (ανίχνευση φάσματος), να διαλέγουν τις καλύτερες φασματικές ευκαιρίες ώστε να καλύψουν τις ανάγκες τους (διαχείριση φάσματος), να μεταβάλλουν τα χαρακτηριστικά τους ώστε να εκμεταλλεύονται τις ευκαιρίες αυτές (κινητικότητα στο φάσμα) και να συγχρονίζουν την πρόσβαση στα επιλεγμένα κανάλια με τους γειτονικούς κόμβους γνωστικών δικτύων (διαμοιρασμός φάσματος). Ταυτόχρονα πρέπει να λειτουργούν χωρίς διακοπές κατά τις εναλλαγές καναλιών και να μην παρεμβάλλονται στην επικοινωνία των πρωτευόντων χρηστών. Για να ανταποκριθούν στις προαναφερθείσες απαιτήσεις αυτόματης ρύθμισης και βελτιστοποίησης των χαρακτηριστικών λειτουργίας τους, οι κόμβοι των γνωστικών δικτύων πρέπει να μπορούν να μαθαίνουν με βάση την προηγούμενη εμπειρία τους (learning based on past experience), να διαθέτουν λογική (reasoning) και κυρίως να μπορούν να μεταβάλλουν δυναμικά τις παραμέτρους μετάδοσης τους. Αυτές οι λειτουργίες μπορούν να πραγματοποιούνται από ένα στρώμα δικτύου, όπως π.χ. η ανίχνευση φάσματος που πραγματοποιείται από το φυσικό στρώμα (PHY), ή να πραγματοποιούνται από περισσότερα στρώματα δικτύου, όπως π.χ. η λειτουργία της κινητικότητας φάσματος που απαιτεί τη συνεργασία όλων των στρωμάτων του δικτύου.

Με βάση την αρχιτεκτονική τους, τα γνωστικά δίκτυα χωρίζονται σε κεντρικά ελεγχόμενα (centralized), γνωστά και ως γνωστικά δίκτυα κορμού (infrastructure network), όπου η εκχώρηση και πρόσβαση στο φάσμα ελέγχονται από μια κεντρική οντότητα, και σε κατανεμημένα (distributed), γνωστά και ως ad-hoc γνωστικά δίκτυα, όπου κάθε κόμβος πρέπει να ενσωματώνει τις ικανότητες ενός cognitive radio και η εκχώρηση φάσματος μπορεί να γίνει είτε συνεργατικά είτε μη-συνεργατικά.



Σχήμα 1-4 - Λειτουργίες γνωστικών ραδιοσυσκευών στα στρώματα του δικτύου.

Η γνωστική ικανότητα των κόμβων γνωστικού δικτύου τους επιτρέπει την αλληλεπίδραση με το περιβάλλον του σε πραγματικό χρόνο, για τον προσδιορισμό των κατάλληλων παραμέτρων και την προσαρμογή τους στο δυναμικό ασύρματο περιβάλλον. Οι ενέργειες που απαιτούνται για την προσαρμοστική λειτουργία είναι γνωστές ως γνωστικός κύκλος (cognitive cycle). Ο γνωστικός κύκλος περιλαμβάνει τρία βήματα: ανίχνευση φάσματος (spectrum sensing), ανάλυση φάσματος (spectrum analysis) και επιλογή φάσματος (spectrum decision).



Σχήμα 1-5 - Ο γνωστικός κύκλος.

Στην φάση της ανίχνευσης φάσματος, ο δέκτης παρατηρεί τις διαθέσιμες ζώνες συχνοτήτων και εντοπίζει τα κενά φάσματος ενώ, στη φάση της ανάλυσης φάσματος, εκτιμώνται τα χαρακτηριστικά των ήδη εντοπισμένων οπών του φάσματος. Τέλος, στη

φάση της επιλογής φάσματος, προσδιορίζεται ο ρυθμός μετάδοσης, η ισχύς εκπομπής, η μέθοδος διαμόρφωσης και το εύρος της ζώνης μετάδοσης και επιλέγεται η κατάλληλη ζώνη συχνοτήτων ανάλογα με τα χαρακτηριστικά του φάσματος και τις ανάγκες επικοινωνίας.

Όταν επιλεγεί η κατάλληλη ζώνη συχνοτήτων, πραγματοποιείται η επικοινωνία σε αυτή. Λόγω της συνεχούς μεταβολής, τόσο στο χρόνο όσο και στο χώρο, του ασύρματου περιβάλλοντος, ο γνωστικός πομποδέκτης πρέπει να αντιλαμβάνεται αυτές τις αλλαγές και να αντιδρά κατάλληλα.

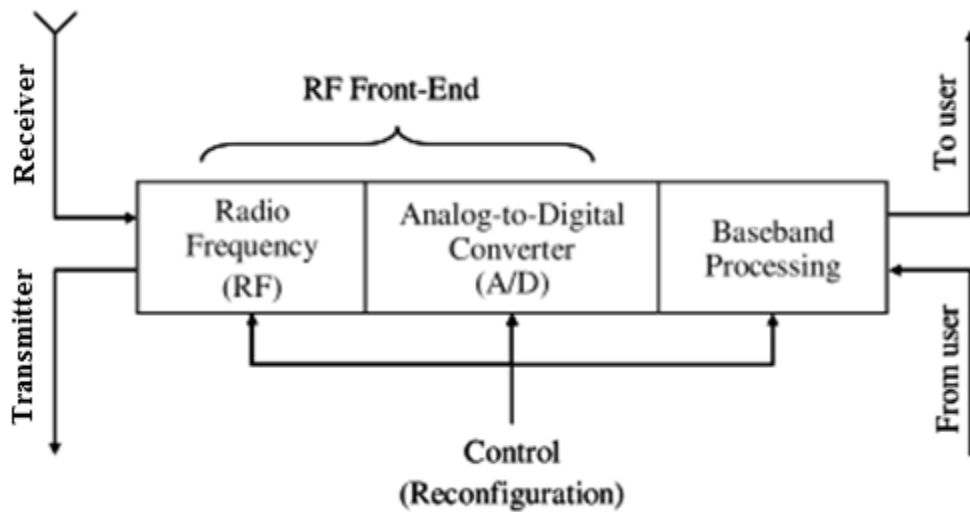
Αναδιάρθρωση (Reconfigurability)

Για να επιτευχθούν τα προαναφερθέντα απαιτείται να υπάρχει η ικανότητα επαναπροσδιορισμού των παραμέτρων λειτουργίας του πομποδέκτη, χωρίς να μεταβάλλονται τμήματα του εξοπλισμού. Η ικανότητα αυτής της αναδιάρθρωσης επιτρέπει στο cognitive radio να προσαρμόζεται εύκολα στο δυναμικό ασύρματο περιβάλλον. Μερικές από τις παραμέτρους που μπορούν να μεταβάλλονται είναι:

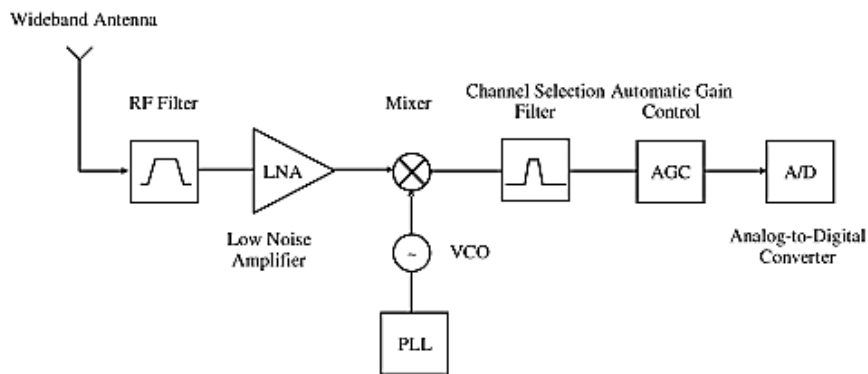
- Συχνότητα λειτουργίας – Με βάση τις πληροφορίες για το ασύρματο περιβάλλον, επιλέγεται η κατάλληλη συχνότητα και η επικοινωνία πραγματοποιείται σε αυτήν.
- Μέθοδος διαμόρφωσης – Το σχήμα διαμόρφωσης μεταβάλλεται ανάλογα με τις ανάγκες του χρήστη και τις συνθήκες του καναλιού. Για παράδειγμα, για εφαρμογές ευαίσθητες σε καθυστερήσεις επιλέγονται σχήματα διαμόρφωσης με υψηλή φασματική απόδοση, ενώ για εφαρμογές ευαίσθητες σε απώλειες, επιλέγονται σχήματα διαμόρφωσης που εγγυώνται χαμηλό ποσοστό σφαλμάτων .
- Ισχύς εκπομπής –Ο έλεγχος της ισχύος εκπομπής επιτρέπει τη δυναμική ρύθμισή της μέσα στα προβλεπόμενα όρια. Όταν δεν είναι αναγκαία η λειτουργία υψηλής ισχύος, ο πομποδέκτης μειώνει την ισχύ εκπομπής για να μειωθούν οι παρεμβολές έτσι ώστε περισσότεροι χρήστες να μπορούν να χρησιμοποιούν το φάσμα .
- Τεχνολογία επικοινωνίας – Ένα cognitive radio μπορεί να χρησιμοποιηθεί για την επίτευξη διαλειτουργικότητας ανάμεσα σε διαφορετικά συστήματα επικοινωνιών.

Οι παράμετροι μετάδοσης ενός γνωστικού πομποδέκτη μπορούν να καθορίζονται κατά την αρχή της μετάδοσης αλλά και να μεταβάλλονται κατά τη διάρκειά της.

1.2 Αρχιτεκτονική ενός cognitive radio



Σχήμα 1-6 - Αναδιάταξη γνωστικής ραδιοσυσκευής.



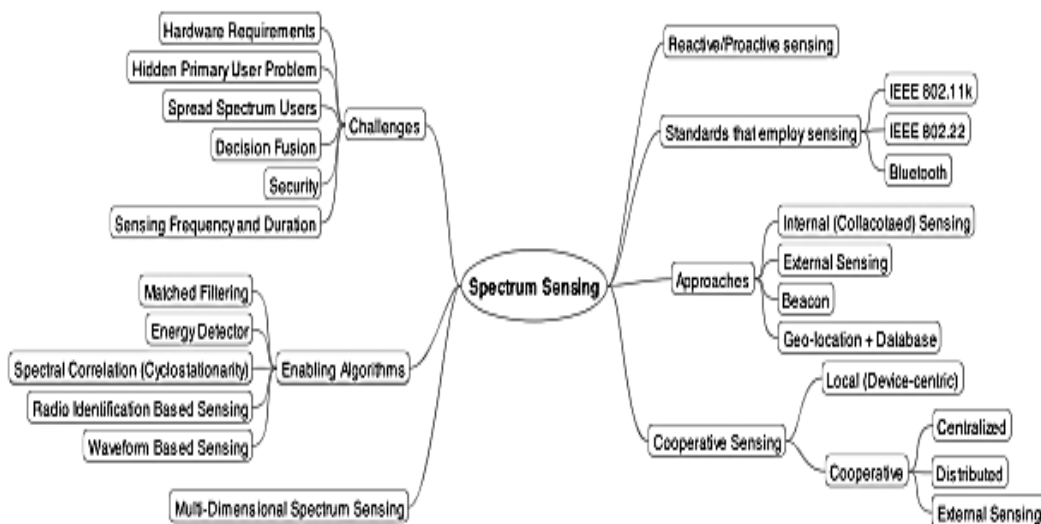
Σχήμα 1-7 - Φυσική αρχιτεκτονική μιας γνωστικής ραδιοσυσκευής.

Οι κύριες μονάδες ενός γνωστικού πομποδέκτη είναι η βαθμίδα ραδιοσυχνοτήτων (RF front-end) και το τμήμα επεξεργασίας βασικής ζώνης (baseband processing unit). Κάθε μονάδα μπορεί να ρυθμιστεί μέσω ενός διαύλου ελέγχου (control bus), ώστε ο πομποδέκτης να προσαρμόζεται στο χρονικά μεταβαλλόμενο RF περιβάλλον. Στην RF βαθμίδα εισόδου το σήμα λήψης ενισχύεται και μετατρέπεται από αναλογικό σε ψηφιακό. Το τμήμα επεξεργασίας βασικής ζώνης ενός γνωστικού πομποδέκτη, δεν διαφέρει ουσιαστικά από τους κλασικούς πομποδέκτες, αφού και εδώ το σήμα διαμορφώνεται/αποδιαμορφώνεται και κωδικοποιείται/αποκωδικοποιείται. Το νέο χαρακτηριστικό των γνωστικών πομποδεκτών είναι η δυνατότητα ευρυζωνικής ανίχνευσης φάσματος της RF βαθμίδας εισόδου. Αυτή η δυνατότητα σχετίζεται κυρίως με τεχνολογίες υλικού RF όπως ευρυζωνικές κεραίες, ενισχυτές ισχύος και προσαρμοστικά φίλτρα. Το υλικό της βαθμίδας ραδιοσυχνοτήτων ενός γνωστικού πομποδέκτη πρέπει να μπορεί να συντονιστεί σε οποιοδήποτε τμήμα μιας ευρείας ζώνης συχνοτήτων, επιτρέποντας έτσι την ανίχνευση και αξιολόγηση του φάσματος σε πραγματικό χρόνο.

Σε αυτή την αρχιτεκτονική ένα ευρυζωνικό σήμα λαμβάνεται από το τμήμα RF, δειγματοληπτείται από τον υψηλής ταχύτητας μετατροπέα Αναλογικού/Ψηφιακού σήματος (A/D converter) ώστε να καταστεί δυνατός ο εντοπισμός σημάτων των αδειοδοτημένων χρηστών. Η ευρυζωνική RF κεραία λαμβάνει σήματα από πολλαπλές εκπομπές, με διαφορετικά επίπεδα ισχύος, διαφορετικά εύρη ζώνης και από διαφορετικές διευθύνσεις. Για το λόγο αυτό η βαθμίδα RF πρέπει να έχει τη δυνατότητα να εντοπίζει ασθενή σήματα σε μια μεγάλη δυναμική περιοχή, κάτι το οποίο απαιτεί μετατροπείς A/D ταχύτητας πολλών GHz, με υψηλή ανάλυση. Με τη σειρά του, ο μετατροπέας υψηλής ταχύτητας απαιτεί το εύρος του σήματος να έχει μειωθεί όσο το δυνατόν περισσότερο πριν γίνει η μετατροπή, κάτι το οποίο μπορεί να επιτευχθεί με προσαρμοζόμενα φίλτρα.

1.3 Τεχνικές Ανίχνευσης Φάσματος (Spectrum Sensing)

Ένα από τα σημαντικότερα προβλήματα στα γνωστικά δίκτυα είναι η ανίχνευση φάσματος που αποσκοπεί στην απόκτηση γνώσης σχετικά με τη χρήση του ηλεκτρομαγνητικού φάσματος και την παρουσία πρωτευόντων χρηστών στην περιοχή. Ο προσδιορισμός των στατιστικών χαρακτηριστικών χρήσης του φάσματος πρέπει να γίνει σε πολλές διαστάσεις, όπως ο χρόνος, ο χώρος, η συχνότητα και ο κώδικας. Επίσης, πρέπει να προσδιοριστεί ο τύπος των σημάτων που καταλαμβάνουν το φάσμα και τα χαρακτηριστικά τους όπως διαμόρφωση, εύρος ζώνης, η συχνότητα φέροντος κλπ. Όλα τα ανωτέρω απαιτούν ισχυρές τεχνικές ανάλυσης σημάτων και μεγάλη υπολογιστική δύναμη.



Σχήμα 1-8 - Μέθοδοι ανίχνευσης φάσματος.

Οι παραδοσιακές μέθοδοι ανίχνευσης φάσματος μπορούν να εντοπίσουν σήματα σε 3 διαστάσεις του φάσματος: συχνότητα, χρόνο και χώρο. Αυτές οι μέθοδοι όμως δεν

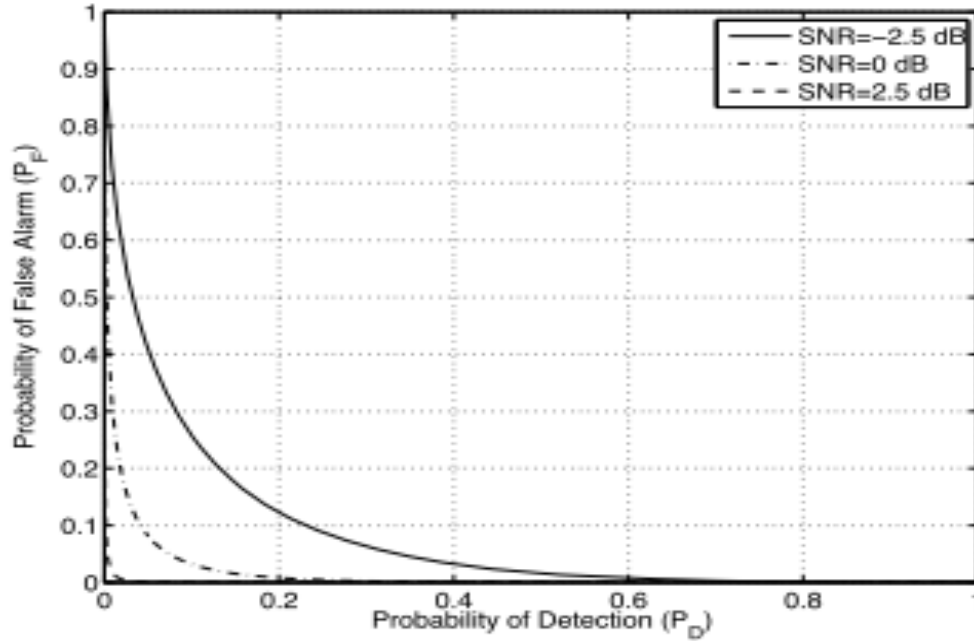
λαμβάνουν υπόψη άλλες διαστάσεις του σήματος (όπως π.χ. τον κώδικα σημάτων διασκορπισμένου φάσματος), και συνεπώς δεν μπορούν να ανιχνεύσουν αποτελεσματικά σήματα διασκορπισμένα στο φάσμα (spread spectrum signals) ή κώδικες μεταπήδησης συχνότητας (frequency hopping codes). Για το λόγο αυτό πρέπει να οριστεί ένας πολυδιάστατος χώρος για την ανίχνευση φάσματος καθώς πρέπει να ανιχνεύεται η χρησιμοποίηση του φάσματος σε όλες τις πιθανές διαστάσεις.

Ακολουθούν οι περισσότερο διαδομένες τεχνικές ανίχνευσης φάσματος που αναφέρονται στην βιβλιογραφία των γνωστικών δικτύων.

1.3.1 Ανιχνευτής Ενέργειας (Energy Detector)

Η τεχνική αυτή είναι γνωστή και ως ραδιομετρία ή περιοδόγραμμα, και αποτελεί κοινή τεχνική ανίχνευσης φάσματος λόγω της χαμηλής υπολογιστικής πολυπλοκότητας και των μικρών απαιτήσεων της σε υλικό. Επιπλέον, είναι γενικής εφαρμογής καθώς οι δέκτες δεν χρειάζεται να έχουν γνώση του σήματος του πρωτεύοντος χρήστη.

Το σήμα του πρωτεύοντος χρήστη ανιχνεύεται συγκρίνοντας την έξοδο του ανιχνευτή ενέργειας με ένα κατώφλι το οποίο εξαρτάται από το επίπεδο θορύβου που συνήθως επικρατεί στο δέκτη. Ο ανιχνευτής ενέργειας παρουσιάζει ορισμένες σχεδιαστικές προκλήσεις όπως είναι η επιλογή του κατωφλίου ανίχνευσης σήματος, η δυσκολία διαχωρισμού παρεμβολών, εκπομπών πρωτευόντων χρηστών και θορύβου, καθώς και οι χαμηλές επιδόσεις του ανιχνευτή ενέργειας σε χαμηλές σηματοθορυβικές σχέσεις στην είσοδο του δέκτη. Επιπλέον, ο ανιχνευτής ενέργειας δεν μπορεί να εντοπίσει αποτελεσματικά σήματα που είναι απλωμένα στο φάσμα (spread spectrum).



Σχήμα 1-9 - Πιθανότητα εντοπισμού του ανιχνευτή ενέργειας.

Έστω ότι το σήμα λήψης είναι της μορφής

$$y(n) = s(n) + w(n) \quad (1.1)$$

όπου $s(n)$ είναι το προς εντοπισμό σήμα και $w(n)$ είναι λευκός προσθετικός Gaussian θόρυβος (AWGN). Όταν $s(n) = 0$ τότε δεν υπάρχει εκπομπή σήματος από πρωτεύοντα χρήστη. Το μέτρο για λήψη απόφασης για τον ανιχνευτή ενέργειας μπορεί να γραφεί υπό τη μορφή

$$M = \sum_{n=0}^N |y(n)|^2 \quad (1.2)$$

όπου N το μέγεθος του διανύσματος παρατήρησης. Η απόφαση για το αν το κανάλι χρησιμοποιείται γίνεται συγκρίνοντας το μέτρο M με ένα κατώφλι λ_E . Αν ισχύει $M > \lambda_E$, το κανάλι θεωρείται κατελημμένο. Η απόδοση του αλγορίθμου ανίχνευσης μπορεί να συνοψιστεί σε 2 πιθανότητες: την πιθανότητα ορθής ανίχνευσης P_D και την πιθανότητα λανθασμένης ανίχνευσης P_F . Είναι επιθυμητό η πιθανότητα ορθής ανίχνευσης P_D να είναι όσο το δυνατό υψηλότερη και η πιθανότητα λανθασμένης ανίχνευσης P_F όσο το δυνατό χαμηλότερη. Το κατώφλι λ_E μπορεί να επιλεγεί τόσο ώστε να υπάρχει βέλτιστη ισορροπία μεταξύ P_D και P_F . Αυτό απαιτεί γνώση της ισχύος τόσο του θορύβου όσο και των σημάτων των πρωτευόντων χρηστών. Στην πράξη, το όριο λ_E επιλέγεται έτσι ώστε να επιτευχθεί συγκεκριμένη πιθανότητα λανθασμένης ανίχνευσης, οπότε η γνώση της διακύμανσης του θορύβου επαρκεί για την επιλογή του κατωφλίου.

1.3.2 Ανίχνευση κυματομορφής (Waveform-Based Sensing)

Γνωστές ακολουθίες συμβόλων (patterns) χρησιμοποιούνται συνήθως στις ασύρματες επικοινωνίες για να υποβοηθήσουν το συγχρονισμό ή άλλες λειτουργίες (π.χ. εκτίμηση του καναλιού). Τέτοιες ακολουθίες είναι οι επικεφαλίδες πακέτων δεδομένων (preambles), συχνά μεταδιδόμενες πιλοτικές ακολουθίες (pilot patterns), κώδικες διασποράς (spreading codes) κλπ. Όταν το σήμα περιλαμβάνει γνωστή ακολουθία, η ανίχνευση μπορεί να γίνει συσχετίζοντας το σήμα λήψης με ένα αντίγραφο της γνωστής ακολουθίας. Αυτή η μέθοδος έχει εφαρμογή μόνο σε συστήματα όπου υπάρχουν γνωστές ακολουθίες συμβόλων και ονομάζεται ανίχνευση κυματομορφής (waveform-based sensing) ή συνεκτική ανίχνευση (coherent sensing). Η ανίχνευση κυματομορφής έχει γενικά καλύτερη απόδοση από τον ανιχνευτή ενέργειας από άποψη αξιοπιστίας και χρόνου σύγκλισης. Επιπλέον, η απόδοση των αλγορίθμων ανίχνευσης αυξάνεται όσο αυξάνεται το μήκος της γνωστής ακολουθίας.

Χρησιμοποιώντας το ίδιο μοντέλο με την εξίσωση (1.1), το μέτρο ανίχνευσης είναι της μορφής

$$M = \text{Re} \left[\sum_{n=1}^N y(n)s^*(n) \right] \quad (1.3)$$

όπου με $s^*(n)$ συμβολίζεται ο συζυγής του $s(n)$. Όταν δεν εκπέμπει πρωτεύων χρήστης στο κανάλι το μέτρο ανίχνευσης είναι ίσο με

$$M = \text{Re} \left[\sum_{n=1}^N w(n)s^*(n) \right] \quad (1.4)$$

ενώ όταν εκπέμπει πρωτεύων χρήστης στο κανάλι το μέτρο ανίχνευσης είναι

$$\begin{aligned} M &= \text{Re} \left[\sum_{n=1}^N (s(n) + w(n))s^*(n) \right] = \text{Re} \left[\sum_{n=1}^N s(n)s^*(n) + \sum_{n=1}^N w(n)s^*(n) \right] \\ &= \sum_{n=1}^N |s(n)|^2 + \text{Re} \left[\sum_{n=1}^N w(n)s^*(n) \right] \end{aligned} \quad (1.5)$$

Η απόφαση για το αν υπάρχει σήμα πρωτεύοντος χρήστη στο κανάλι λαμβάνεται συγκρίνοντας το M με ένα κατώφλι λ_w .

1.3.3 Ανίχνευση Κυκλοστατικότητας (Cyclostationarity-Based Sensing)

Η ανίχνευση χαρακτηριστικών κυκλοστατικότητας αποτελεί μέθοδο ανίχνευσης των εκπομπών πρωτευόντων χρηστών που εκμεταλλεύεται τα χαρακτηριστικά κυκλοστατικότητας των λαμβανόμενων σημάτων. Τα χαρακτηριστικά κυκλοστατικότητας

οφείλονται στην περιοδικότητα του σήματος ή των στατιστικών χαρακτηριστικών του, όπως η μέση τιμή και η αυτοσυσχέτιση (autocorrelation), ή μπορεί να έχουν ενσωματωθεί στο σήμα ακριβώς για το σκοπό της ανίχνευσης φάσματος. Για τον εντοπισμό σημάτων στο ραδιοφάσμα, αντί της φασματικής πυκνότητας ισχύος (Power spectral density, PSD), χρησιμοποιείται η κυκλική συνάρτηση συσχέτισης (cyclic correlation function). Οι αλγόριθμοι ανίχνευσης κυκλοστατικότητας μπορούν να διαχωρίσουν το θόρυβο από τα σήματα των πρωτευόντων χρηστών, αφού ο θόρυβος είναι στατικός υπό την ευρεία έννοια (wide-sense stationary, WSS) με χρονικά στατική φασματική πυκνότητα ισχύος ενώ τα διαμορφωμένα σήματα είναι συνήθως μη στατικά αλλά κυκλοστατικά με φασματική πυκνότητα ισχύος που μεταβάλλεται περιοδικά με συχνότητα κύκλου (cycle frequency) ίση με α , η οποία σχετίζεται με την περιοδικότητα του σήματος. Πράγματι, πολλά τηλεπικοινωνιακά σήματα λόγω της επεξεργασίας που υφίστανται, όπως δειγματοληψία, διαμόρφωση, πολυπλεξία και κωδικοποίηση, εμφανίζουν περιοδικότητα των στατιστικών τους παραμέτρων (μέση τιμή, συνάρτηση αυτοσυσχέτισης κ.α.). Για παράδειγμα το διακριτό PAM σήμα $s(t) = \sum a_k p(t - kT_0)$ με WSS ακολουθία συμβόλων $\{a_k\}$, ή το DSB σήμα $s(t) = m(t) \cos(2\pi f_c t)$ με $m(t)$ WSS, είναι κυκλοστατικά διότι έχουν περιοδική τη μέση τιμή και τη συνάρτηση αυτοσυσχέτισης με συχνότητες κύκλου $\alpha = 1/T_0$ και $\alpha = f_c$, αντίστοιχα. Επίσης, η ύπαρξη κυκλοστατικότητας μπορεί να χρησιμοποιηθεί για τη διαφοροποίηση διαφορετικών τύπων μεταδόσεων και διαφορετικών πρωτευόντων χρηστών.

Η συνάρτηση κυκλικής φασματικής πυκνότητας (Cyclic spectral density, CSD) ενός σήματος λήψης της μορφής (1.1) μπορεί να υπολογιστεί μέσω της σχέσης

$$S(f, \alpha) = \sum_{\tau=-\infty}^{\infty} R_y^{\alpha}(\tau) e^{-j2\pi f \tau} \quad (1.6)$$

όπου

$$R_y^{\alpha}(\tau) = E[y(n + \tau)y^*(n - \tau)e^{j2\pi \alpha n}] \quad (1.7)$$

είναι η κυκλική συνάρτηση αυτοσυσχέτισης (cyclic autocorrelation function, CAF) και α η συχνότητα κύκλου (cycle frequency) ενώ f είναι η φασματική συχνότητα. Η συνάρτηση κυκλικής φασματικής πυκνότητας (CSD) λαμβάνει μέγιστη τιμή όταν η συχνότητα κύκλου είναι ίση με τις βασικές συχνότητες του μεταδιδόμενου σήματος $s(n)$. Οι συχνότητες κύκλου μπορεί να θεωρηθούν γνωστές ή μπορεί να εξαχθούν και να χρησιμοποιηθούν ως χαρακτηριστικά αναγνώρισης μεταδιδόμενων σημάτων.

1.3.4 Ανίχνευση Ραδιοαναγνώρισης (Radio Identification Based Sensing)

Πλήρης γνώση των φασματικών χαρακτηριστικών μπορεί να αποκτηθεί αναγνωρίζοντας τις τεχνολογίες μετάδοσης που χρησιμοποιούν οι πρωτεύοντες χρήστες. Μια τέτοια ταυτοποίηση επιτρέπει στους κόμβους γνωστικών δικτύων να ανιχνεύσουν το φάσμα σε περισσότερες διαστάσεις και με μεγαλύτερη ακρίβεια. Για παράδειγμα, αν ένας πρωτεύων χρήστης διαπιστωθεί ότι χρησιμοποιεί Bluetooth, ένας δευτερεύων χρήστης μπορεί να αξιοποιήσει αυτή την πληροφορία γνωρίζοντας ότι το Bluetooth έχει εμβέλεια περίπου 10 μέτρα και ότι χρησιμοποιεί μεταπήδηση συχνότητας.

Στην ανίχνευση ραδιοαναγνώρισης πολλά χαρακτηριστικά του σήματος λήψης προσδιορίζονται και χρησιμοποιούνται για την εκτίμηση της πιθανότερης τεχνολογίας εκπομπής του πρωτεύοντος χρήστη, χρησιμοποιώντας διάφορες μεθόδους ταξινόμησης. Χαρακτηριστικά του σήματος λήψης που χρησιμοποιούνται για την αναγνώριση του είναι η ενέργεια του, η κατανομή του στο φάσμα, το εύρος ζώνης, οι κεντρικές συχνότητες, η τυπική απόκλιση της στιγμιαίας συχνότητας, η μέγιστη χρονική διάρκεια του σήματος, οι κυκλικές συχνότητες και τα κυκλοστατικά χαρακτηριστικά του σήματος. Από τα παραπάνω το σημαντικότερο χαρακτηριστικό είναι το εύρος ζώνης του σήματος.

1.3.5 Ανίχνευση προσαρμοσμένου φίλτρου (Matched Filtering)

Τα προσαρμοσμένα φίλτρα αποτελούν τη βέλτιστη μέθοδο εντοπισμού πρωτευόντων χρηστών όταν η μορφή του μεταδιδόμενου σήματος είναι γνωστή. Το μεγάλο πλεονέκτημα των προσαρμοσμένων φίλτρων είναι ο μικρός χρόνος που απαιτείται για να επιτευχθεί συγκεκριμένη πιθανότητα λανθασμένης ανίχνευσης σε σύγκριση με τις υπόλοιπες μεθόδους.

Εντούτοις, η χρήση προσαρμοσμένων φίλτρων απαιτεί την αποδιαμόρφωση των σημάτων λήψης και, επομένως, την ακριβή γνώση των χαρακτηριστικών μετάδοσης των πρωτευόντων χρηστών όπως είναι το εύρος ζώνης, η συχνότητα λειτουργίας, το είδος και η τάξη της διαμόρφωσης, η μορφοποίηση παλμών και η μορφή πλαισίων. Επίσης, οι κόμβοι γνωστικών δικτύων χρειάζονται δέκτες για όλα τα είδη σημάτων, απαίτηση που καθιστά την πολυπλοκότητα εφαρμογής του προσαρμοσμένου φίλτρου πολύ υψηλή. Ένα άλλο μειονέκτημα των προσαρμοσμένων φίλτρων είναι η υψηλή κατανάλωση ενέργειας καθώς, για να υπάρξει ανίχνευση, πολυάριθμοι αλγόριθμοι πρέπει να εκτελεστούν στο δέκτη.

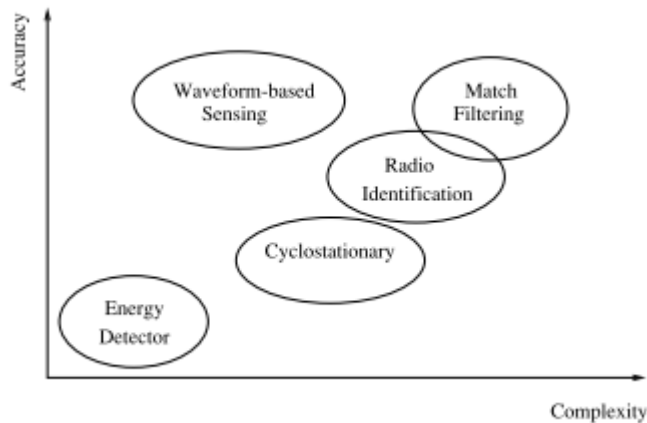
1.3.6 Άλλες μέθοδοι ανίχνευσης

Άλλες μέθοδοι ανίχνευσης φάσματος είναι η πολυκωνική εκτίμηση φάσματος (multi-taper signal estimation), η εκτίμηση μετασχηματισμού κυματιδίων (wavelet transform

based estimation), ο μετασχηματισμός Hough και χρόνο-συχνοτική ανάλυση (time-frequency analysis).

1.3.7 Σύγκριση μεθόδων ανίχνευσης

Μια βασική σύγκριση των προαναφερθεισών μεθόδων φαίνεται στο σχήμα 1-10:



Σχήμα 1-10 - Σύγκριση μεθόδων ανίχνευσης φάσματος.

Η μέθοδος ανίχνευσης με βάση την κυματομορφή είναι περισσότερο ακριβής τόσο από την τεχνική ανίχνευσης ενέργειας όσο και από την ανίχνευση κυκλοστατικότητας. Όμως, απαιτεί a priori πληροφορία για τα χαρακτηριστικά μετάδοσης των πρωτεύοντων χρηστών. Η απόδοση του ανιχνευτή ενέργειας είναι περιορισμένη όταν 2 κοινές παραδοχές, η στατικότητα του θορύβου και η γνώση της διακύμανσής του, δεν ισχύουν. Όταν ο θόρυβος είναι στατικός, οι μέθοδοι ανίχνευσης κυκλοστατικότητας έχουν χειρότερη επίδοση από τον ανιχνευτή ενέργειας. Όταν υπάρχουν παρεμβολές στο ίδιο κανάλι(CCI) ή σε γειτονικά κανάλια(ACI), ο θόρυβος γίνεται μη-στατικός, οπότε υπερτερούν οι κυκλοστατικές μέθοδοι. Από την άλλη πλευρά, τα χαρακτηριστικά κυκλοστατικότητας ενός σήματος μπορεί να χαθούν λόγω διαλείψεων και σκίασης, οπότε κάτω από ένα κατώφλι σηματοθορυβικού λόγου, παρόμοιο με αυτό στους ανιχνευτές ενέργειας, όπου η πιθανότητα ανίχνευσης γίνεται πολύ μικρή. Επιπλέον, οι μέθοδοι ανίχνευσης κυκλοστατικότητας είναι ευάλωτες σε αποκλίσεις (offsets) του ρολογιού δειγματοληψίας.

Κατά την επιλογή μεθόδου ανίχνευσης φάσματος πρέπει να λαμβάνονται υπόψη τόσο τα χαρακτηριστικά των πρωτεύοντων χρηστών όσο και η επιθυμητή ακρίβεια, η υπολογιστική πολυπλοκότητα και οι ανάγκες του δικτύου (π.χ. συχνότητα και ποσότητα ανταλλασσόμενων πληροφοριών μεταξύ γνωστικών χρηστών).

1.4 Συνεργατική Ανίχνευση (Cooperative Sensing)

Η συνεργασία μεταξύ χρηστών γνωστικών δικτύων προτείνεται ως λύση στα προβλήματα που προκύπτουν λόγω αβεβαιοτήτων σχετικά με το θόρυβο, τις διαλείψεις και τις σκιάσεις. Η συνεργατική ανίχνευση μειώνει σημαντικά την πιθανότητα λανθασμένης ανίχνευσης, μπορεί να επιλύσει το πρόβλημα του «κρυμμένου» πρωτεύοντος χρήστη (hidden primary user problem) και μπορεί να μειώσει το χρόνο ανίχνευσης. Ο διαμοιρασμός των αποτελεσμάτων της ανίχνευσης και των πληροφοριών ως προς την κατανομή των καναλιών μπορεί να γίνει μέσω ενός καναλιού ελέγχου (control channel). Η συνεργατική ανίχνευση φάσματος είναι αποτελεσματικότερη όταν οι συνεργαζόμενοι χρήστες παρατηρούν ανεξάρτητες διαλείψεις. Επιπλέον, είναι περισσότερο επωφελές ο ίδιος αριθμός συνεργαζόμενων χρηστών να είναι διασκορπισμένος σε μεγάλη περιοχή αντί για μικρή.

Η συνεργασία μεταξύ χρηστών γνωστικών δικτύων ή εξωτερικών αισθητήρων μπορεί να χρησιμοποιηθεί για να δημιουργηθεί ένα δίκτυο συνεργατικής ανίχνευσης (cooperative sensing network). Αυτή η συνεργασία μπορεί να έχει δύο μορφές: κεντρικά ελεγχόμενη (centralized) ή καταναμημένη (distributed).

Στην κεντρικά ελεγχόμενη ανίχνευση ένας κεντρικός χρήστης λαμβάνει τις πληροφορίες ανίχνευσης από όλους τους χρήστες, αναγνωρίζει τα φασματικά κενά και μεταδίδει την πληροφορία αυτή στους χρήστες ή ελέγχει απευθείας τη δικτυακή κίνηση. Στην περίπτωση, όμως, όπου οι χρήστες είναι πολλοί, το εύρος ζώνης που απαιτείται για την ανταλλαγή των πληροφοριών ανίχνευσης γίνεται μεγάλο. Για το λόγο αυτό, μόνο οι χρήστες οι οποίοι έχουν αξιόπιστες πληροφορίες επιτρέπεται να τις μεταδώσουν στον κεντρικό χρήστη.

Στην καταναμημένη ανίχνευση οι χρήστες μοιράζονται μεταξύ τους τις πληροφορίες για το φάσμα, αλλά αποφασίζουν μόνοι τους ποιο κανάλι θα χρησιμοποιήσουν. Η καταναμημένη ανίχνευση υπερτερεί της κεντρικής γιατί δεν απαιτεί κάποιο δίκτυο υποδομής και έχει μειωμένο κόστος.

1.5 Εξωτερική ανίχνευση (External Sensing)

Μια άλλη τεχνική ανίχνευσης φάσματος είναι η εξωτερική ανίχνευση όπου ένας εξωτερικός πράκτορας πραγματοποιεί την ανίχνευση και μεταδίδει τις πληροφορίες για την κατάληψη των καναλιών στους χρήστες γνωστικών δικτύων. Αλγόριθμοι εξωτερικής ανίχνευσης επιλύουν ορισμένα από τα προβλήματα που σχετίζονται με την ανίχνευση που πραγματοποιείται από τους πομποδέκτες των χρηστών γνωστικών δικτύων. Τα κύρια πλεονεκτήματα της εξωτερικής ανίχνευσης είναι η λύση του προβλήματος του κρυμμένου πρωτεύοντος χρήστη, ο οποίος ενδεχομένως να μη γινόταν αντιληπτός από μερικούς δευτερεύοντες χρήστες, και η μείωση της αβεβαιότητας λόγω διαλείψεων. Επιπλέον, οι

δευτερεύοντες χρήστες δεν σπαταλούν χρόνο για ανίχνευση φάσματος και έτσι αυξάνεται η φασματική αποδοτικότητά τους.

1.6 Χρησιμοποίηση του ιστορικού για προβλέψεις

Για τη μείωση της παρεμβολής στους πρωτεύοντες χρήστες με ταυτόχρονη αποδοτική εκμετάλλευση του ελεύθερου φάσματος, οι χρήστες γνωστικών δικτύων πρέπει να καταγράφουν τις μεταβολές στη διαθεσιμότητα του φάσματος και να πραγματοποιούν προβλέψεις. Ένας γνωστικός πομποδέκτης παρακολουθεί το φάσμα με σταθερό ρυθμό και έχει τη δυνατότητα μάθησης. Επομένως, ένα ιστορικό της χρήσης του φάσματος μπορεί να χρησιμοποιηθεί για να προβλεφθεί η μελλοντική του χρήση. Για το σκοπό αυτό μπορούν να χρησιμοποιηθούν αλγόριθμοι πρόβλεψης οι οποίοι βασίζονται σε στατιστική ανάλυση.

1.7 Διαχείριση Φάσματος (Spectrum Management)

Αχρησιμοποίητες ζώνες συχνοτήτων κατανέμονται σε όλη την έκταση του φάσματος, τόσο σε αδειοδοτημένες ζώνες όσο και σε μη-αδειοδοτημένες. Αυτές οι αχρησιμοποίητες ζώνες, που εντοπίζονται μέσω της διαδικασίας ανίχνευσης φάσματος, εμφανίζουν διαφορετικά χαρακτηριστικά τα οποία, μάλιστα, μεταβάλλονται με το χρόνο. Για την επιλογή της κατάλληλης ζώνης συχνοτήτων που θα χρησιμοποιηθεί από τη γνωστική ραδιοσυσκευή απαιτούνται τα άλλα δύο βήματα του γνωστικού κύκλου, η ανάλυση φάσματος και η επιλογή του φάσματος. Τα βήματα αυτά, σε αντίθεση με την ανίχνευση φάσματος που ασχολείται κυρίως με το φυσικό στρώμα (PHY), είναι στενά συνδεδεμένα και επηρεάζουν τις λειτουργίες των υψηλότερων στρωμάτων του δικτύου, όπως για παράδειγμα η χωρητικότητα του διαύλου.

1.7.1 Ανάλυση φάσματος (Spectrum Analysis)

Η ανάλυση φάσματος επιτρέπει στο γνωστικό πομποδέκτη να αναγνωρίσει τα χαρακτηριστικά και να αξιολογήσει με βάση συγκεκριμένα κριτήρια τις διαθέσιμες ζώνες συχνοτήτων που μπορεί να χρησιμοποιήσει. Κάθε διαθέσιμη περιοχή (κενό) του φάσματος χαρακτηρίζεται όχι μόνο από το χρονικά μεταβαλλόμενο ασύρματο περιβάλλον διάδοσης, αλλά και από τη δραστηριότητα των πρωτευόντων χρηστών. Για να χαρακτηριστεί η ποιότητα μιας ελεύθερης ζώνης συχνοτήτων και για να εκτιμηθεί η χωρητικότητά της, δηλαδή ο διαθέσιμος ρυθμός μετάδοσης δεδομένων, χρησιμοποιούνται τα ακόλουθα μεγέθη:

- *Παρεμβολές* – Ορισμένες ζώνες συχνοτήτων εμφανίζουν μεγαλύτερη δραστηριότητα από τις υπόλοιπες, εμφανίζοντας έτσι δυσμενή χαρακτηριστικά παρεμβολών. Από το βαθμό παρεμβολής στους πρωτεύοντες χρήστες,

προσδιορίζεται η επιτρεπόμενη ισχύς εκπομπής η οποία χρησιμοποιείται για την εκτίμηση της χωρητικότητας του καναλιού.

- *Απώλειες ζεύξης* – Οι απώλειες της ζεύξης αυξάνονται όσο αυξάνεται η συχνότητα λειτουργίας. Έτσι, αν η ισχύς μετάδοσης παραμένει σταθερή, η εμβέλεια του γνωστικού πομποδέκτη μειώνεται στις υψηλότερες συχνότητες. Η αύξηση, όμως, της ισχύος εκπομπής, για την αντιστάθμιση των απωλειών μιας ζεύξης οδηγεί συνήθως σε μεγαλύτερα επίπεδα παρεμβολής στους υπόλοιπους χρήστες, γεγονός που υποβαθμίζει την καταλληλότητα της.
- *Λάθη ασύρματης ζεύξης* – Το ποσοστό λαθών μεταβάλλεται αντίστοιχα με το σχήμα διαμόρφωσης και το επίπεδο παρεμβολών του επιλεγμένου καναλιού.
- *Καθυστέρηση στρώματος ζεύξης* – Για να αντιμετωπιστούν διαφορετικά επίπεδα απωλειών ζεύξης, λαθών και παρεμβολών, απαιτούνται διαφορετικοί τύποι πρωτοκόλλων στρώματος ζεύξης για διαφορετικά κανάλια. Αυτό οδηγεί σε διαφορετική καθυστέρηση μετάδοσης πακέτων.
- *Διάρκεια κατάληψης καναλιού* – Οι δραστηριότητες των πρωτευόντων χρηστών επηρεάζουν την ποιότητα ενός καναλιού. Η διάρκεια κατάληψης καναλιού αναφέρεται στην αναμενόμενη διάρκεια που η γνωστική ραδιοσυσκευή μπορεί να καταλάβει μια αδειοδοτημένη συχνότητα πριν αναγκαστεί να διακόψει την εκπομπή της. Όσο μεγαλύτερη είναι η διάρκεια κατάληψης τόσο υψηλότερη η ποιότητα επικοινωνίας στο κανάλι αυτό.

Η χωρητικότητα του καναλιού, που μπορεί να προκύψει μέσω των ανωτέρω παραμέτρων, αποτελεί το σημαντικότερο παράγοντα χαρακτηρισμού ενός καναλιού. Συνήθως, για την εκτίμηση της χωρητικότητας χρησιμοποιείται ο σηματοθορυβικός λόγος λήψης στο δέκτη του δευτερεύοντος χρήστη, που, όμως, λαμβάνει υπόψη μόνο τις τοπικές εκτιμήσεις των χρηστών γνωστικών δικτύων οι οποίες δεν επαρκούν για να εξασφαλιστεί η αποφυγή παρεμβολών στους δέκτες των πρωτευόντων χρηστών, επειδή συνήθως δεν είναι γνωστή η σχετική θέση τους. Εκτός από τη χωρητικότητα του καναλιού, παράγοντες όπως το ποσοστό λαθών της ζεύξης με τη συνεπακόλουθη καθυστέρηση και η διάρκεια κατάληψης του καναλιού έχουν σημαντική επίπτωση στην ποιότητα των υπηρεσιών και, κατά συνέπεια, στην επιλογή του καταλλήλου καναλιού για την επικοινωνία των δευτερευόντων χρηστών. Επιπλέον, η χωρητικότητα είναι στενά συνδεδεμένη με το επίπεδο παρεμβολών και τις απώλειες ζεύξης, καθώς ενδεχόμενη αύξησή της συνεπάγεται αύξηση της ισχύος εκπομπής των δευτερευόντων χρηστών. Μια πλήρης ανάλυση και ένα πλήρες μοντέλο του φάσματος σε γνωστικά δίκτυα δεν έχει αναπτυχθεί ακόμα.

1.7.2 Επιλογή φάσματος

Μετά το χαρακτηρισμό όλων των διαθέσιμων καναλιών μέσω της διαδικασίας ανάλυσης φάσματος, το κατάλληλο κανάλι για την επικοινωνία των δευτερευόντων χρηστών επιλέγεται λαμβάνοντας υπόψη τις απαιτήσεις ποιότητας υπηρεσίας (Quality of Service, QoS) και τα χαρακτηριστικά του καναλιού, με βάση συγκεκριμένο κανόνα απόφασης (decision rule).

1.8 Κινητικότητα στο φάσμα (Spectrum Mobility)

Ως κινητικότητα στο φάσμα ορίζεται η διαδικασία αλλαγής της συχνότητας λειτουργίας ενός γνωστικού πομποδέκτη. Η μετακίνηση στο φάσμα ενεργοποιείται όταν οι συνθήκες στο τρέχον κανάλι χειροτερεύουν ή εμφανίζεται ένας πρωτεύων χρήστης. Τότε, πραγματοποιείται μεταπομπή φάσματος (spectrum handoff). Τα πρωτόκολλα στα διάφορα στρώματα πρέπει να προσαρμοστούν στις παραμέτρους του νέου καναλιού. Η μεταπομπή αυτή πρέπει να πραγματοποιείται με ομαλό τρόπο ώστε ο χρήστης ενός γνωστικού δικτύου και οι εφαρμογές που λειτουργούν σε αυτό να αντιληφθούν ελάχιστη υποβάθμιση της συνολικής επίδοσης του δικτύου. Είναι αναγκαίο όλα τα πρωτόκολλα κινητικότητας φάσματος να ενημερωθούν, από τον αλγόριθμο ανίχνευσης φάσματος, εγκαίρως πριν από τη μεταπομπή φάσματος. Για το λόγο αυτό χρησιμοποιούνται πολυστρωματικά πρωτόκολλα διαχείρισης κινητικότητας, που υποστηρίζουν κινητικότητα προσαρμοζόμενη προς τις εφαρμογές ή τις συνθήκες δικτύου.

1.9 Διαμοιρασμός φάσματος (Spectrum sharing)

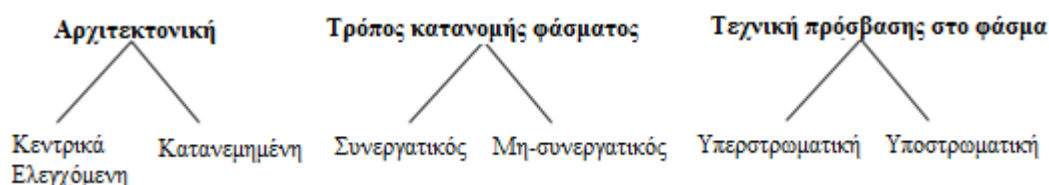
Μια από τις κύριες προκλήσεις στα γνωστικά δίκτυα είναι ο διαμοιρασμός του φάσματος. Το πρόβλημα του διαμοιρασμού του φάσματος είναι παρόμοιο με τα γενικά προβλήματα πρόσβασης στο μέσο (medium access control) σε ήδη υπάρχοντα συστήματα. Όμως, η συνύπαρξη με πρωτεύοντες χρήστες και το μεγάλο εύρος του διαθέσιμου φάσματος δημιουργούν στους δευτερεύοντες χρήστες ιδιαίτερες προκλήσεις με στόχο την ικανοποίηση των αναγκών τους χωρίς να επηρεάζουν την επικοινωνία των πρωτευόντων.

Η διαδικασία διαμοιρασμού φάσματος περιλαμβάνει πέντε βήματα

- *Ανίχνευση Φάσματος* – Ένας γνωστικός πομποδέκτης μπορεί να αξιοποιεί τμήμα του φάσματος μόνο εάν αυτό δεν χρησιμοποιείται από πρωτεύοντα χρήστη ή άλλο χρήστη γνωστικών δικτύων. Επομένως, ένας γνωστικός πομποδέκτης οφείλει να γνωρίζει τη χρήση φάσματος στην περιοχή του.
- *Εκχώρηση φάσματος* – Ανάλογα με τη διαθεσιμότητα του φάσματος, ένας γνωστικός πομποδέκτης μπορεί να προσπελάσει ένα κανάλι. Η προσπέλαση αυτή δεν εξαρτάται μόνο από τη διαθεσιμότητα του καναλιού, αλλά οφείλει να συμμορφώνεται με διάφορες εσωτερικές (και πιθανώς εξωτερικές) πολιτικές.

- *Πρόσβαση στο φάσμα* – Η πρόσβαση στο φάσμα πρέπει να είναι συντονισμένη ώστε να αποφεύγεται ο συνωστισμός και η σύγκρουση πολλών χρηστών σε επικαλυπτόμενα μέρη του φάσματος.
- *Χειραψία πομπού-δέκτη* – Μόλις αποφασιστεί η χρήση ενός καναλιού για επικοινωνία, ο δέκτης πρέπει να ενημερωθεί για το κανάλι αυτό. Έτσι είναι απαραίτητη η ύπαρξη ενός πρωτοκόλλου χειραψίας πομπού-δέκτη.
- *Κινητικότητα στο φάσμα* – Καθώς οι χρήστες γνωστικών δικτύων θεωρούνται «επισκέπτες» στις συχνότητες που καταλαμβάνουν, επιβάλλεται η άμεση αποχώρησή τους, μόλις οι συχνότητες αυτές πρέπει να χρησιμοποιηθούν από πρωτεύοντα χρήστη. Τότε η επικοινωνία των δευτερευόντων χρηστών πρέπει να μετατεθεί σε άλλο κανάλι.

Οι ήδη υπάρχουσες τεχνικές διαμοιρασμού φάσματος (spectrum sharing) μπορούν να ταξινομηθούν σύμφωνα με τρία κριτήρια: την αρχιτεκτονική τους, το τρόπο κατανομής του φάσματος και την τεχνική πρόσβασης στο φάσμα.



Σχήμα 1-11 - Ταξινόμηση τεχνικών διαμοιρασμού φάσματος.

Ως προς την αρχιτεκτονική, ο διαμοιρασμός φάσματος μπορεί να ταξινομηθεί σε κεντρικά ελεγχόμενο και κατανεμημένο. Στον κεντρικά ελεγχόμενο διαμοιρασμό φάσματος, μια κεντρική οντότητα ελέγχει την εκχώρηση και τις διαδικασίες πρόσβασης στο φάσμα, και καταstrώνει ένα χάρτη εκχώρησης φάσματος. Στον κατανεμημένο διαμοιρασμό φάσματος, κάθε κόμβος είναι αποκλειστικός υπεύθυνος για την εκχώρηση φάσματος και η πρόσβαση σε αυτό εξαρτάται από τοπικές (ή και πιθανώς δικτυακές) πολιτικές. Τεχνικές κατανεμημένου διαμοιρασμού προτείνονται συνήθως σε περιπτώσεις όπου η κατασκευή δικτύου κορμού δεν είναι εφικτή ή επιθυμητή.

Ως προς τον τρόπο κατανομής του φάσματος, ο διαμοιρασμός φάσματος μπορεί να ταξινομηθεί σε συνεργατικό ή μη-συνεργατικό. Στις συνεργατικές τεχνικές κάθε κόμβος λαμβάνει υπόψη του την επίδραση της επικοινωνίας του στους άλλους κόμβους, δηλαδή οι μετρήσεις παρεμβολών κάθε κόμβου γνωστοποιούνται στους υπόλοιπους και οι αλγόριθμοι εκχώρησης φάσματος χρησιμοποιούν τις σχετικές πληροφορίες. Οι κεντρικά ελεγχόμενες τεχνικές μπορούν να θεωρηθούν ως συνεργατικές, αλλά υπάρχουν και κατανεμημένες συνεργατικές τεχνικές. Σε αντίθεση με τις συνεργατικές, οι μη-συνεργατικές τεχνικές λαμβάνουν υπόψη μόνο ένα κόμβο. Αναφέρονται, επίσης, και ως εγωιστικές τεχνικές. Παρά τη μειωμένη φασματική χρήση που προσφέρουν οι μη-

συνεργατικές τεχνικές, η ελάχιστη ανάγκη επικοινωνίας μεταξύ των χρηστών αποτελεί πλεονέκτημα, ιδιαίτερα σε πρακτικές εφαρμογές.

Ως προς την τεχνική πρόσβασης στο φάσμα η κατανομή του φάσματος μπορεί να ταξινομηθεί σε υπερστρωματική (overlay) και υποστρωματική (underlay). Στον υπερστρωματικό διαμοιρασμό φάσματος ένας κόμβος χρησιμοποιεί ένα μέρος του φάσματος το οποίο δεν χρησιμοποιείται από πρωτεύοντες χρήστες, επιτυγχάνοντας έτσι τα ελάχιστα επίπεδα παρεμβολής στο πρωτεύον σύστημα. Στον υποστρωματικό διαμοιρασμό φάσματος χρησιμοποιούνται τεχνικές διασποράς φάσματος (spread spectrum), που έχουν αναπτυχθεί για κινητά δίκτυα. Τέτοιες τεχνικές είναι το CDMA και UWB. Κάθε κόμβος μεταδίδει κατά τρόπο τέτοιο ώστε η ισχύς μετάδοσής του να θεωρείται ως θόρυβος για τους πρωτεύοντες χρήστες. Ο υποστρωματικός διαμοιρασμός φάσματος απαιτεί εξελιγμένες τεχνικές διασποράς φάσματος και μπορεί να χρησιμοποιεί μεγαλύτερο εύρος ζώνης σε σύγκριση με τον υπερστρωματικό διαμοιρασμό φάσματος.

1.10 Σύγχρονες τεχνολογίες

Ασύρματα πρότυπα (standards) προσφάτως εκδοθέντα ή υπό ανάπτυξη έχουν αρχίσει να ενσωματώνουν χαρακτηριστικά γνωστικών δικτύων. Τέτοια πρότυπα αποτελούν το IEEE 802.11k, το IEEE 802.22 καθώς και νέες δυνατότητες που έχουν περιληφθεί στο πρότυπο του Bluetooth.

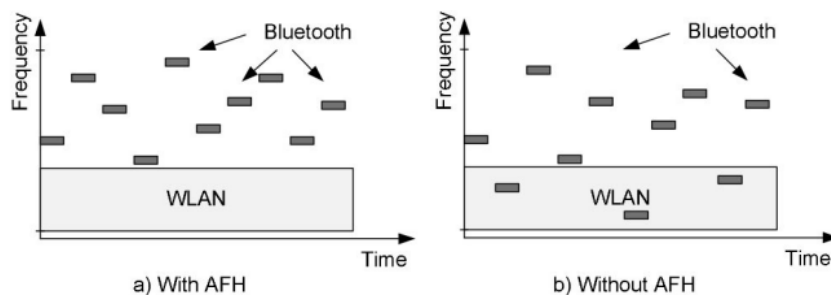
1.10.1 IEEE 802.11k

Το πρότυπο 802.11k αποτελεί επέκταση του προτύπου 802.11 και ορίζει πολλούς τύπους μετρήσεων, όπως αναφορές φόρτου καναλιού (channel load reports), αναφορές ιστογράμματος θορύβου (noise histogram reports) και αναφορές στατιστικών σταθμού επικοινωνίας (station statistic report). Η αναφορά ιστογράμματος θορύβου παρέχει μεθόδους μέτρησης των επιπέδων παρεμβολής που οφείλονται στην ενέργεια που δεν προέρχεται από συστήματα 802.11 στο κανάλι, όπως αυτά λαμβάνονται από τους χρήστες. Αυτά τα δεδομένα χρησιμοποιούνται από το σταθμό βάσης (AP) για να ρυθμίσουν την πρόσβαση των τερματικών χρηστών στο κανάλι. Οι πληροφορίες από την ανίχνευση χρησιμοποιούνται και για να βελτιώσουν την κατανομή κίνησης μέσα στο δίκτυο. Οι συσκευές WLAN συνήθως συνδέονται στο σταθμό βάσης ο οποίος έχει το δυνατότερο σήμα. Πολλές φορές αυτό μπορεί να μην είναι βέλτιστο διότι μπορεί να προκαλέσει την συμφόρηση ενός σταθμού βάσης και την υπο-χρησιμοποίηση άλλων. Στο 802.11k, όταν ο σταθμός βάσης με το ισχυρότερο σήμα προσεγγίζει τα όρια της χωρητικότητας του, οι νέοι χρήστες ανατίθενται σε κάποιον από τους υπο-χρησιμοποιούμενους σταθμούς βάσης. Καίτοι το σήμα λήψης των τερματικών είναι

ασθενέστερο, η συνολική ρυθμαπόδοση του συστήματος είναι καλύτερη λόγω της αποδοτικότερης χρήσης των πόρων του δικτύου.

1.10.2 Bluetooth

Μια νέα δυνατότητα, που ονομάζεται Προσαρμοζόμενη Μεταπήδηση Συχνότητας (Adaptive Frequency Hopping, AFH) έχει εισαχθεί στο πρότυπο του Bluetooth, προκειμένου να μειωθούν οι παρεμβολές μεταξύ ασύρματων τεχνολογιών που μοιράζονται το μη-αδειοδοτημένο φάσμα των 2.4GHz. Στη περιοχή συχνοτήτων αυτή λειτουργούν συσκευές IEEE 802.11b/g, ασύρματα τηλέφωνα, φούρνοι μικροκυμάτων και συσκευές Bluetooth. Ο μηχανισμός της προσαρμοζόμενης μεταπήδησης συχνότητας εντοπίζει τις συχνότητες εκπομπών στις ελεύθερες ζώνες ISM (Industrial Scientific and Medical) και αποφεύγει τη χρησιμοποίησή τους. Κατά αυτό τον τρόπο αποφεύγεται η παρεμβολή στενής ζώνης σε ταυτόχρονες εκπομπές στη ζώνη ISM(π.χ. Wi-Fi), όπως φαίνεται στο σχήμα 1-12, επιτυγχάνονται μικρότερα ποσοστά λανθασμένων bit (Bit error rate, BER) και μειώνεται η απαιτούμενη ισχύς μετάδοσης, για όλα τα εμπλεκόμενα μέρη (Bluetooth και WLAN, στο παράδειγμα του σχήματος 1-12).



Σχήμα 1-12 - Το φάσμα που καταλαμβάνει το Bluetooth με και χωρίς Προσαρμοζόμενη Μεταπήδηση Συχνότητας.

Η προσαρμοζόμενη μεταπήδηση συχνότητας απαιτεί ένα αλγόριθμο ανίχνευσης φάσματος ώστε να εντοπίζονται άλλες συσκευές που χρησιμοποιούν τις συχνότητες ISM και να προσδιορίζεται αν υπάρχει ανάγκη αποφυγής συγκεκριμένων συχνοτήτων. Ο αλγόριθμος ανίχνευσης φάσματος βασίζεται σε παραμέτρους όπως το ποσοστό λανθασμένων πακέτων (Packet error rate) ή και το ποσοστό λανθασμένων bits (Bit error rate), ο δείκτης ισχύος σήματος λήψης (Received signal strength indicator, RSSI), ο λόγος φέροντος προς παρεμβολή και θόρυβο (Carrier-to-interference-plus-noise ratio, CINR) ή άλλα μέτρα επίδοσης που χρησιμοποιούνται για να χαρακτηριστεί ένα κανάλι ως «καλό», «κακό» ή «άγνωστης κατάστασης».

1.10.3 IEEE 802.22

Το πρότυπο 802.22, γνωστό ως το πρότυπο γνωστικών ραδιοεπικοινωνιών λόγω των γνωστικών χαρακτηριστικών που ενσωματώνει, βρίσκεται ακόμα στο στάδιο της ανάπτυξης. Ένα από τα σημαντικότερα χαρακτηριστικά του 802.22 είναι η απαίτηση για ανίχνευση φάσματος. Οι ασύρματες συσκευές περιφερειακής λειτουργίας (Wireless regional area network, WRAN) που βασίζονται στο 802.22 ανιχνεύουν τα κανάλια

τηλεόρασης και βρίσκουν ευνοϊκές διάρκειες μετάδοσης σε αυτά. Οι λειτουργικές απαιτήσεις του προτύπου απαιτούν τουλάχιστον 90% πιθανότητα ορθής ανίχνευσης και το πολύ 10% πιθανότητα λανθασμένης ανίχνευσης για σήματα τηλεόρασης που λαμβάνονται σε δέκτες 802.22 με ισχύ -116dBm ή υψηλότερη.

Η ανίχνευση σχεδιάζεται να γίνεται σε δύο στάδια: ταχεία ανίχνευση και λεπτομερής ανίχνευση (fine sensing). Στο στάδιο ταχείας ανίχνευσης χρησιμοποιείται ένας χονδροειδής αλγόριθμος, π.χ. ανιχνευτής ενέργειας. Το στάδιο λεπτομερούς ανίχνευσης ενεργοποιείται βάσει των αποτελεσμάτων της ταχείας. Η λεπτομερής ανίχνευση περιλαμβάνει ισχυρότερες μεθόδους όπως ανίχνευση κυκλοστατικότητας, ανίχνευση κυματομορφής και ανίχνευση προσαρμοσμένου φίλτρου. Ένας σταθμός βάσης μπορεί να καταναίμει την ανίχνευση σε πολλούς χρήστες. Τα αποτελέσματα επιστρέφουν στο σταθμό βάσης που τα επεξεργάζεται για να διαχειριστεί τις μεταδόσεις.

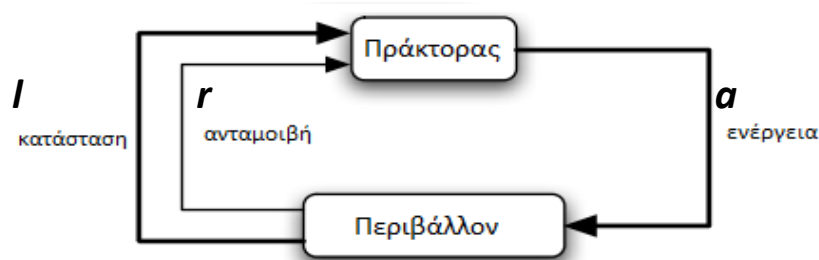
Μια άλλη προσέγγιση διαχείρισης φάσματος βασίζεται σε συστήματα GPS, όπου οι σταθμοί βάσης θα μπορούν να αναφέρουν τη θέση τους και να λαμβάνουν πληροφορίες για τα διαθέσιμα κανάλια τηλεόρασης ανάλογα με αυτή, από ένα κεντρικό εξυπηρετητή. Για συσκευές χαμηλής ισχύος, όπως ασύρματα μικρόφωνα και κάμερες, που λειτουργούν στις συχνότητες μετάδοσης της τηλεόρασης προτείνεται η εξωτερική ανίχνευση. Οι συσκευές αυτές μεταδίδουν περιοδικά σήματα-φάρους, τα οποία έχουν υψηλότερη ισχύ μετάδοσης από το κανονικό. Αυτά τα σήματα-φάροι λαμβάνονται από τις συνήθεις συσκευές 802.22, που αντιλαμβάνονται την παρουσία των συσκευών χαμηλής ισχύος, οι οποίες αλλιώς θα ήταν δύσκολο να ανιχνευτούν λόγω της χαμηλής ισχύος εκπομπής τους.

2 Κεφάλαιο 2 – Ενισχυτική Μάθηση

Η Ενισχυτική Μάθηση (Reinforcement Learning) εμφανίστηκε στα του 20^{ου} αιώνα μαζί με την Κυβερνητική (Cybernetics – Η επιστημονική μελέτη ελέγχου και επικοινωνίας σε βιολογικά και μηχανικά συστήματα) ως απόρροια ερευνών σε στατιστική, ψυχολογία, επιστήμη των υπολογιστών καθώς και στην νευροεπιστήμη. Τα τελευταία χρόνια έχει προσελκύσει ιδιαίτερα το ενδιαφέρον των κοινοτήτων μηχανικής μάθησης και τεχνητής νοημοσύνης. Η κεντρική ιδέα της ενισχυτικής μάθησης είναι ο προγραμματισμός πρακτόρων (agents) μέσω ανταμοιβής ή τιμωρίας χωρίς να χρειάζεται να προσδιοριστεί ο τρόπος με τον οποίο πρέπει να επιτευχθεί ο στόχος. Η ενισχυτική μάθηση ασχολείται με τα προβλήματα που εμφανίζονται όταν ένας πράκτορας πρέπει να εκπαιδευτεί σε συγκεκριμένη συμπεριφορά μέσω δοκιμών σε κάποιο δυναμικό περιβάλλον. Υπάρχουν δύο βασικές στρατηγικές για την επίλυση προβλημάτων ενισχυτικής μάθησης. Η πρώτη είναι η αναζήτηση στο πεδίο των δυνατών συμπεριφορών έτσι ώστε να βρεθεί μία η οποία να λειτουργεί αποδοτικά στο περιβάλλον. Η στρατηγική αυτή βρίσκει εφαρμογές στους χώρους των γενετικών αλγορίθμων και του γενετικού προγραμματισμού. Η δεύτερη στρατηγική είναι η χρησιμοποίηση στατιστικών τεχνικών και μεθόδων δυναμικού προγραμματισμού για την εκτίμηση των δράσεων στις καταστάσεις του περιβάλλοντος.

2.1 Το μοντέλο της ενισχυτικής μάθησης

Στο κλασικό μοντέλο της ενισχυτικής μάθησης, ένας πράκτορας (agent) αλληλεπιδρά με το περιβάλλον κάνοντας παρατηρήσεις και προβαίνοντας σε ενέργειες (actions).



Σχήμα 2-1 - Το κλασικό μοντέλο της ενισχυτικής μάθησης.

Σε κάθε βήμα της αλληλεπίδρασης, ο πράκτορας λαμβάνει μια είσοδο I , η οποία εξαρτάται από την παρούσα κατάσταση του περιβάλλοντος s . Έπειτα ο πράκτορας επιλέγει μια ενέργεια a , ως έξοδό του. Αυτή η δράση αλλάζει την κατάσταση του περιβάλλοντος και αυτή η αλλαγή γνωστοποιείται στον πράκτορα μέσω ενός βαθμωτού σήματος ενίσχυσης (ή ανταμοιβής) r (reinforcement signal). Η συμπεριφορά του πράκτορα εστιάζεται στην επιλογή δράσεων οι οποίες αποσκοπούν στην αύξηση του μακροπρόθεσμου αθροίσματος των τιμών του σήματος ενίσχυσης. Ο πράκτορας μπορεί

να εκπαιδευτεί σε αυτή τη συμπεριφορά έπειτα από συστηματικές δοκιμές, καθοδηγούμενες από μια μεγάλη ποικιλία αλγορίθμων.

Το τυπικό μοντέλο αποτελείται από:

- Ένα διακριτό σύνολο καταστάσεων του περιβάλλοντος, S
- Ένα διακριτό σύνολο ενεργειών του πράκτορα A και
- Ένα σύνολο βαθμωτών σημάτων ενίσχυσης – συνήθως $\{0,1\}$ ή το σύνολο των πραγματικών αριθμών.

Ένας διαισθητικός τρόπος για να κατανοηθεί η σχέση μεταξύ περιβάλλοντος και πράκτορα είναι ο ακόλουθος διάλογος:

Περιβάλλον: Είσαι στην κατάσταση 65. Έχεις 4 πιθανές ενέργειες.

Πράκτορας: Διαλέγω την ενέργεια 2.

Περιβάλλον: Λαμβάνεις ενίσχυση 7 μονάδων. Τώρα είσαι στην κατάσταση 15. Έχεις 2 πιθανές ενέργειες.

Πράκτορας: Διαλέγω τη ενέργεια 1.

Περιβάλλον: Λαμβάνεις ενίσχυση -4 μονάδων. Τώρα είσαι στην κατάσταση 65. Έχεις 4 πιθανές ενέργειες.

Πράκτορας: Διαλέγω τη ενέργεια 2.

Περιβάλλον: Λαμβάνεις ενίσχυση 5 μονάδων. Τώρα είσαι στην κατάσταση 44. Έχεις 5 πιθανές ενέργειες.

...

Ο ρόλος του πράκτορα είναι να προσδιορίσει μια πολιτική π , επιλέγοντας από κάθε κατάσταση ενέργειες που τον οδηγούν σε καταστάσεις τέτοιες ώστε να μεγιστοποιείται κάποιο μακροπρόθεσμο μέτρο ενίσχυσης. Γενικά, αναμένεται το περιβάλλον να είναι μη-αιτιοκρατικό, δηλαδή η εκτέλεση της ίδιας ενέργειας από την ίδια κατάσταση αλλά σε διαφορετικές περιπτώσεις μπορεί να οδηγήσει σε διαφορετικές καταστάσεις και/ή διαφορετικές τιμές ενίσχυσης. Αυτό συμβαίνει στο προηγούμενο παράδειγμα: η ενέργεια 2 από την κατάσταση 65 οδηγεί σε διαφορετική κατάσταση και διαφορετική ενίσχυση σε δύο διαφορετικές περιπτώσεις. Εντούτοις, το περιβάλλον θεωρείται στατικό, δηλαδή οι πιθανότητες αλλαγής κατάστασης ή λήψης συγκεκριμένων σημάτων ενίσχυσης δεν μεταβάλλονται με το χρόνο.

Η ενισχυτική μάθηση διαφέρει από την εποπτευόμενη μάθηση (supervised learning) με πολλούς τρόπους. Η κυριότερη διαφορά είναι η απουσία συγκεκριμένων ζευγών εισόδου/εξόδου. Αντιθέτως, έπειτα από τη λήψη απόφασης για κάποια ενέργεια ο πράκτορας λαμβάνει την άμεση ανταμοιβή και πληροφορείται την επόμενη κατάσταση, αλλά δεν γνωρίζει ποιά ενέργεια θα ήταν η βέλτιστη μακροπρόθεσμα. Είναι αναγκαίο για

τον πράκτορα να αποκτήσει εμπειρία για τις καταστάσεις του περιβάλλοντος, τις ενέργειες, τις μεταβάσεις και τις ανταμοιβές και να δράσει βέλτιστα. Μια ακόμα διαφορά από την εποπτευόμενη μάθηση είναι ότι οι επιδόσεις σε πραγματικό χρόνο έχουν μεγάλη σημασία, καθώς η εκτίμηση του συστήματος είναι ταυτόχρονη με τη μάθηση.

Η ενισχυτική μάθηση μπορεί να αφορά ένα πράκτορα (Single-Agent reinforcement learning, SARL), όπως περιγράφεται ανωτέρω, ή πολλούς. Στην ενισχυτική μάθηση πολλαπλών πρακτόρων (Multi-Agent reinforcement learning, MARL), υπάρχουν πολλοί πράκτορες οι οποίοι ενεργούν στο περιβάλλον. Κάθε πράκτορας επηρεάζει το περιβάλλον οπότε οι ενέργειες ενός πράκτορα μπορεί να έχουν διαφορετικά αποτελέσματα επηρεαζόμενα και από τις ενέργειες των υπόλοιπων πρακτόρων. Αυτή η διαφορά είναι η αιτία προβλημάτων στην εφαρμογή τεχνικών ενισχυτικής μάθησης σε περιβάλλον με πολλαπλούς πράκτορας, καθώς οι τεχνικές αυτές είναι σχεδιασμένες για στατικά περιβάλλοντα και από την οπτική γωνία κάθε πράκτορα το περιβάλλον δεν είναι στατικό. Για να περιγραφούν (μαθηματικά) τέτοια περιβάλλοντα, χρησιμοποιούνται μοντέλα από τη θεωρία παιγνίων, με στόχο την επίλυση προβλημάτων πολλαπλών πρακτόρων στα οποία οι λύσεις περιέχουν συμβιβασμούς και συνεργασίες. Συγκεκριμένα τα περισσότερο γνωστά μοντέλα είναι αυτά των στοχαστικών παιγνίων (stochastic games) ή η υποκατηγορία των παιγνίων μητρών (matrix games).

2.2 Μοντέλα Βέλτιστης Συμπεριφοράς

Πριν την επιλογή αλγορίθμου μάθησης πρέπει να αποφασιστεί ποιο μοντέλο βέλτιστης συμπεριφοράς θα χρησιμοποιηθεί, δηλαδή πρέπει να αποφασιστεί πώς θα λαμβάνει ο πράκτορας υπόψη του τις μελλοντικές επιπτώσεις των αποφάσεων του με βάση την παρούσα συμπεριφορά του. Τα κυρίαρχα μοντέλα είναι τρία.

2.2.1 Το μοντέλο πεπερασμένου ορίζοντα (finite-horizon model).

Σε κάθε χρονική στιγμή ο πράκτορας πρέπει να μεγιστοποιεί την αναμενόμενη ανταμοιβή του για τα επόμενα h βήματα:

$$E \left(\sum_{t=0}^h r_t \right) \quad (2.1)$$

όπου r_t δηλώνει τη βαθμωτή ανταμοιβή που λαμβάνεται μετά από t βήματα στο μέλλον. Το μοντέλο αυτό μπορεί να χρησιμοποιηθεί με δύο τρόπους. Σύμφωνα με τον πρώτο τρόπο ο πράκτορας έχει μια μη-στατική πολιτική. Στο πρώτο βήμα επιλέγει μια βέλτιστη ενέργεια h -βημάτων, που ορίζεται ως η καλύτερη ενέργεια που είναι δυνατή δεδομένου ότι έχει ακόμα h βήματα για να προβεί σε ενέργειες και να λάβει ενισχύσεις. Στο επόμενο βήμα πραγματοποιεί μια βέλτιστη ενέργεια $(h-1)$ βημάτων κ.ο.κ. μέχρι να πραγματοποιήσει μια βέλτιστη ενέργεια 1 βήματος και να τερματίσει. Σύμφωνα με το δεύτερο τρόπο ο πράκτορας πραγματοποιεί έλεγχο απομακρυσμένου-ορίζοντα (receding-horizon control) όπου πραγματοποιεί πάντα τη βέλτιστη ενέργεια h -βημάτων. Ο

πράκτορας συμπεριφέρεται πάντα με βάση την ίδια πολιτική αλλά η τιμή του h περιορίζει το μελλοντικό βάθος που θα εξετάζεται για τη λήψη των αποφάσεων του. Το μοντέλο πεπερασμένου ορίζοντα δεν είναι πάντοτε ιδανικό καθώς μπορεί να μην είναι εξαρχής γνωστός ο χρόνος ζωής (δηλαδή το συνολικό χρονικό διάστημα κατά το οποίο θα εκτελείται ένας αλγόριθμος μάθησης) του πράκτορα.

2.2.2 Το απομειούμενο μοντέλο απείρου ορίζοντα (infinite-horizon discounted model).

Σε αυτό το μοντέλο ο πράκτορας λαμβάνει υπόψη το μακροπρόθεσμο όφελος, αλλά οι μελλοντικές ανταμοιβές υπολογίζονται με γεωμετρικά φθίνουσα βαρύτητα με βάση ένα παράγοντα απομείωσης γ (discount factor), $0 \leq \gamma < 1$. Η αναμενόμενη ανταμοιβή τότε γίνεται

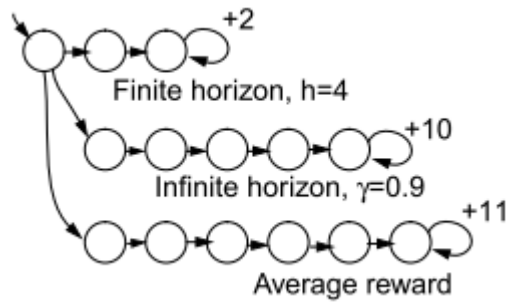
$$E \left(\sum_{t=0}^{\infty} \gamma^t r_t \right) \quad (2.2)$$

2.2.3 Το μοντέλο μέσης ανταμοιβής (average-reward model)

Σύμφωνα με το μοντέλο αυτό ο πράκτορας λαμβάνει αποφάσεις οι οποίες βελτιστοποιούν τη μακροπρόθεσμη μέση ανταμοιβή του. Η ανταμοιβή ισούται με το όριο

$$\lim_{h \rightarrow \infty} E \left(\frac{1}{h} \sum_{t=0}^h r_t \right) \quad (2.3)$$

Μια τέτοια πολιτική αναφέρεται ως πολιτική βέλτιστου κέρδους και μπορεί να θεωρηθεί ως το όριο του απομειούμενου μοντέλου απείρου ορίζοντα όταν ο παράγοντας απομείωσης τείνει προς τη μονάδα. Ένα πρόβλημα με αυτό το μοντέλο είναι ότι δυο πολιτικές, στη μια εκ των οποίων λαμβάνονται υψηλές ανταμοιβές στα αρχικά στάδια ενώ στην άλλη δεν συμβαίνει αυτό, δεν διαφοροποιούνται, αφού οι ενδεχόμενες υψηλές ανταμοιβές που λαμβάνονται στα αρχικά στάδια της ζωής του πράκτορα «καλύπτονται» από τη μακροπρόθεσμη μέση απόδοση. Είναι δυνατή μια γενίκευση αυτού του μοντέλου ώστε να λαμβάνονται υπόψη τόσο η μακροπρόθεσμη μέση ανταμοιβή όσο και οι αρχικές ανταμοιβές. Στο γενικευμένο μοντέλο βέλτιστης πόλωσης (bias optimal model) προτιμάται η πολιτική η οποία μεγιστοποιεί τη μακροπρόθεσμη ανταμοιβή και σε περιπτώσεις όπου δύο ή περισσότερες πολιτικές αποφέρουν την ίδια μακροπρόθεσμη ανταμοιβή προτιμάται αυτή η οποία έχει υψηλές αρχικές ανταμοιβές.



Σχήμα 2-2 - Σύγκριση μοντέλων βέλτιστης συμπεριφοράς.

2.3 Μέτρηση της ποιότητας της μάθησης

Ενώ τα προηγούμενα μοντέλα χρησιμοποιούνται για να εκτιμήσουν τις επιδόσεις των πολιτικών στις οποίες καταλήγει ένας αλγόριθμος σύμφωνα με τη διαδικασία της μάθησης, υπάρχει ανάγκη να εκτιμηθεί και η ποιότητα της μάθησης αυτής καθ' εαυτής. Για το λόγο αυτό χρησιμοποιούνται τα ακόλουθα κριτήρια.

2.3.1 Τελική σύγκλιση στο βέλτιστο. (Eventual Convergence to optimality)

Πολλοί αλγόριθμοι εγγυώνται ασυμπτωτική σύγκλιση στη βέλτιστη συμπεριφορά (Watkins & Dayan, 1992 [29]). Το κριτήριο αυτό δεν είναι και πολύ χρήσιμο πρακτικά, καθώς ένας πράκτορας ο οποίος μπορεί να φθάσει γρήγορα στο 99% του βέλτιστου είναι προτιμότερος από ένα πράκτορα ο οποίος εγγυάται τη βέλτιστη λύση αλλά με αργό ρυθμό.

2.3.2 Ταχύτητα σύγκλισης στο βέλτιστο. (Speed of convergence to optimality)

Καθώς η βέλτιστη λύση είναι συνήθως ένα ασυμπτωτικό αποτέλεσμα, ένα περισσότερο πρακτικό κριτήριο είναι η ταχύτητα σύγκλισης κοντά στο βέλτιστο, όπου βέβαια πρέπει να οριστεί ποια προσέγγιση του βέλτιστου θεωρείται ικανοποιητική. Ένα άλλο σχετικό μέτρο είναι το επίπεδο της απόδοσης μετά από συγκεκριμένο χρόνο και στην περίπτωση αυτή χρειάζεται να οριστεί ποιός χρόνος θεωρείται κατάλληλος.

Ένα μειονέκτημα των κριτηρίων αυτών είναι ότι, αν ένας αλγόριθμος προσπαθεί απλώς να φθάσει στη βέλτιστη λύση όσο ταχύτερα μπορεί, είναι πιθανό να λάβει πολύ υψηλές τιμωρίες (penalties) στη φάση της μάθησης. Μια λιγότερο επιθετική στρατηγική, όπου ο πράκτορας χρειάζεται περισσότερο χρόνο για να φθάσει στη βέλτιστη λύση αλλά ενδιάμεσως λαμβάνει μεγαλύτερες ανταμοιβές, πιθανώς να είναι προτιμότερη.

2.3.3 Μεταμέλεια - Regret

Ένα καταλληλότερο κριτήριο είναι η αναμενόμενη μείωση στις ανταμοιβές λόγω της εφαρμογής του αλγορίθμου μάθησης αντί της βέλτιστης συμπεριφοράς από την αρχή.

Αυτό το κριτήριο είναι γνωστό ως μεταμέλεια (regret). Όποια λάθη γίνονται κατά τη διάρκεια της μάθησης «τιμωρούνται». Δυστυχώς, ο ποσοτικός προσδιορισμός αυτού του κριτηρίου είναι δύσκολος.

2.3.4 Ο συμβιβασμός μεταξύ εξερεύνησης και εκμετάλλευσης

Μια κύρια διαφορά μεταξύ ενισχυτικής μάθησης και εποπτευόμενης μάθησης είναι το γεγονός ότι ο πράκτορας ενισχυτικής μάθησης πρέπει να εξερευνήσει (exploration) το περιβάλλον του. Το πρόβλημα που μπορεί να προκύψει γίνεται φανερό από το επόμενο παράδειγμα.

Το πρόβλημα των k κουλοχέρηδων.

Ένας πράκτορας βρίσκεται σε ένα δωμάτιο με ένα σύνολο από k κουλοχέρηδες. Ο πράκτορας επιτρέπεται να παίξει ένα πεπερασμένο αριθμό φορών, h . Ο πράκτορας μπορεί να επιλέξει οποιοδήποτε κουλοχέρη σε κάθε γύρο, και το μόνο κόστος είναι η σπατάλη ενός παιχνιδιού όταν επιλέγει μη-βέλτιστο κουλοχέρη. Όταν ο πράκτορας παίζει στον κουλοχέρη i , λαμβάνει ανταμοιβή 1 ή 0, σύμφωνα με μια πιθανότητα p_i . Όταν οι ανταμοιβές είναι ανεξάρτητα γεγονότα και οι πιθανότητες p_i είναι άγνωστες, ποια πρέπει να είναι η στρατηγική του πράκτορα;

Το πρόβλημα αυτό παρουσιάζει τη βασική συμβιβαστική λύση μεταξύ εξερεύνησης και εκμετάλλευσης. Ο πράκτορας μπορεί να πιστεύει ότι συγκεκριμένος κουλοχέρης έχει σχετικά μεγάλη πιθανότητα ανταμοιβής, πρέπει όμως να διαλέγει το συγκεκριμένο κουλοχέρη συνέχεια ή να διαλέξει ένα άλλο για τον οποίο διαθέτει λιγότερες πληροφορίες αλλά σύμφωνα με αυτές τις (ελλιπείς πιθανότατα) πληροφορίες η πιθανότητα ανταμοιβής του είναι μικρότερη; Οι απαντήσεις σε αυτά τα ερωτήματα εξαρτώνται από τη διάρκεια του παιχνιδιού. Όσο μεγαλύτερη η διάρκεια τόσο μεγαλύτερη η πιθανότητα σύγκλισης σε ένα μη βέλτιστο κουλοχέρη και τόσο περισσότερο πρέπει ο πράκτορας να εξερευνήσει.

Το πρόβλημα των k κουλοχέρηδων αποτελεί ένα περιβάλλον ενισχυτικής μάθησης στο οποίο έχουμε μία κατάσταση με μεταβάσεις μόνο στον εαυτό της. Για αυτό το πρόβλημα της μιας κατάστασης υπάρχουν 2 είδη λύσεων. Το πρώτο είδος, οι αποδεδειγμένα ορθές τεχνικές (formally justified techniques), αποτελούν λύσεις οι οποίες έχουν αποδεδειγμένα ορθά αποτελέσματα αλλά δεν κλιμακώνονται εύκολα, έτσι ώστε να χρησιμοποιηθούν σε περισσότερο πολύπλοκα προβλήματα. Τέτοιες τεχνικές είναι ο δυναμικός προγραμματισμός, οι δείκτες κατανομής του Gittins (Gittins Allocation Indices) καθώς και τα Αυτόματα μάθησης (Learning automata). Το δεύτερο είδος, οι Ad-Hoc τεχνικές, αποτελούν τεχνικές οι οποίες, καίτοι σπάνια αποτελούν την καλύτερη επιλογή για τα μοντέλα βέλτιστης συμπεριφοράς, μπορούν να θεωρηθούν ως λογικοί και απλοί (τόσο λογικά όσο και υπολογιστικά) ευριστικοί μηχανισμοί (heuristics). Παραδείγματα τέτοιων τεχνικών είναι οι άπληστες στρατηγικές (greedy strategies), οι τυχαίες στρατηγικές

(Randomized Strategies) και οι τεχνικές που είναι βασισμένες σε διαστήματα εμπιστοσύνης (interval-based techniques).

2.3.5 Καθυστερημένη Ανταμοιβή

Γενικά, στα προβλήματα ενισχυτικής μάθησης οι ενέργειες του πράκτορα επηρεάζουν όχι μόνο την άμεση ανταμοιβή του αλλά και την επόμενη κατάσταση του περιβάλλοντος. Τέτοια περιβάλλοντα μπορούν να θεωρηθούν ως δίκτυα προβλημάτων κουλοχέρηδων, αλλά ο πράκτορας πρέπει να λαμβάνει υπόψη τόσο την άμεση ανταμοιβή όσο και την επόμενη κατάσταση όταν αποφασίζει για την ενέργεια του. Το μοντέλο μακροπρόθεσμης βέλτιστης συμπεριφοράς που χρησιμοποιεί ο πράκτορας καθορίζει πώς και κατά πόσο λαμβάνει υπόψη του το μέλλον. Ο πράκτορας πρέπει να είναι ικανός να μαθαίνει από την καθυστερημένη ανταμοιβή: μπορεί να χρειαστεί μεγάλος αριθμός ενεργειών οι οποίες αποδίδουν μικρή ανταμοιβή μέχρι να φθάσει ο πράκτορας σε μια κατάσταση που να αποδίδει υψηλή ανταμοιβή. Επίσης, ο πράκτορας πρέπει να είναι ικανός να μάθει ποιές από τις ενέργειες του είναι επιθυμητές βασιζόμενος στην ανταμοιβή που μπορεί να λάβει στο μακρινό του μέλλον.

2.4 Διαδικασίες αποφάσεων Markov (Markov Decision Processes, MDPs)

Προβλήματα καθυστερημένης ανταμοιβής μπορούν να μοντελοποιηθούν ικανοποιητικά ως διαδικασίες απόφασης Markov (MDP). Τέτοιες διαδικασίες αποτελούνται από:

- Ένα σύνολο καταστάσεων S
- Ένα σύνολο ενεργειών A
- Μια συνάρτηση ανταμοιβής $R: S \times A \rightarrow \mathfrak{R}$
- Μια συνάρτηση μετάβασης κατάστασης $T: S \times A \rightarrow \Pi(S)$, όπου το $\Pi(S)$ περιέχει κατανομές πιθανοτήτων για το σύνολο S (απεικονίζει καταστάσεις σε πιθανότητες). Με $T(s,a,s')$ συμβολίζεται η πιθανότητα μετάβασης από την κατάσταση s στην κατάσταση s' λόγω της ενέργειας a .

Η συνάρτηση μετάβασης κατάστασης καθορίζει την πιθανότητα της επόμενης κατάστασης του περιβάλλοντος του πράκτορα. Η συνάρτηση ανταμοιβής καθορίζει την αναμενόμενη άμεση ανταμοιβή με βάση την τρέχουσα κατάσταση και την ενέργεια. Το μοντέλο είναι Μαρκοβιανό εφόσον οι μεταβάσεις από κατάσταση σε κατάσταση είναι ανεξάρτητες από προηγούμενες καταστάσεις του περιβάλλοντος και προηγούμενες ενέργειες του πράκτορα.

2.5 Εύρεση πολιτικής για δεδομένο Μοντέλο μάθησης

Δεδομένου ενός τυπικού μοντέλου, απαιτούνται τεχνικές για την εύρεση της βέλτιστης πολιτικής που πρέπει να ακολουθήσει ένας πράκτορας, σε MDP περιβάλλοντα. Αυτές

είναι τεχνικές δυναμικού προγραμματισμού και αποτελούν τη βάση για τους αλγόριθμους μάθησης.

Ως βέλτιστη τιμή μιας κατάστασης ορίζεται το αναμενόμενο άπειρο απομειούμενο άθροισμα (infinite discounted sum) ανταμοιβών που λαμβάνει ένας πράκτορας αν εκκινώντας από αυτή την κατάσταση ακολουθεί τη βέλτιστη πολιτική. Η βέλτιστη τιμή αυτή γράφεται

$$V^*(s) = \max_{\pi} E \left(\sum_{t=0}^{\infty} \gamma^t r_t \right) \quad (2.4)$$

όπου με π συμβολίζεται μια συνολική πολιτική αποφάσεων. Η συνάρτηση βέλτιστης τιμής είναι μοναδική και προσδιορίζεται ως η κοινή λύση των εξισώσεων

$$V^*(s) = \max_a \left(R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V^*(s') \right) \forall s \in S \quad (2.5)$$

Η ανωτέρω εξίσωση δηλώνει ότι η τιμή μιας κατάστασης s είναι το άθροισμα της αναμενόμενης άμεσης ανταμοιβής και της αναμενόμενης μειωμένης τιμής της επόμενης κατάστασης, υποθέτοντας ότι έχει επιλεγεί η καλύτερη δυνατή ενέργεια. Δεδομένης της συνάρτησης βέλτιστης τιμής $V^*(s)$, η βέλτιστη πολιτική ορίζεται ως

$$\pi^*(s) = \arg \max_a \left(R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V^*(s') \right) \quad (2.6)$$

2.5.1 Επανάληψη τιμών (Value iteration)

Ένας τρόπος για να προσδιοριστεί η βέλτιστη πολιτική είναι να βρεθεί η συνάρτηση βέλτιστης τιμής. Αυτή με τη σειρά της μπορεί να προσδιοριστεί εφαρμόζοντας τον ακόλουθο απλό επαναληπτικό αλγόριθμο, ο οποίος ονομάζεται επανάληψη τιμών (value iteration), ο οποίος μπορεί να δείχτει ότι συγκλίνει στις σωστές τιμές του V^* .

Δώσε αρχικές τιμές στο $V(s)$

Επανάλαβε μέχρι η πολιτική να γίνει ικανοποιητική (μεγιστοποίηση της 2.6)

Επανάλαβε για $s \in S$

Επανάλαβε για $a \in A$

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

$$V(s) := \max_a Q(s, a)$$

Τέλος επανάληψης

Τέλος επανάληψης

Η συνθήκη τερματισμού του αλγορίθμου δεν είναι προφανής. Ένα αποτελεσματικό κριτήριο τερματισμού δόθηκε από τους Williams & Baird ([30]). Σύμφωνα με αυτό, όταν η μέγιστη διαφορά μεταξύ δύο συνεχόμενων συναρτήσεων τιμής είναι μικρότερη από ϵ , όπου ϵ αρκετά μικρό, η τιμή της άπληστης (greedy) πολιτικής (δηλαδή της πολιτικής με την οποία επιλέγεται σε κάθε κατάσταση η ενέργεια που μεγιστοποιεί την αναμενόμενη ανταμοιβή, χρησιμοποιώντας την παρούσα εκτίμηση της συνάρτησης τιμής) διαφέρει από την συνάρτηση τιμής της βέλτιστης πολιτικής λιγότερο από $2\epsilon\gamma/(1-\gamma)$ σε οποιαδήποτε κατάσταση. Αυτή η άπληστη πολιτική θα γίνει εγγυημένα βέλτιστη σε πεπερασμένο πλήθος βημάτων ακόμα και αν η συνάρτηση τιμής δεν έχει φθάσει σε σύγκλιση (Bertsekas, 1987 [31]). Στην πράξη, η άπληστη πολιτική γίνεται συχνά βέλτιστη πολύ πριν τη σύγκλιση της συνάρτησης τιμής.

Η επανάληψη τιμών είναι πολύ ευέλικτη τεχνική καθώς ο υπολογισμός των τιμών του V δεν απαιτείται να πραγματοποιηθεί με αυστηρή σειρά αλλά μπορεί να γίνει ασύγχρονα και παράλληλα, εφόσον κάθε κατάσταση ενημερώνεται άπειρες φορές σε άπειρες επαναλήψεις.

Ενημερώσεις βασισμένες στην εξίσωση (2.5) είναι γνωστές ως πλήρη αντίγραφα ασφαλείας (full backups) και χρησιμοποιούν πληροφορίες από όλες τις πιθανές διάδοχες καταστάσεις. Μπορεί να δειχτεί ότι μπορούν να χρησιμοποιηθούν ενημερώσεις της μορφής

$$Q(s, a) := Q(s, a) + \tilde{\alpha} \left(r + \gamma \max_a Q(s', a') - Q(s, a) \right) \quad (2.7)$$

εφόσον κάθε ζεύγος των a και s ενημερώνεται απείρως συχνά, το s' είναι δείγμα τυχαίας μεταβλητής με κατανομή $T(s, a, s')$, το r είναι δείγμα τυχαίας μεταβλητής που ακολουθεί την κανονική κατανομή με μέση τιμή $R(s, a)$ και πεπερασμένη διακύμανση και ο ρυθμός μάθησης $\tilde{\alpha}$ (learning rate) μειώνεται σταδιακά.

2.5.2 Επανάληψη πολιτικών (Policy iteration)

Ο αλγόριθμος επανάληψης πολιτικών μεταβάλλει απευθείας την πολιτική ενός πράκτορα, αντί να την προσδιορίζει έμμεσα μέσω της βέλτιστης συνάρτησης τιμής.

Διάλεξε μια τυχαία πολιτική π'

Επανάλαβε

$\pi := \pi'$

Υπολόγισε την συνάρτηση τιμής της π :

Λύσε τις γραμμικές εξισώσεις

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

Βελτίωσε την πολιτική σε κάθε κατάσταση:

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

Μέχρι $\pi = \pi'$

Η συνάρτηση τιμής μιας πολιτικής είναι η αναμενόμενη, συνεχώς απομειούμενη, ανταμοιβή η οποία θα ληφθεί σε κάθε κατάσταση εφόσον ακολουθηθεί η συγκεκριμένη πολιτική και αυτή η συνάρτηση τιμής μπορεί να προσδιοριστεί επιλύοντας ένα σύνολο γραμμικών εξισώσεων. Όταν γίνει γνωστή η τιμή κάθε κατάστασης υπό την τρέχουσα πολιτική, μπορεί να κριθεί αν η τιμή μπορεί να βελτιωθεί διαφοροποιώντας την αρχική ενέργεια του πράκτορα. Αν μπορεί να βελτιωθεί, μεταβάλλεται η πολιτική. Αυτό το βήμα εγγυάται τη βελτίωση της απόδοσης της πολιτικής. Εφόσον δεν υπάρχουν βελτιώσεις, η πολιτική είναι βέλτιστη.

2.6 Μαθαίνοντας μια βέλτιστη πολιτική: Μέθοδοι χωρίς μοντέλα

Στη γενική περίπτωση ένα μοντέλο περιλαμβάνει την πιθανότητα μετάβασης κατάστασης (state transition probability) $T(s,a,s')$ και τη συνάρτηση ενίσχυσης $R(s,a)$. Η ενισχυτική μάθηση ασχολείται κυρίως με την εύρεση της βέλτιστης πολιτικής όταν ένα τέτοιο μοντέλο δεν είναι γνωστό. Ο πράκτορας πρέπει να αλληλεπιδράσει απευθείας με το περιβάλλον του για να αποκτήσει πληροφορίες οι οποίες, με τη χρησιμοποίηση κατάλληλου αλγορίθμου, μπορούν να οδηγήσουν στη βέλτιστη πολιτική.

Το μεγαλύτερο πρόβλημα που καλείται να αντιμετωπίσει ένας πράκτορας ενισχυτικής μάθησης είναι η προσωρινή εκχώρηση τιμής (temporal credit assignment). Πώς είναι δυνατό να γνωρίζει αν η ενέργεια που μόλις εκτέλεσε είναι καλή, όταν αυτή μπορεί να προκαλέσει μακροχρόνιες επιπτώσεις; Μια στρατηγική είναι να περιμένει μέχρι το τέλος και να ανταμείψει αναδρομικά τις ενέργειες που έκανε όταν το αποτέλεσμα είναι ικανοποιητικό ή να τις τιμωρήσει όταν είναι κακό. Σε τρέχουσες εργασίες δεν είναι γνωστό το τέλος, και η γνώση του πιθανώς να χρειάζεται μεγάλο μέγεθος μνήμης. Για το λόγο αυτό χρησιμοποιείται η εμπειρία που αποκτάται από την μέθοδο της επανάληψης τιμών (value iteration) για τη προσαρμογή της εκτιμώμενης τιμής μιας κατάστασης με βάση την άμεση ανταμοιβή και την εκτιμώμενη τιμή της επόμενης κατάστασης. Αυτή η κατηγορία αλγορίθμων είναι γνωστή ως μέθοδοι χρονικών διαφορών (temporal difference methods, Sutton, 1988 [42]), ένας από τους οποίους είναι η Q-μάθηση (Q-learning).

2.6.1 Q-μάθηση (Q-learning)

Ο αλγόριθμος Q-learning πρωτοπαρουσιάστηκε από τον Watkins το 1989[32]. Το βασικό χαρακτηριστικό του είναι η ικανότητα επιλογής μεταξύ άμεσης και καθυστερημένης ανταμοιβής. Για την πληρέστερη κατανόηση του αλγορίθμου χρησιμοποιούνται οι ακόλουθοι συλλογισμοί και συμβολισμοί.

Έστω $Q^*(s,a)$ η αναμενόμενη απομειούμενη ανταμοιβή που θα λάμβανε ο πράκτορας εφόσον εκτελούσε την ενέργεια a στην κατάσταση s , και συνέχιζε διαλέγοντας τις βέλτιστες ενέργειες. $V^*(s)$ είναι η τιμή της s θεωρώντας ότι επιλέγεται αρχικά η καλύτερη ενέργεια, δηλαδή $V^*(s) = \max_a Q^*(s,a)$. Η $Q^*(s,a)$ μπορεί να γραφεί αναδρομικά υπό τη μορφή

$$Q^*(s, a) = R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') \max_{a'} Q^*(s', a') \quad (2.8)$$

όπου $R(s, a)$ η συνάρτηση ανταμοιβής και γ ($0 \leq \gamma \leq 1$) η παράμετρος απομείωσης, που εξασφαλίζει τη σύγκλιση του αθροίσματος. Για $\gamma \approx 0$ ο πράκτορας τείνει να λαμβάνει υπόψη μόνο την άμεση ανταμοιβή. Για $\gamma \approx 1$ ο πράκτορας δίνει μεγαλύτερη βαρύτητα στις μελλοντικές ανταμοιβές.

Εφόσον $V^*(s) = \max_a Q^*(s, a)$ η βέλτιστη πολιτική στην κατάσταση s θα είναι η ενέργεια a που μεγιστοποιεί την $Q^*(s, a)$, δηλαδή $\pi^*(s) = \arg \max_a Q^*(s, a)$.

Επειδή η συνάρτηση Q καθορίζει πλήρως την ενέργεια, οι τιμές Q (Q-values) μπορούν να εκτιμηθούν απευθείας χρησιμοποιώντας μια μέθοδο παρόμοια με την TD(0) του Sutton (Sutton, 1988 [42]) η οποία ακολουθεί τον κανόνα ανανέωσης

$$V(s) := V(s) + \tilde{\alpha}(r + \gamma V(s') - V(s)) \quad (2.9)$$

όπου s' η επόμενη κατάσταση, r η ανταμοιβή και $\tilde{\alpha}$ ο ρυθμός μάθησης.

Επίσης, οι τιμές Q μπορούν να χρησιμοποιηθούν για να καθορίσουν την πολιτική, καθώς μια ενέργεια μπορεί να επιλεγεί απλά διαλέγοντας αυτήν που έχει τη μέγιστη τιμή Q για την τρέχουσα κατάσταση.

Ο κανόνας Q-μάθησης είναι:

$$Q(s, a) := Q(s, a) + \tilde{\alpha}(r + \gamma \max_{a'} Q(s', a') - Q(s, a)) \quad (2.10)$$

όπου $\langle s, a, r, s' \rangle$ είναι μια πλειάδα εμπειρίας (experience tuple) που συμβολίζει μια μετάβαση από μια κατάσταση του περιβάλλοντος στην επόμενη. Ως s ορίζεται η τρέχουσα κατάσταση του πράκτορα, a είναι η ενέργεια, r είναι η άμεση ανταμοιβή που λαμβάνει και s' η κατάσταση στην οποία καταλήγει. Αν κάθε ενέργεια εκτελείται σε κάθε κατάσταση άπειρες φορές σε άπειρο αριθμό βημάτων (run) και ο ρυθμός μάθησης $\tilde{\alpha}$ μειώνεται κατάλληλα, τότε οι τιμές Q θα συγκλίνουν με πιθανότητα 1 στο Q^* (Watkins, 1989 [32]; Tsitsiklis, 1994 [33]; Jaakkola, Jordan & Singh, 1994 [34]). Η Q-μάθηση μπορεί να επεκταθεί έτσι ώστε να ενημερώνει καταστάσεις τις οποίες επισκέφτηκε ο πράκτορας σε περισσότερα από 1 βήματα πριν (Peng & Williams, 1994 [35]).

Όταν οι τιμές Q βρίσκονται κοντά στη σύγκλιση προς τις βέλτιστες τιμές τους, προσφέρεται η ευκαιρία ο πράκτορας να δράσει άπληστα, εκτελώντας σε κάθε περίπτωση την ενέργεια που έχει τη μεγαλύτερη τιμή Q . Κατά τη διάρκεια της μάθησης, όμως, υπάρχει ένας δύσκολος συμβιβασμός (tradeoff) μεταξύ εξερεύνησης και εκμετάλλευσης. Καθώς δεν υπάρχουν ικανοποιητικές τυπικά ορθές μέθοδοι για αυτό το πρόβλημα χρησιμοποιούνται συνήθως ad-hoc τεχνικές.

Η Q-μάθηση δεν είναι ευαίσθητη στην στρατηγική της εξερεύνησης (exploration insensitive): Οι τιμές Q θα συγκλίνουν στις βέλτιστες ανεξάρτητα από τη συμπεριφορά του πράκτορα κατά τη συλλογή των δεδομένων (αρκεί όλα τα ζεύγη κατάστασης-ενέργειας να δοκιμάζονται αρκετά συχνά). Αυτό σημαίνει ότι, μολονότι πρέπει να αντιμετωπιστεί το ζήτημα εξερεύνησης-εκμετάλλευσης, οι λεπτομέρειες της στρατηγικής της εξερεύνησης δεν θα επηρεάσουν τη σύγκλιση του αλγορίθμου μάθησης. Για τους λόγους αυτούς η Q-μάθηση είναι η πλέον δημοφιλής και φαίνεται να είναι η πλέον αποτελεσματική από τους αλγόριθμους χωρίς μοντέλο για τη μάθηση από καθυστερημένες ανταμοιβές. Η Q-μάθηση όμως δεν αντιμετωπίζει τα προβλήματα που προκύπτουν από τη γενίκευση μεγάλων χώρων καταστάσεων ή/και ενεργειών. Επίσης, η σύγκλιση σε μια καλή πολιτική μπορεί να είναι αργή.

Επιπλέον, η Q-μάθηση αποδεικνύεται χρήσιμη ακόμα και σε περιβάλλοντα με πολλούς πράκτορες. Κάθε πράκτορας θεωρείται ανεξάρτητος, εφόσον δεν μπορεί να παρακολουθήσει τις ενέργειες ούτε τις ανταμοιβές των υπολοίπων και εφόσον συμπεριφέρεται σύμφωνα μόνο με την προσωπική του αλληλεπίδραση με το περιβάλλον.

Ο αλγόριθμος για τη δημιουργία του πίνακα $Q(s,a)$ της Q-μάθησης είναι ο ακόλουθος:

Δώσε αρχικές τιμές στο γ και στη συνάρτηση ανταμοιβής $R(s,a)$

Δώσε αρχικές τιμές στο $Q(s,a)$ (συνήθως $Q(s,a)=\{0\}$)

Επανάλαβε για κάθε προσπάθεια

Επίλεξε αρχική κατάσταση s

Επανάλαβε

Διάλεξε μια ενέργεια a που οδηγεί σε νέα κατάσταση s'

Υπολόγισε την μέγιστη τιμή Q για την s' για όλες τις δυνατές ενέργειες a'

Υπολόγισε

$$Q(s, a) := Q(s, a) + \tilde{\alpha}(r + \gamma \max_{a'} Q(s', a') - Q(s, a))$$

$s := s'$

Μέχρι η s να είναι η τελική κατάσταση

Ο αλγόριθμος χρησιμοποίησης του πίνακα $Q(s,a)$ είναι ο ακόλουθος:

Επέλεξε τυχαία αρχική κατάσταση s

Επανάλαβε

Βρες από τον πίνακα $Q(s,a)$ την ενέργεια με τη μεγαλύτερη τιμή Q

Παρατήρησε την επόμενη κατάσταση s'

$s := s'$

Μέχρι η s να είναι η τελική κατάσταση

2.7 Εφαρμογές της ενισχυτικής μάθησης

Ένας από τους λόγους της δημοφιλίας της ενισχυτικής μάθησης είναι η χρήση της, ως θεωρητικού εργαλείου, στη μελέτη των αρχών των πρακτόρων που μαθαίνουν μια συμπεριφορά. Επίσης, έχει χρησιμοποιηθεί από ερευνητές ως πρακτικό υπολογιστικό εργαλείο για την ανάπτυξη αυτόνομων συστημάτων που βελτιώνονται ενόσω αποκτούν

εμπειρία, όπως ρομπότ, βιομηχανικές κατασκευές και συνδυαστικά προβλήματα αναζήτησης.

Οι πρακτικές εφαρμογές αποτελούν ένα τεστ για την αποτελεσματικότητα και τη χρησιμότητα των αλγορίθμων μάθησης. Επιπλέον, βοηθούν στο διαχωρισμό των συνιστωσών του πλαισίου της ενισχυτικής μάθησης που έχουν πρακτική χρησιμότητα. Τέτοιες πρακτικές εφαρμογές είναι τα παίγνια, ένας τομέας που απασχολεί τον κόσμο της τεχνητής νοημοσύνης από το ξεκίνημα του κλάδου, και η ρομποτική και ο έλεγχος.

3 Κεφάλαιο 3 – Περιγραφή Μοντέλου

Στο πλαίσιο της διπλωματικής εργασίας δημιουργήθηκε πρόγραμμα προσομοίωσης, σε C#, το οποίο προσομοιώνει το ασύρματο περιβάλλον που περιγράφεται στο εδάφιο 3.1. Με τη βοήθεια αυτού του προγράμματος προσομοίωσης μελετάται η Δυναμική Επιλογή Καναλιού (Dynamic Channel Selection, DCS) από δευτερεύοντες χρήστες, βασισμένη σε μεθόδους ενισχυτικής μάθησης (Reinforcement Learning-based DCS, RL-DCS). Σκοπός της εργασίας είναι η βελτίωση της απόδοσης της ενισχυτικής μάθησης απλού πράκτορα (που περιγράφεται στην εργασία [37]), που παρουσιάζεται στο 3.1.4.1, χρησιμοποιώντας διαφορετική τεχνική εξερεύνησης (3.1.5) και μέσω ενός νέου πρωτοκόλλου πρόσβασης στο μέσο (3.2.3), το οποίο δεν απαιτεί επικοινωνία μεταξύ των δευτερευόντων χρηστών, όπως συμβαίνει στην ενισχυτική μάθηση πολλαπλών πρακτόρων (3.1.4.2) και το πρωτόκολλο MMAC (Multi agent Medium Access Control, 3.2.2 – περιγράφεται επίσης στην εργασία [37]). Σε πρώτη φάση, το γνωστικό δίκτυο θεωρείται στατικό ως προς την πιθανότητα κατάληψης των καναλιών από τους πρωτεύοντες χρήστες· οι πρωτεύοντες χρήστες έχουν σταθερό μέσο ρυθμό άφιξης πακέτων σε όλη τη διάρκεια της προσομοίωσης.

3.1 Το μοντέλο προσομοίωσης

Το μοντέλο προσομοίωσης είναι βασισμένο στο ασύρματο περιβάλλον και το κατανεμημένο γνωστικό δίκτυο που περιγράφεται στην εργασία [37] και αποτελείται από:

- Ένα σύνολο καναλιών μετάδοσης δεδομένων, $C = [C_1, \dots, C_k]$
- Ένα κανάλι ελέγχου (control channel, CC), κοινό για όλους τους δευτερεύοντες χρήστες
- Ένα σύνολο πρωτευόντων χρηστών (Primary Users, PUs), $PU = [PU_1, \dots, PU_k]$
- Ένα σύνολο δευτερευόντων χρηστών (Secondary Users, SUs), $SU = [SU_1, \dots, SU_n]$

3.1.1 Το κανάλι

Το ασύρματο κανάλι αποτελεί το φυσικό μέσο μετάδοσης δεδομένων για τους πρωτεύοντες και τους δευτερεύοντες χρήστες. Ένα κανάλι C_i χαρακτηρίζεται από την κεντρική συχνότητα F_i και το εύρος ζώνης W_i . Κάθε κανάλι C_i εμφανίζει διαφορετικό ποσοστό λανθασμένων πακέτων (Packet Error Ratio) $PER_{i,u}$ όταν χρησιμοποιείται από διαφορετικούς δευτερεύοντες χρήστες. Έτσι, οι δευτερεύοντες χρήστες αντιλαμβάνονται διαφορετικά κάθε κανάλι. Το ποσοστό λανθασμένων πακέτων εκφράζει κατά ενοποιημένο τρόπο την επίδραση του σηματοθορυβικού λόγου, των παρεμβολών, των διαλείψεων και σκιάσεων και άλλων παραμέτρων επικοινωνίας για κάθε δευτερεύοντα χρήστη.

3.1.2 Το κανάλι ελέγχου

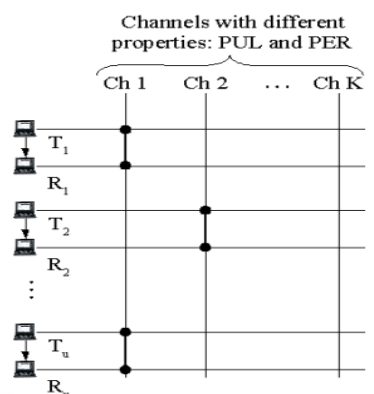
Το κανάλι ελέγχου είναι ένα κανάλι όμοιο με τα κανάλια του συνόλου C , το οποίο μοιράζονται όλοι οι δευτερεύοντες χρήστες για επικοινωνία που δεν αφορά ανταλλαγή δεδομένων αλλά ανταλλαγή μηνυμάτων ελέγχου μεταξύ τους.

3.1.3 Πρωτεύοντες Χρήστες (Primary Users)

Ως πρωτεύων χρήστης PU_i θεωρείται το ζεύγος πομπού-δέκτη το οποίο έχει την άδεια να χρησιμοποιήσει το κανάλι δεδομένων C_i αποκλειστικά. Κάθε πρωτεύων χρήστης (PU) χαρακτηρίζεται από ένα επίπεδο χρησιμοποίησης (PU Utilization Level,) PUL_i , όπου $0 < PUL_i < 1$. Όσο υψηλότερο το PUL_i τόσο λιγότερα τα κενά διαστήματα (whitespaces) στο κανάλι C_i . Κάθε PU διαθέτει δική του ουρά αναμονής (queue), και όποτε υπάρχουν πακέτα σε αυτή, μεταδίδει. Ο ρυθμός άφιξης των πακέτων στην ουρά αναμονής ακολουθεί την κατανομή Poisson, με μέση τιμή (ή ρυθμό) εξαρτώμενη από το PUL, ανεξάρτητη και όμοια κατανεμημένη (i.i.d.) για όλους τους πρωτεύοντες χρήστες (PUs). Επίσης, θεωρείται ότι οι πρωτεύοντες χρήστες δεν χρησιμοποιούν τετραμερή χειραψία (four-way handshake) στην αρχή των μεταδόσεων τους αλλά μεταδίδουν αμέσως.

3.1.4 Δευτερεύοντες Χρήστες (Secondary Users)

Αντίστοιχα με τους πρωτεύοντες, ο δευτερεύων χρήστης (ή πράκτορας) SU είναι το ζεύγος του πομπού και του δέκτη που χρησιμοποιούν κάποιο κανάλι για τη μετάδοση δεδομένων. T_i είναι ο πομπός και R_i ο δέκτης του πράκτορα SU_i . Στη γενική περίπτωση το πλήθος των δευτερευόντων χρηστών είναι μεγαλύτερο ή ίσο με το πλήθος των καναλιών μετάδοσης δεδομένων $|C| \leq |SU|$, οπότε οι δευτερεύοντες χρήστες ανταγωνίζονται και μεταξύ τους για τη χρήση των καναλιών. Το κανάλι ελέγχου χρησιμοποιείται από τους δευτερεύοντες χρήστες για τη μετάδοση επιβεβαιώσεων (πακέτων ACK), πληροφοριών σχετικά με το ποιο κανάλι έχει επιλεγεί για μετάδοση μεταξύ πομπού και δέκτη καθώς και άλλων δευτερευόντων μηνυμάτων. Κάθε στιγμή θεωρείται ότι οι δευτερεύοντες χρήστες διαθέτουν πακέτα προς μετάδοση (backlogged). Επομένως, η ουρά αναμονής τους δεν είναι ποτέ άδεια.

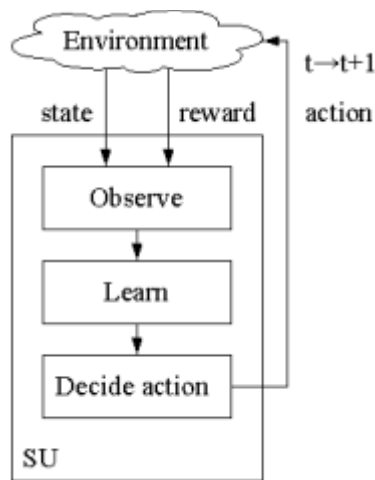


Σχήμα 3-1 - Ζεύγη πομπού-δέκτη δευτερευόντων χρηστών.

Οι δευτερεύοντες χρήστες χρησιμοποιούν ενισχυτική μάθηση για να αποκτήσουν επίγνωση του ασύρματου περιβάλλοντος και να επιλέξουν το κανάλι στο οποίο θα πραγματοποιήσουν τη μετάδοσή τους (Δυναμική Επιλογή Καναλιού βασισμένη στην ενισχυτική μάθηση, RL-based DCS). Χρησιμοποιούνται δύο είδη ενισχυτικής μάθησης: η ενισχυτική μάθηση απλού πράκτορα (Single-Agent) και η ενισχυτική μάθηση πολλαπλών πρακτόρων (Multi-Agent).

3.1.4.1 Ενισχυτική Μάθηση Απλού Πράκτορα - SARL

Ο αλγόριθμος μάθησης που χρησιμοποιείται από τους δευτερεύοντες χρήστες για τον προσδιορισμό των χαρακτηριστικών του περιβάλλοντος είναι η Q-μάθηση, καθώς προσφέρει τη δυνατότητα απλής μαθηματικής περιγραφής του ασύρματου περιβάλλοντος μειώνοντας έτσι την απαιτούμενη πολυπλοκότητα.



Σχήμα 3-2 - Διάγραμμα λειτουργίας πράκτορα SARL.

Έστω ότι ο χρόνος χωρίζεται σε χρονικές περιστάσεις (γνωστές και ως εποχές) $t \in T = \{1, 2, \dots\}$ διάρκειας t_D . Κάθε πράκτορας i διατηρεί ένα πίνακα με τιμές Q , ο οποίος διαθέτει $|A|$ καταχωρίσεις, όπου A το σύνολο πιθανών ενεργειών του πράκτορα. Στην περίπτωση της διπλωματικής εργασίας ως ενέργεια θεωρείται η επιλογή ενός καναλιού για μετάδοση, έτσι $|A| = |C|$. Κάθε καταχώρηση του πίνακα περιέχει μια τιμή $Q_t(a)$ στο διάστημα $[0, Q_{max}]$, όπου $a \in A$. Η τιμή Q αποτελεί μια εκτίμηση της ανταμοιβής που θα λάβει ο πράκτορας για κάποια ενέργεια a . Σε κάθε περίπτωση (εποχή) t , ο πράκτορας i επιλέγει μια ενέργεια a_t και λαμβάνει μια (καθυστερημένη) ανταμοιβή $r_{t+1}(a_t)$ την επόμενη εποχή. Ο πράκτορας i ενημερώνει την αντίστοιχη καταχώριση του πίνακα τιμών Q σύμφωνα με τη σχέση

$$Q_{t+1}^i(a_t^i) \leftarrow (1 - \tilde{\alpha})Q_t^i(a_t^i) + \tilde{\alpha}r_{t+1}^i(a_t^i) \quad (3.1)$$

όπου $0 \leq \tilde{\alpha} \leq 1$ είναι ο ρυθμός μάθησης (learning rate). Υψηλές τιμές του $\tilde{\alpha}$ οδηγούν σε υψηλή εξάρτηση από τις τελευταίες ανταμοιβές και μικρότερη εξάρτηση από την πρότερη

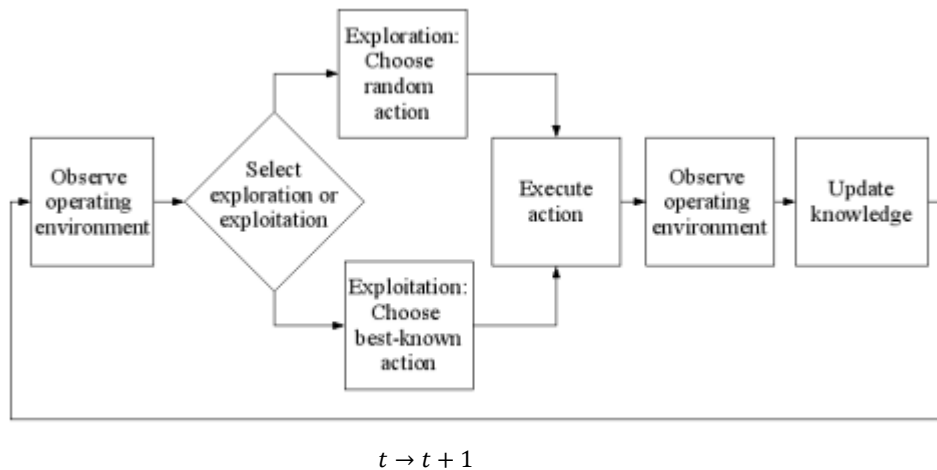
εμπειρία του πράκτορα. Ο πράκτορας αποφασίζει την ενέργεια του ανάλογα με τη βέλτιστη πολιτική που μεγιστοποιεί τη συνάρτηση τιμής

$$V^\pi = \max_{a \in A} (Q_t^i(a)) \quad (3.2)$$

Δηλαδή κάθε φορά επιλέγεται η ενέργεια η οποία έχει την υψηλότερη τιμή στον πίνακα τιμών Q. Η ανταμοιβή $r_{t+1}(a_t)$ είναι η ρυθμαπόδοση (throughput) που παρατηρείται από το σύστημα πομπού-δέκτη κατά τη διάρκεια μιας εποχής:

$$r_{t+1}(a_t) = \frac{N_D}{t_D} \quad (3.3)$$

όπου N_D το πλήθος των πακέτων δεδομένων που μεταδόθηκαν επιτυχώς από τον πομπό κατά την διάρκεια t_D της εποχής t . Μια μετάδοση πακέτων δεδομένων θεωρείται επιτυχής όταν ληφθεί η αντίστοιχη επιβεβαίωση (ACK) στο κανάλι ελέγχου CC.



Σχήμα 3-3 - Αναλυτικό διάγραμμα λειτουργίας δευτερεύοντος χρήστη.

3.1.4.2 Ενισχυτική Μάθηση Πολλαπλών Πρακτόρων - MARL

Η ενισχυτική μάθηση πολλαπλών πρακτόρων αποτελεί συνδυασμό της ενισχυτικής μάθησης απλού πράκτορα και ενός μηχανισμού διάδοσης ανταμοιβών (payoff propagation). Η ενισχυτική μάθηση απλού πράκτορα χρησιμοποιείται για την εύρεση της ατομικής ανταμοιβής ενώ ο μηχανισμός διάδοσης ανταμοιβών για την επικοινωνία μεταξύ των πρακτόρων.

Εκτός από τον πίνακα τιμών Q, κάθε πράκτορας i διαθέτει και ένα πίνακα μηνυμάτων ανταμοιβής με διαστάσεις $|\Gamma(i)| \times |A|$, όπου $\Gamma(i)$ το σύνολο των γειτονικών χρηστών του πράκτορα i . Γειτονικοί πράκτορες $\Gamma(i)$ θεωρούνται αυτοί των οποίων οι εκπομπές επηρεάζουν και επηρεάζονται από τις εκπομπές του πράκτορα i . Κάθε πράκτορας υπολογίζει τη δική του ενέργεια a_t^i με βάση τις δικές του τιμές Q από τον πίνακα τιμών Q $\{Q_t^i(a_t^i)\}$ και τις τιμές Q των γειτόνων του από τον πίνακα μηνυμάτων ανταμοιβής

$\{Q_t^{j \in \Gamma(i)}(a_t^j)\}$. Επιπλέον, κάθε πράκτορας i αποστέλλει συνεχώς μηνύματα ανταμοιβής (payoff message) $\mu_{ij}(a_t^j)$ στους γειτονικούς του πράκτορες $j \in \Gamma(i)$ για να ενημερώνουν τους δικούς τους πίνακες, αντίστοιχα. Το μήνυμα ανταμοιβής είναι η ανανεωμένη τιμή Q της επιλεγμένης ενέργειας του πράκτορα

$$\mu_{ij}(a_t^j) = [Q_t^i(a_t^i)] \quad (3.4)$$

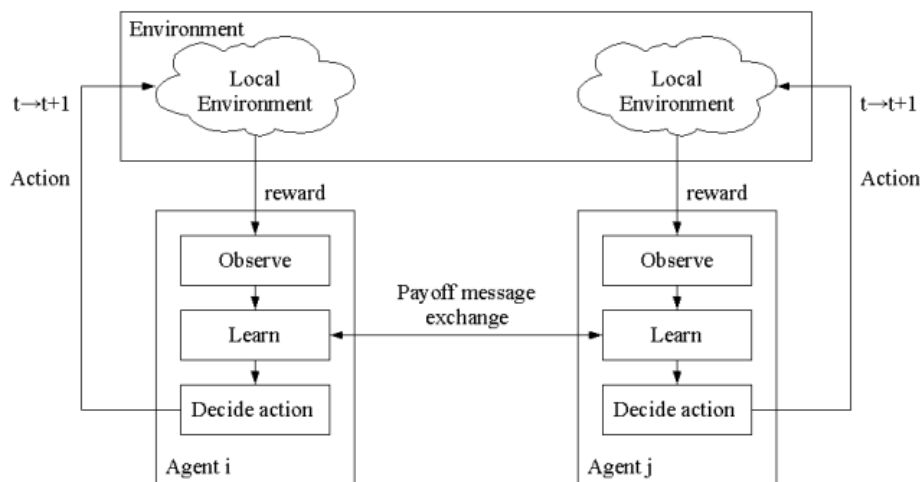
Κατ'αυτόν τον τρόπο, κάθε πράκτορας επιλέγει τη δική του βέλτιστη ενέργεια με στόχο να μεγιστοποιήσει το προσωπικό του όφελος

$$g_t^i(a_t^i) = \max_{a \in A} \left[Q_t^i(a) + \sum_{j \in \Gamma(i)} \mu_{ji}(a) \right] \quad (3.5)$$

όπου $Q_t^i(a)$ είναι η τιμή για την ενέργεια a όπως την έχει εκτιμήσει και αποθηκεύσει ο πράκτορας i ενώ $\mu_{ji}(a)$ είναι η τιμή για την ενέργεια a όπως την έχει λάβει από το γειτονικό πράκτορα j . Η βέλτιστη ενέργεια επιλέγεται μέσω της σχέσης

$$a_t^i = \arg \max_{a \in A} g_t^i(a) \quad (3.6)$$

Ο αλγόριθμος ενισχυτικής μάθησης και η μετάδοση ανταμοιβών συνεχίζονται μέχρι να επέλθει σύγκλιση σε κάποια σταθερή τιμή της καθολικής Q τιμής όλων των χρηστών (Global Q-value).



Σχήμα 3-4 - Διάγραμμα λειτουργίας πράκτορα MARL.

3.1.5 Μηχανισμοί εξερεύνησης

Σε κάθε εποχή ενημερώνονται οι τιμές Q των ενεργειών που επιλέχθηκαν. Καθώς αυτός ο μηχανισμός επιλογής καναλιού (μηχανισμός αξιοποίησης) επιλέγει τη βέλτιστη ενέργεια

σύμφωνα με τα διαθέσιμα δεδομένα (γνωστή και ως άπληστη πολιτική), προκύπτει η ανάγκη ύπαρξης ενός μηχανισμού εξερεύνησης που να ωθεί στην αξιολόγηση όλων των ενεργειών. Ο μηχανισμός εξερεύνησης είναι εκείνος που διαλέγει μια από τις υπόλοιπες μη-βέλτιστες ενέργειες σε μια τυχαία εποχή, ώστε να προσδιοριστούν ενδεχομένως καλύτερες ενέργειες. Γενικά, ένας πράκτορας πραγματοποιεί εξερεύνηση με μια μικρή πιθανότητα ϵ και εκμετάλλευση με πιθανότητα $1-\epsilon$.

Κατά την εξερεύνηση, οι πιθανότητες των ενεργειών ακολουθούν ομοιόμορφη κατανομή, δηλαδή η πιθανότητα επιλογής κάποιας ενέργειας a είναι

$$Pr(a) = \frac{1}{|A|} \quad (3.7)$$

όπου A το σύνολο των πιθανών ενεργειών του πράκτορα.

Στην παρούσα εργασία προσομοίωσης ερευνάται και η επίδοση ενός άλλου μηχανισμού εξερεύνησης. Όπως και προηγουμένως, σύμφωνα με τον μηχανισμό αυτόν ο πράκτορας πραγματοποιεί εξερεύνηση με μια μικρή πιθανότητα αλλά η πιθανότητα επιλογής μιας ενέργειας a εξαρτάται από την τιμή της στον πίνακα Q διαιρεμένη με το άθροισμα των τιμών του πίνακα Q για όλες τις ενέργειες του συνόλου A

$$Pr(a) = \frac{Q_t(a)}{\sum_{x \in A} Q_t(x)} \quad (3.8)$$

Στην περίπτωση πολλαπλών πρακτόρων αντί του $Q_t(a)$ χρησιμοποιείται το άθροισμα: $Q_t(a) + \sum_{j \in \Gamma(i)} \mu_j(a)$

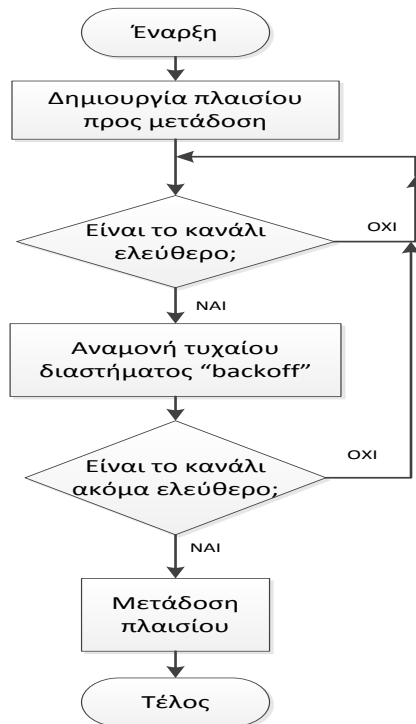
3.1.6 Μετρήσιμα μεγέθη

Τα κύρια μεγέθη τα οποία μετρώνται από τους δευτερεύοντες χρήστες και μπορούν να χρησιμοποιηθούν προς εκτίμηση των επιδόσεων τους είναι η συνολική ρυθμαπόδοση, δηλαδή το σύνολο των πακέτων που στάλθηκαν επιτυχώς κατά τη διάρκεια της προσομοίωσης, ο αριθμός των εναλλαγών καναλιού δεδομένων και το ποσοστό των χαμένων πακέτων (PER). Καθώς η προσομοίωση λειτουργεί στο στρώμα MAC του δικτύου, μετράται επίσης και το πλήθος των υπαναχωρήσεων εκπομπής (back offs) που πραγματοποιεί κάθε δευτερεύων χρήστης λόγω της χρήσης του πρωτοκόλλου CSMA-CA.

3.2 Πρωτόκολλα Πρόσβασης στο Μέσο (MAC)

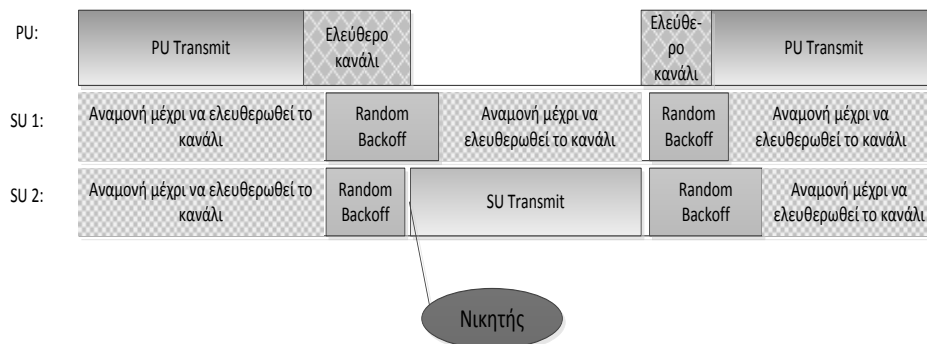
Κάθε δευτερεύων χρήστης διαθέτει δύο πομποδέκτες, ένα για τη μετάδοση δεδομένων και ένα αποκλειστικά για το κανάλι ελέγχου. Το κανάλι ελέγχου χρησιμοποιείται από τον πομπό για να στείλει πληροφορίες για την εναλλαγή καναλιού δεδομένων στο δέκτη. Ο πομποδέκτης δεδομένων μπορεί να συντονιστεί σε οποιοδήποτε από τα διαθέσιμα κανάλια δεδομένων. Καθώς η πρόσβαση στο φάσμα είναι υπερστρωματική (το φάσμα χρησιμοποιείται όταν δεν μεταδίδει πρωτεύων ή άλλος δευτερεύων χρήστης σε αυτό. Εδάφιο 1.9), οι δευτερεύοντες χρήστες χρησιμοποιούν μια απλοποιημένη έκδοση του

CSMA-CA (Carrier Sense Multiple Access – Collision Avoidance) για την αποφυγή συγκρούσεων μεταξύ τους αλλά και με τους πρωτεύοντες χρήστες.



Σχήμα 3-5 - Διάγραμμα λειτουργίας απλοποιημένου CSMA-CA.

Για λόγους απλότητας οι χρόνοι επεξεργασίας των δεδομένων και ενημέρωσης των τιμών Q κάθε χρήστη θεωρούνται αμελητέοι. Επιπλέον, επειδή τα πακέτα επιβεβαίωσης (ACK) είναι μικρά, η μετάδοσή τους θεωρείται ακαριαία και χωρίς λάθη. Οι πρωτεύοντες χρήστες έχουν κοινό χρόνο μετάδοσης πακέτων δεδομένων, $t_{DATA,PU}$. Το ίδιο ισχύει και για όλους τους δευτερεύοντες που έχουν κοινό χρόνο μετάδοσης πακέτου δεδομένων $t_{DATA,SU}$. Επίσης, όλοι οι πρωτεύοντες και οι δευτερεύοντες χρήστες θεωρείται ότι βρίσκονται σε μικρές αποστάσεις, οπότε οι μεταδόσεις τους επηρεάζουν το ασύρματο περιβάλλον όλων.



Σχήμα 3-6 - CSMA-CA.

Παρά τη χρήση του CSMA-CA από τους δευτερεύοντες χρήστες, υπάρχει η πιθανότητα σύγκρουσης δευτερευόντων και πρωτεύοντων χρηστών καθώς οι πρωτεύοντες χρήστες μεταδίδουν όποτε διαθέτουν πακέτα χωρίς να ανιχνεύουν το κανάλι πριν μεταδώσουν. Η πιθανότητα αποτυχημένης μετάδοσης πακέτου, ενός δευτερεύοντος χρήστη στο κανάλι C_i προκύπτει από την ακόλουθη σχέση:

$$\begin{aligned} Pr\{\text{Αποτυχημένη Μετάδοση}\} &= 1 - \left\{ (1 - PER_i) \left(1 - \int_0^{t_{DATA,SU}} PUL_i \cdot e^{-t \cdot PUL_i} dt \right) \right\} = \\ &= 1 - \{ (1 - PER_i) \cdot e^{-t_{DATA,SU} \cdot PUL_i} \} \end{aligned} \quad (3.9)$$

όπου PER_i το ποσοστό λανθασμένων μεταδόσεων του δευτερεύοντος χρήστη για το κανάλι C_i και PUL_i το επίπεδο χρησιμοποίησης πρωτεύοντος χρήστη για το ίδιο κανάλι. Η ποσότητα $1 - \int_0^{t_{DATA,SU}} PUL_i \cdot e^{-t \cdot PUL_i} dt$ είναι η πιθανότητα να μην εκκινήσει πρωτεύων χρήστης μετάδοση κατά τη διάρκεια μετάδοσης ενός πακέτου του δευτερεύοντος.

Κατά τις προσομοιώσεις δοκιμάστηκαν τρία διαφορετικά πρωτόκολλα πρόσβασης στο μέσο. Σε όλα τα πρωτόκολλα μπορούν να χρησιμοποιηθούν οι μηχανισμοί εξερεύνησης που αναφέρθηκαν στο 3.1.5.

3.2.1 SMAC (Single agent Medium Access Control)

Στο πρωτόκολλο SMAC ένας δευτερεύων χρήστης λειτουργεί ως πράκτορας SARL για την επιλογή του καναλιού δεδομένων. Κάθε SU διαιρεί το χρονικό του ορίζοντα σε εποχές που διαρκούν $t_D = t_{D,SMAC}$. Σε κάθε εποχή, ο SU μετρά το πλήθος των πακέτων που στάλθηκαν με επιτυχία, N_D , για να βρει την ανταμοιβή $r_{t+1}(a_t)$ και να ενημερώσει τον πίνακα με τις τιμές Q βάσει της εξίσωσης (3.1). Ένας SU επιλέγει με πιθανότητα $1-\epsilon$ το κανάλι δεδομένων για την επόμενη εποχή, ανάλογα με τις τιμές του πίνακα Q. Για τη αποφυγή συνεχών εναλλαγών, εφόσον η τιμή Q της νέας επιλογής καναλιού δεν είναι μεγαλύτερη από την τιμή Q της προηγούμενης κατά τουλάχιστον ένα κατώφλι β , δεν πραγματοποιείται αλλαγή καναλιού. Η εξερεύνηση πραγματοποιείται με πιθανότητα ϵ και δεν πραγματοποιείται σε δύο συνεχόμενες εποχές. Η απόφαση για την αλλαγή καναλιού λαμβάνεται από το χρήστη στην αρχή της εποχής αλλά η μετάδοση δεδομένων πραγματοποιείται μόνο αφού έχει αποσταλεί μήνυμα, μέσω του καναλιού ελέγχου, από τον πομπό στο δέκτη, ενημερώνοντας για την αλλαγή καναλιού. Όταν πραγματοποιηθεί αυτή η ενημέρωση, ο χρήστης θέτει $N_D = 0$ και εκκινεί τη μετάδοση δεδομένων. Στην πρώτη εποχή λειτουργίας το κανάλι δεδομένων επιλέγεται τυχαία.

3.2.2 MMAC (Multi agent Medium Access Control)

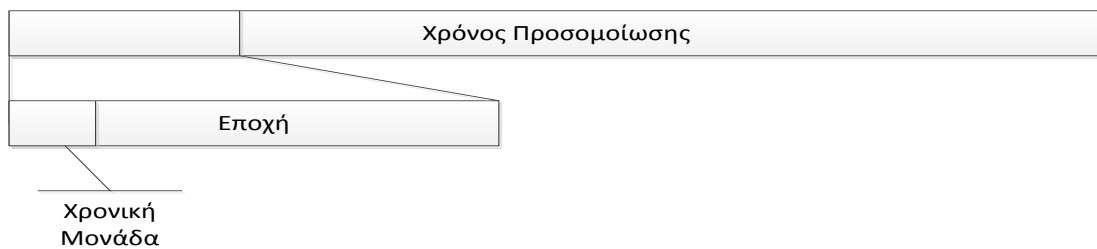
Στο πρωτόκολλο MMAC ένας δευτερεύων χρήστης λειτουργεί ως πράκτορας MARL για την επιλογή του καναλιού δεδομένων. Ο χρονικός ορίζοντας διαιρείται σε εποχές διάρκειας $t_D = t_{D,MMAC}$. Κατά τη διάρκεια μιας εποχής ο SU μετρά τα πακέτα που στάλθηκαν επιτυχώς αλλά και αποστέλλει στους γειτονικούς του χρήστες την τιμή Q της τρέχουσας ενέργειας του. Η μετάδοση αυτή γίνεται μέσω του καναλιού ελέγχου και πραγματοποιείται μόνο κατά την εκμετάλλευση. Το κανάλι στο οποίο πραγματοποιείται η μετάδοση δεδομένων επιλέγεται βάσει της εξίσωσης (3.6), και η μετάδοση δεδομένων εκκινεί αφού σταλεί κατάλληλο μήνυμα από τον πομπό στο δέκτη για την αλλαγή του καναλιού, όπως και στο πρωτόκολλο SMAC. Η διάρκεια της εποχής είναι μεγαλύτερη από αυτή του πρωτοκόλλου SMAC για να αντισταθμιστεί ο χρόνος αποστολής των μηνυμάτων ανταμοιβής. Το πρωτόκολλο MMAC επιτυγχάνει μειωμένο πλήθος εναλλαγών καναλιού και κατά συνέπεια μεγαλύτερη σταθερότητα στο σύστημα.

3.2.3 SMAC με έλεγχο υπαναχωρήσεων (Backoff)

Παρά την καλή λειτουργία τους, τα πρωτόκολλα SMAC και MMAC εμφανίζουν μειωμένη απόδοση στις περιπτώσεις όπου το κανάλι στο οποίο λειτουργούν εμφανίζει συμφόρηση από άλλους δευτερεύοντες χρήστες. Προς βελτίωση της επίδοσης σε καταστάσεις συμφόρησης επινοήθηκε στο πλαίσιο της διπλωματικής εργασίας μια παραλλαγή του πρωτοκόλλου SMAC, το πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων, το οποίο υλοποιήθηκε στον προσομοιωτή και του οποίου οι επιδόσεις επιβεβαιώθηκαν από τα αποτελέσματα των προσομοιώσεων. Το πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων λειτουργεί παρόμοια με το πρωτόκολλο SMAC, αφού ο SU λειτουργεί και ως πράκτορας SARL. Ο χρονικός ορίζοντας διαιρείται σε εποχές διάρκειας $t_D = t_{D,SMAC}$ και σε κάθε εποχή, εκτός από τα πακέτα που στάλθηκαν επιτυχώς, ο χρήστης μετρά και τις φορές που αναγκάστηκε να κάνει υπαναχώρηση (backoff) στο κανάλι δεδομένων. Αν το πλήθος φορών ανάμεσα σε εποχές, υπερβεί ένα κατώφλι, θεωρείται ότι έχουν αφιχθεί και άλλοι δευτερεύοντες χρήστες στο κανάλι, οπότε ο πράκτορας εκκινεί τη διαδικασία αλλαγής καναλιού μέσω εξερεύνησης ή μέσω εκμετάλλευσης. Σε κάθε εποχή ενημερώνεται ο πίνακας Q παρομοίως με το πρωτόκολλο SMAC. Επίσης, η διαδικασία αλλαγής καναλιού ενεργοποιείται όταν παρουσιαστεί μεγάλη πτώση της τιμής Q του καναλιού στο οποίο λειτουργεί εκείνη τη στιγμή ο χρήστης. Καθώς ο πράκτορας σε πρώτη φάση ελέγχει το κανάλι για συμφόρηση (μέσω της μέτρησης του πλήθους των υπαναχωρήσεων) υπάρχει η πιθανότητα να παραμείνει για πολλή ώρα σε ένα κανάλι που ενδεχομένως δεν εμφανίζει ευνοϊκές συνθήκες για μετάδοση (υψηλό PER ή/και PUL). Για την αποφυγή μιας τέτοιας κατάστασης ο χρήστης πραγματοποιεί πρώτα μια εξερεύνηση όλων των πιθανών καναλιών, έτσι ώστε να διαθέτει μια αρχική τιμή Q για έκαστο εξ αυτών.

3.3 Ο προσομοιωτής

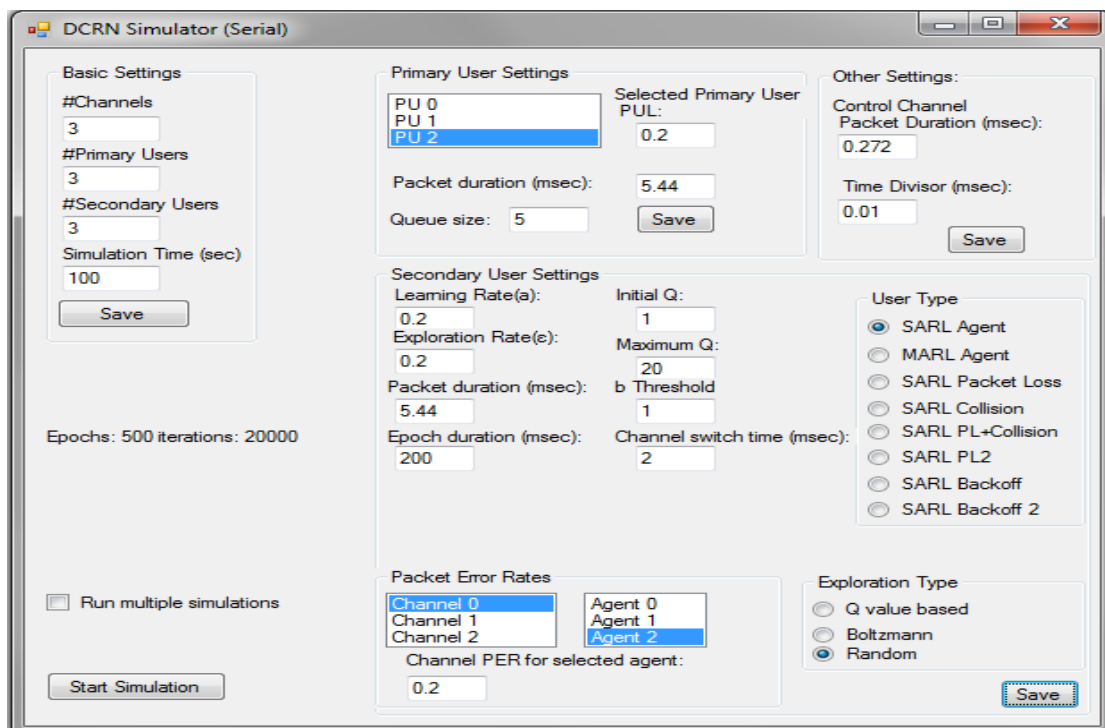
Το πρόγραμμα προσομοίωσης χωρίζει το χρονικό ορίζοντα της προσομοίωσης σε εποχές διάρκειας t_D . Στη συνέχεια, χωρίζει αυτές τις εποχές σε μικρότερες χρονικές διάρκειες.



Σχήμα 3-7 - Διαμερισμός χρόνου στον προσομοιωτή.

Σε κάθε χρονική μονάδα και για κάθε κανάλι ο προσομοιωτής ελέγχει αν εκπέμπει κάποιος πρωτεύων ή δευτερεύων χρήστης και ενημερώνει τις καταστάσεις των αντίστοιχων καναλιών.

Στο κατωτέρω σχήμα εμφανίζεται το γραφικό περιβάλλον του προσομοιωτή:



Σχήμα 3-8 - Γραφικό περιβάλλον προσομοιωτή

Από το γραφικό περιβάλλον δίνεται η δυνατότητα επιλογής των ακόλουθων παραμέτρων

- πλήθος καναλιών δεδομένων
- πλήθος πρωτευόντων χρηστών
- πλήθος δευτερευόντων χρηστών
- συνολικός χρόνος της προσομοίωσης.

Για τους πρωτεύοντες χρήστες δίνεται η δυνατότητα ρύθμισης των ακόλουθων παραμέτρων

- μέγεθος της ουράς αναμονής (Queue size)
- βαθμός χρησιμοποίησης (PUL – PU Utilization Level)
- χρόνος μετάδοσης που απαιτείται για τη μετάδοση των πακέτων τους σε msec (Packet duration).

Για τους δευτερεύοντες χρήστες δίνεται η δυνατότητα ρύθμισης των ακόλουθων παραμέτρων

- ρυθμός μάθησης (learning rate, σ)
- ρυθμός εξερεύνησης (exploration rate, ϵ)
- αρχική τιμή Q για κάθε κανάλι (Initial Q)
- μέγιστη πιθανή τιμή του Q (maximum Q, Q_{max})
- η τιμή του κατωφλίου β
- διάρκεια εποχής σε msec (Epoch duration)
- διάστημα αλλαγής καναλιού σε msec (Channel switch time)
- χρόνος μετάδοσης πακέτου σε msec (Packet duration)
- το πρωτόκολλο MAC (SMAC,MMAC, SMAC με έλεγχο υπαναχωρήσεων)
- το είδος εξερεύνησης (τυχαίο ή εξαρτώμενο από τις τιμές Q) και
- το ποσοστό λαθών (PER) που αντιλαμβάνεται κάθε χρήστης σε κάθε κανάλι.

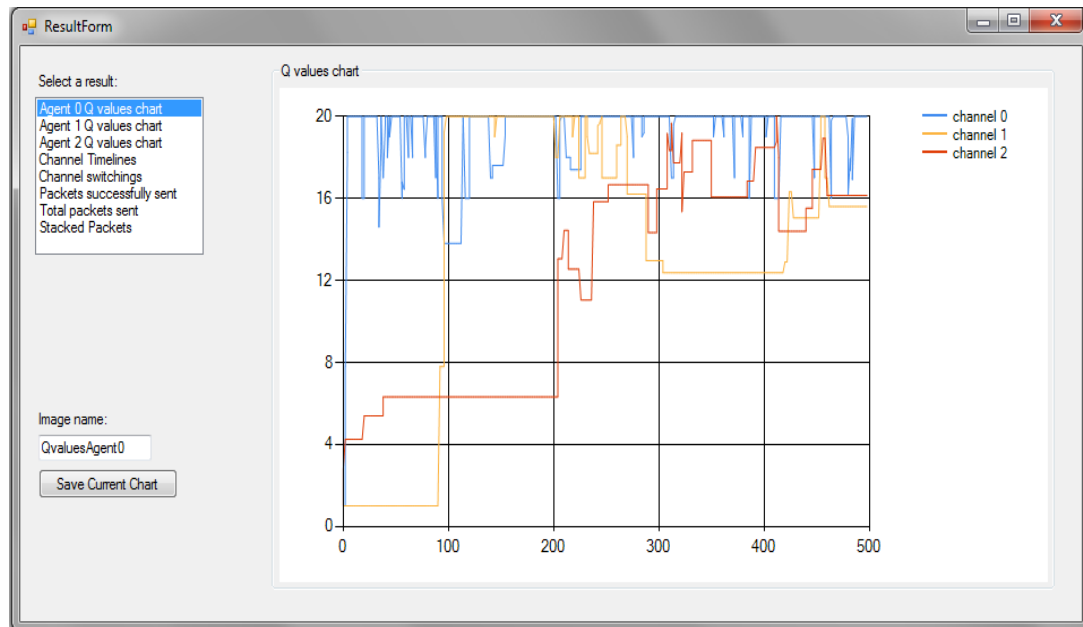
Στην περίπτωση του πρωτοκόλλου SMAC με έλεγχο υπαναχωρήσεων, παρέχεται η δυνατότητα επιλογής του κατωφλίου των υπαναχωρήσεων από το οποίο εξαρτάται η ενεργοποίηση του μηχανισμού αλλαγής καναλιού στο συγκεκριμένο πρωτόκολλο. Επιπλέον, παρέχεται η δυνατότητα ρύθμισης του χρόνου μετάδοσης των πακέτων στο κανάλι ελέγχου σε msec (Control Channel Packet Duration) και της χρονικής μονάδας με την οποία διαιρείται κάθε εποχή σε msec (Time Divisor).

Μετά την ολοκλήρωση της προσομοίωσης εμφανίζεται ένα παράθυρο αποτελεσμάτων (Σχήμα 3-9), το οποίο περιέχει:

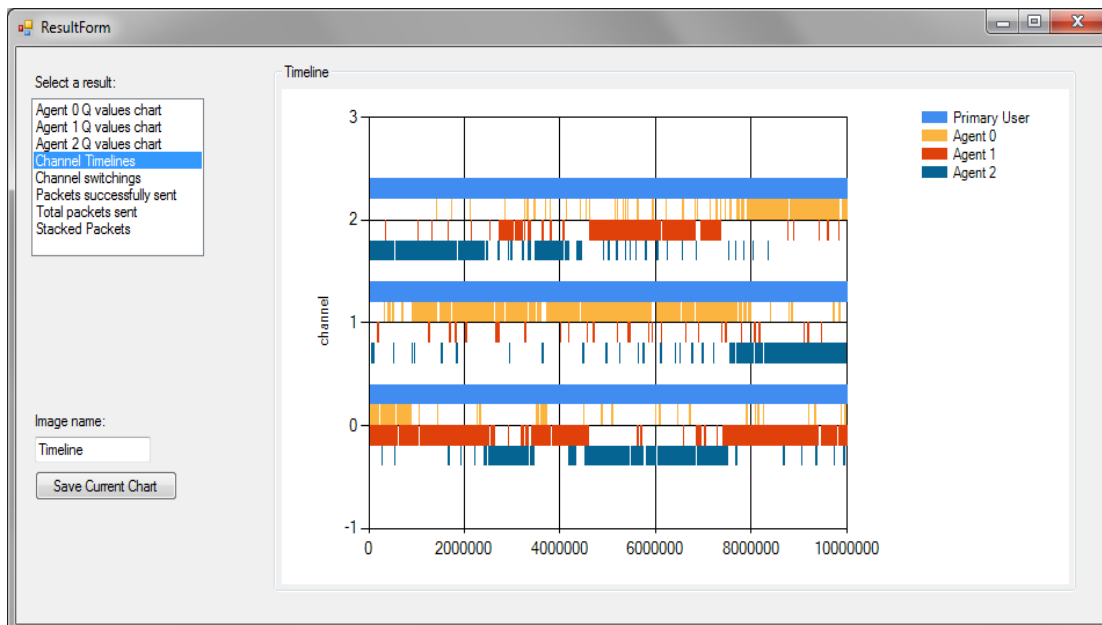
- διαγράμματα των τιμών Q συναρτήσει των εποχών για κάθε χρήστη
- το πλήθος εναλλαγών καναλιού για κάθε χρήστη
- τα πακέτα που στάλθηκαν με επιτυχία
- το συνολικό πλήθος πακέτων που στάλθηκαν από κάθε χρήστη

- και ένα αναλυτικό διάγραμμα της μετάδοσης πακέτων σε κάθε κανάλι, τόσο για τους πρωτεύοντες όσο και για τους δευτερεύοντες χρήστες (Σχήμα 3-10).

Επίσης, παρέχεται η δυνατότητα εκτέλεσης πολλαπλών ομοίων προσομοιώσεων. Τα αποτελέσματα αυτών δίνονται σε αρχείο .txt (Σχήμα 3-11).



Σχήμα 3-9 - Παράθυρο αποτελεσμάτων προσομοιωτή.



Σχήμα 3-10 - Διάγραμμα μετάδοσης πακέτων / χρήσης καναλιών.

```

multi-sim.txt - Notepad
File Edit Format View Help
Simulation Results for 3 runs
21/2/2013 7:16:19 μμ
PUL[0] = 0.5 PUL[1] = 0.5 PUL[2] = 0.5
Channel 0: PER[0]=0.2 PER[1]=0.2 PER[2]=0.2
Channel 1: PER[0]=0.2 PER[1]=0.2 PER[2]=0.2
Channel 2: PER[0]=0.2 PER[1]=0.2 PER[2]=0.2
Type of agents: SARL
Secondary user Settings:
Learning Rate: 0.2 Exploration Rate: 0.2 Packet Duration: 5.44 Epoch Duration: 200
Initial Q: 1 Max Q: 20 b threshold: 1 Channel Switching time: 2
Exploration Type: Q-based
Lines represent different runs, columns represent different agents
-----
Results for Agents:
Preffered Channel:
1 2 0
0 2 1
2 1 0
-----
Channel Switchings:
31 27 26
27 32 24
12 40 43
Means:
23.333333333333 33 31
Total Mean: 29.111111111111
-----
Packets successfullly sent:
4169 3704 3934
3969 3799 4318
4420 4165 4167
Means:
4186 3889.3333333333 4139.6666666667
Total Mean: 4071.6666666667
-----
Total packets sent:
8932 7987 8354
8664 8009 9096
9339 8822 8748
Means:
8978.3333333333 8272.6666666667 8732.6666666667
Total Mean: 8661.2222222222

```

Σχήμα 3-11 - Αρχείο text με τα αποτελέσματα πολλών προσομοιώσεων.

3.4 Συνεισφορές της παρούσας εργασίας

Για τη διεκπεραίωση της διπλωματικής εργασίας και για την εξαγωγή συμπερασμάτων δημιουργήθηκε εξ αρχής ένας νέος προσομοιωτής χρησιμοποιώντας τη γλώσσα προγραμματισμού C# και το λογισμικό Visual Studio 2010 της εταιρίας Microsoft. Η αρχιτεκτονική του προσομοιωτή επιτρέπει την εισαγωγή νέων ειδών πρακτόρων και πρωτοκόλλων πέραν αυτών που περιγράφονται στην παρούσα εργασία.

Το μοντέλο προσομοίωσης καθώς και τα πρωτόκολλα πρόσβασης στο μέσο SMAC (Εδάφιο 3.2.1) και MMAC (Εδάφιο 3.2.2) των πρακτόρων είναι βασισμένα σε αυτά που περιγράφονται στην εργασία [37]. Τα αποτελέσματα για τα ίδια σενάρια προσομοιώσεων είναι όμοια με τα αποτελέσματα της εργασίας [37] γεγονός που φανερώνει ότι ο προσομοιωτής που αναπτύχθηκε είναι εξίσου αξιόπιστος με το εργαλείο λογισμικού OMNET++ που χρησιμοποιήθηκε στην εργασία [37]. Το πρωτόκολλο πρόσβασης στο μέσο SMAC με έλεγχο υπαναχωρήσεων (Εδάφιο 3.2.3) επινοήθηκε και υλοποιήθηκε με τη μορφή προσομοίωσης για πρώτη φορά στην παρούσα εργασία. Επίσης, η τεχνική εξερεύνησης με πιθανότητες βασισμένες στις τιμές Q (Εδάφιο 3.1.5) παρουσιάστηκε για πρώτη φορά στην παρούσα εργασία καθώς και η μελέτη των επιδόσεων της στο πλαίσιο του συγκεκριμένου μοντέλου προσομοίωσης.

4 Κεφάλαιο 4 – Αποτελέσματα

Στο κεφάλαιο αυτό παρουσιάζονται αποτελέσματα που λήφθηκαν από τον προσομοιωτή που δημιουργήθηκε σε C#, με τις παραμέτρους που παρουσιάζονται στο εδάφιο 4.1. Οι δευτερεύοντες χρήστες που χρησιμοποιούν ενισχυτική μάθηση απλού πράκτορα (SARL), υλοποιούνται με τα πρωτόκολλα SMAC και SMAC με έλεγχο για υπαναχωρήσεις, ενώ οι χρήστες που χρησιμοποιούν ενισχυτική μάθηση πολλών πρακτόρων, υλοποιούνται με το πρωτόκολλο MMAC. Στην περίπτωση του MMAC, στο σύνολο γειτόνων $\Gamma(i)$ κάθε χρήστη περιλαμβάνονται όλοι οι υπόλοιποι $|SU|-1$ δευτερεύοντες χρήστες. Οι δευτερεύοντες χρήστες μπορούν σε κάθε εποχή να χρησιμοποιήσουν οποιοδήποτε από τα κανάλια δεδομένων, ενώ οι πρωτεύοντες παραμένουν στο δικό τους κανάλι. Το κανάλι ελέγχου είναι κοινό για όλους τους δευτερεύοντες χρήστες οι οποίοι ανταγωνίζονται μεταξύ τους για τη χρήση του. Στη συνέχεια παρουσιάζονται αποτελέσματα που αφορούν τη σύγκριση των διαφόρων πρωτοκόλλων πρόσβασης στο μέσο καθώς και των διαφόρων τρόπων εξερεύνησης.

4.1 Παράμετροι προσομοίωσης

Οι παράμετροι της προσομοίωσης παρουσιάζονται στον κατωτέρω πίνακα. Το πλήθος των πρωτευόντων χρηστών είναι ίσο με το πλήθος των διαθέσιμων καναλιών. Προσομοιώθηκαν τρία επίπεδα πυκνότητας του γνωστικού δικτύου, τα οποία εξαρτώνται από το πλήθος των δευτερευόντων χρηστών που ανταγωνίζονται για τη χρήση των καναλιών: $d = |SU|/|C| = \{1,2,3\} = \{\text{Χαμηλή}, \text{Μεσαία}, \text{Υψηλή}\}$.

Πίνακας 1 - Παράμετροι προσομοίωσης.

Κατηγορία	Σύμβολο	Λεπτομέρειες	Τιμή
Αρχικές τιμές	$ SU $	Πλήθος δευτερευόντων χρηστών	{3,6,9}
	$ C $	Πλήθος καναλιών	3
	$ PU $	Πλήθος πρωτευόντων χρηστών	3
	PUL_i	Επίπεδο χρησιμοποίησης καναλιού σε κάθε κανάλι	[0,0.9]
	PER_i	Ποσοστό λανθασμένων πακέτων για τον πράκτορα i	[0,0.3]
	T	Συνολικός χρόνος προσομοίωσης	100s
SU	$t_{DATA,SU}$	Διάρκεια πακέτου δεδομένων	5.44ms
	$t_{CTRL,SU}$	Διάρκεια πακέτου καναλιού ελέγχου	272μs
	T_{CSD}	Καθυστέρηση εναλλαγής καναλιού	2ms
PU	$t_{DATA,PU}$	Διάρκεια πακέτου δεδομένων	5.44ms
		Μέγιστο μέγεθος ουράς αναμονής	5
SMAC, MMAC και SMAC με έλεγχο υπαναχωρήσεων	\tilde{a}	Ρυθμός μάθησης	0.2
	ϵ	Ρυθμός εξερεύνησης	0.2
	β	Κατώφλι τιμής Q για την εναλλαγή καναλιού	1
		Αρχική τιμή Q	1
	Q_{max}	Μέγιστη τιμή Q	20
Πρωτόκολλα SMAC	$t_{D,SMAC}$	Διάρκεια εποχής	200ms
	$t_{D,SMAC-B}$	Διάρκεια εποχής	200ms
MMAC	$t_{D,MMAC}$	Διάρκεια εποχής	250ms

4.2 Πειράματα και αποτελέσματα

Όλα τα αποτελέσματα είναι οι μέσοι όροι των τιμών 50 προσομοιώσεων διάρκειας 100s έκαστη. Για κάθε ομάδα 50 προσομοιώσεων λαμβάνονται αποτελέσματα για τις συνολικές εναλλαγές καναλιού, τα συνολικά πακέτα που στάλθηκαν από τους δευτερεύοντες χρήστες και τα πακέτα που λήφθηκαν με επιτυχία (δηλαδή αυτά για τα οποία ο πομπός έλαβε επιβεβαίωση).

Λόγω των σταθερών επιπέδων χρησιμοποίησης των καναλιών από τους πρωτεύοντες χρήστες και του σταθερού ποσοστού λανθασμένων πακέτων σε κάθε κανάλι, ο κύριος παράγοντας ρύθμισης της ρυθμαπόδοσης κάθε δευτερεύοντος χρήστη είναι ο ανταγωνισμός για τη χρήση καναλιών με άλλους δευτερεύοντες χρήστες. Όσο περισσότεροι χρήστες προσπαθούν να μεταδώσουν σε ένα κανάλι τόσο μικρότερη ατομική ρυθμαπόδοση θα πετυχαίνουν.

4.2.1 Επίπεδα χρήσης καναλιού

Το μέγιστο πλήθος πακέτων που μπορούν να σταλούν από ένα χρήστη μέσω ενός καναλιού κατά τη διάρκεια μιας εποχής είναι ίσο με

$$\pi_{total} = \lfloor \text{χρόνος εποχής} / \text{διάρκεια πακέτου} \rfloor \quad (4.1)$$

όπου με $\lfloor x \rfloor$ συμβολίζεται το ακέραιο μέρος του x .

Αν στο κανάλι i υπάρχει πρωτεύων χρήστης με επίπεδο χρησιμοποίησης καναλιού PUL_i , τότε το πλήθος των πακέτων που μπορούν να σταλούν από ένα δευτερεύοντα χρήστη είναι κατά μέσο όρο:

$$\pi_{SU} = (1 - PUL_i)\pi_{total} \quad (4.2)$$

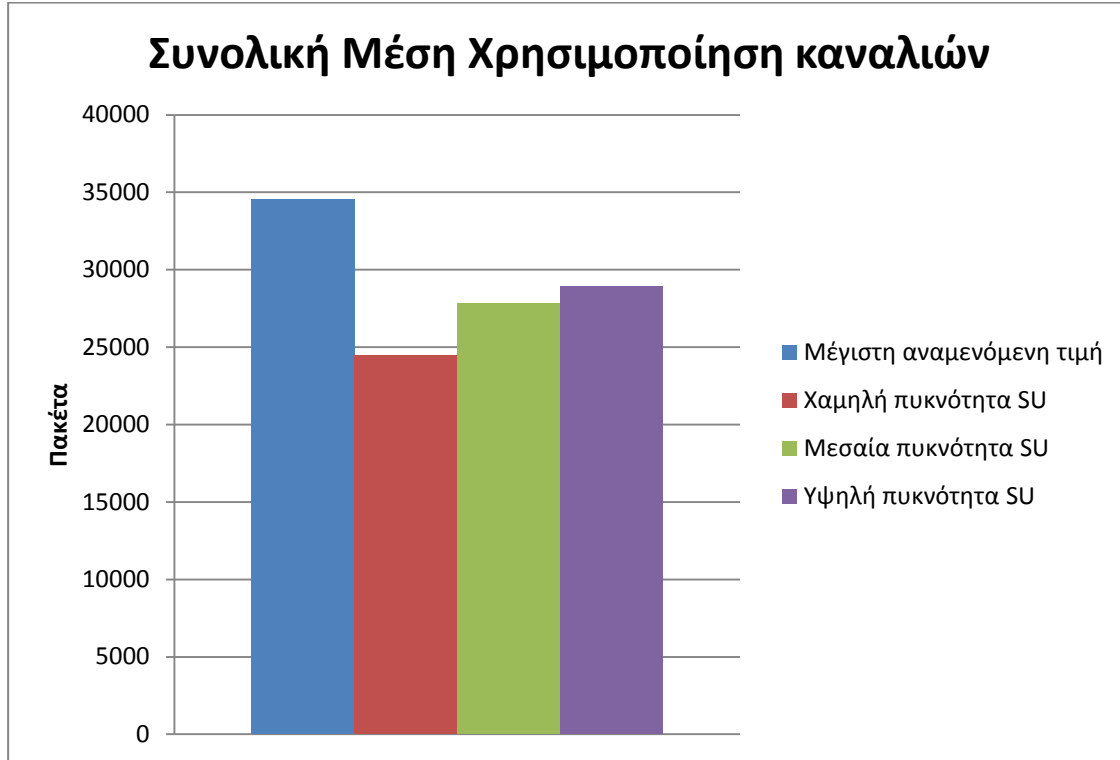
Από τα πακέτα αυτά, λόγω διαλείψεων, σκίασης και παρεμβολών, θα φθάσουν επιτυχώς στο δέκτη περίπου:

$$\pi = (1 - PER_i)\pi_{SU} = (1 - PER_i)(1 - PUL_i) \left\lfloor \frac{\text{χρόνος εποχής}}{\text{διάρκεια πακέτου}} \right\rfloor \quad (4.3)$$

που είναι το μέγιστο αναμενόμενο πλήθος πακέτων που μπορεί να έχει στείλει ένας δευτερεύων χρήστης σε μία εποχή. Στην πραγματικότητα, το μέγιστο πλήθος είναι ελαφρώς μικρότερο του π λόγω καθυστερήσεων στην έναρξη εκπομπής κάθε πακέτου.

Στο διάγραμμα που ακολουθεί φαίνεται το επίπεδο χρησιμοποίησης όλων των καναλιών από τους δευτερεύοντες χρήστες όταν χρησιμοποιούν ενισχυτική μάθηση απλού πράκτορα (πρωτόκολλο SMAC). Το μέσο PUL είναι 0.2 και το μέσο PER σε κάθε κανάλι είναι επίσης 0.2. Το μέγιστο αναμενόμενο πλήθος πακέτων προκύπτει από την τιμή του π

που δίνει η εξίσωση (4.3) πολλαπλασιασμένη με το πλήθος των καναλιών και το πλήθος των εποχών, ενώ οι τιμές των δευτερευόντων χρηστών για τις διάφορες πυκνότητες προκύπτουν από τις μετρούμενες τιμές της προσομοίωσης πολλαπλασιασμένες με το πλήθος των δευτερευόντων χρηστών.



Σχήμα 4-1 - Μέση χρησιμοποίηση καναλιών εκφρασμένη σε αριθμό πακέτων για κανάλια με PER=0.2 και PUL=0.2.

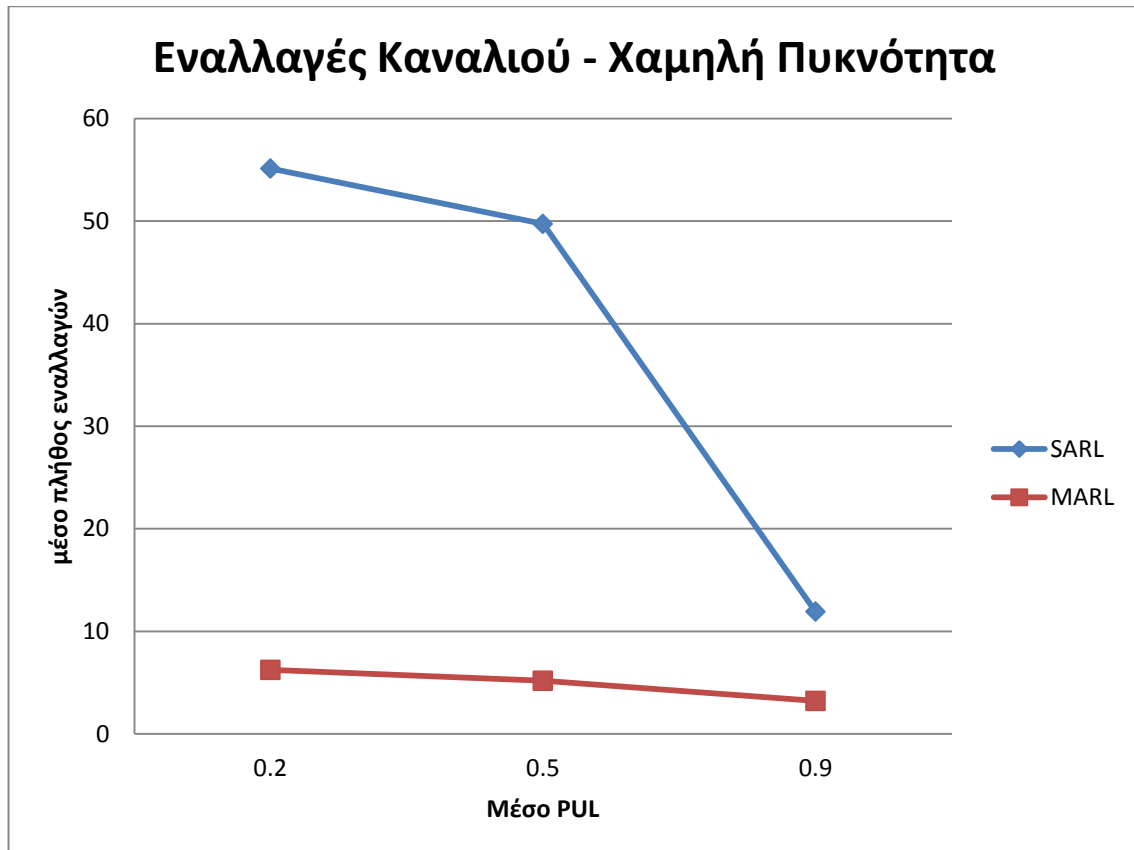
Καίτοι όσο αυξάνει το πλήθος των χρηστών τόσο ελαττώνεται η ατομική τους ρυθμαπόδοση, λόγω του ανταγωνισμού για χρήση του ίδιου φάσματος, η σωρευτική (aggregate) ρυθμαπόδοση των δευτερευόντων χρηστών αυξάνει. Ενδεικτικά για το παράδειγμα του σχήματος 4-1, στη χαμηλή πυκνότητα SU (3 SU) τα κενά φάσματος (white spaces) των διαθέσιμων καναλιών, χρησιμοποιούνται από τους δευτερεύοντες χρήστες κατά 70.7%, στη μεσαία πυκνότητα (6 SU) κατά 80.5% και στη υψηλή πυκνότητα (9 SU) κατά 83.7%.

Αυτό συμβαίνει διότι όσο περισσότεροι χρήστες επιχειρούν να μεταδώσουν σε ένα κανάλι τόσο μεγαλύτερο τμήμα των κενών διαστημάτων, τα οποία αφήνουν οι πρωτεύοντες χρήστες, θα χρησιμοποιείται, λόγω των μικρότερων διαστημάτων υπαναχώρησης. Το διάστημα υπαναχώρησης ενός δευτερεύοντος χρήστη αποτελεί ομοιόμορφα κατανομημένη τυχαία μεταβλητή και η μέση τιμή του ελαχίστου N τυχαίων ομοιόμορφα κατανομημένων μεταβλητών (οι οποίες είναι ανεξάρτητες μεταξύ τους) μειώνεται όσο αυξάνει ο αριθμός τους (βλ. Παράρτημα Α). Έτσι όταν υπάρχουν παραπάνω από ένας δευτερεύοντες χρήστες σε ένα κανάλι η μέση τιμή του ελαχίστου των διαστημάτων

υπαναχώρησης (που καθορίζει πότε θα γίνει η εκπομπή πακέτου) είναι μικρότερη από τη μέση τιμή του διαστήματος υπαναχώρησης του ενός δευτερεύοντος χρήστη.

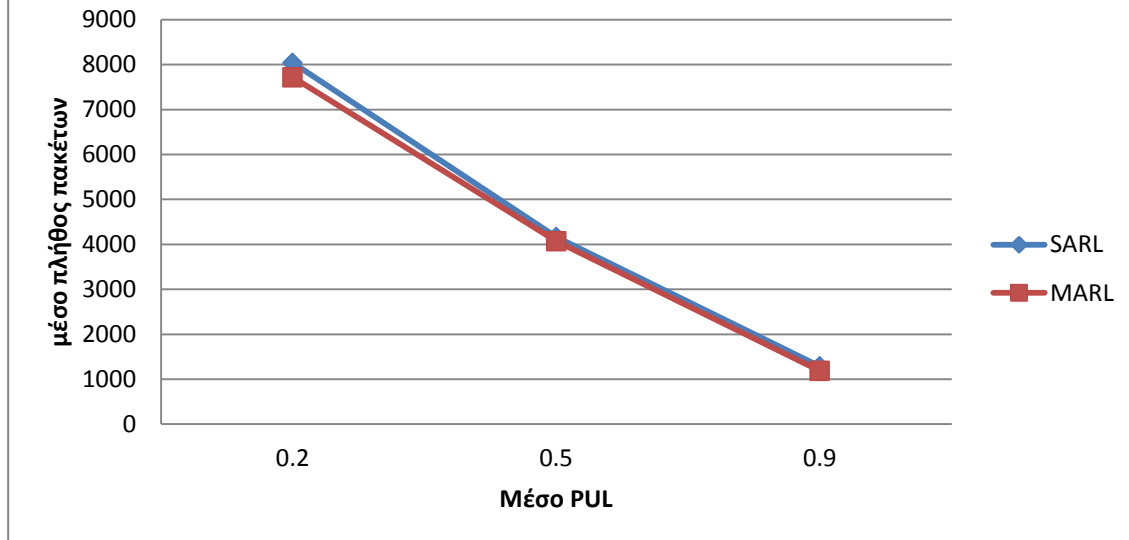
4.2.2 Σύγκριση SARL και MARL

Στη συνέχεια, λαμβάνονται μετρήσεις και συγκρίνονται οι δύο διαφορετικές προσεγγίσεις ενισχυτικής μάθησης που παρουσιάστηκαν στο εδάφιο 3.1.4, η ενισχυτική μάθηση απλού πράκτορα και η ενισχυτική μάθηση πολλών πρακτόρων, μέσω των πρωτοκόλλων SMAC και MMAC, αντίστοιχα. Όλοι οι δευτερεύοντες χρήστες βλέπουν τα κανάλια με μέσο PER ίσο με 0.2 ενώ κάθε φορά οι πρωτεύοντες χρήστες έχουν σταθερό μέσο PUL.



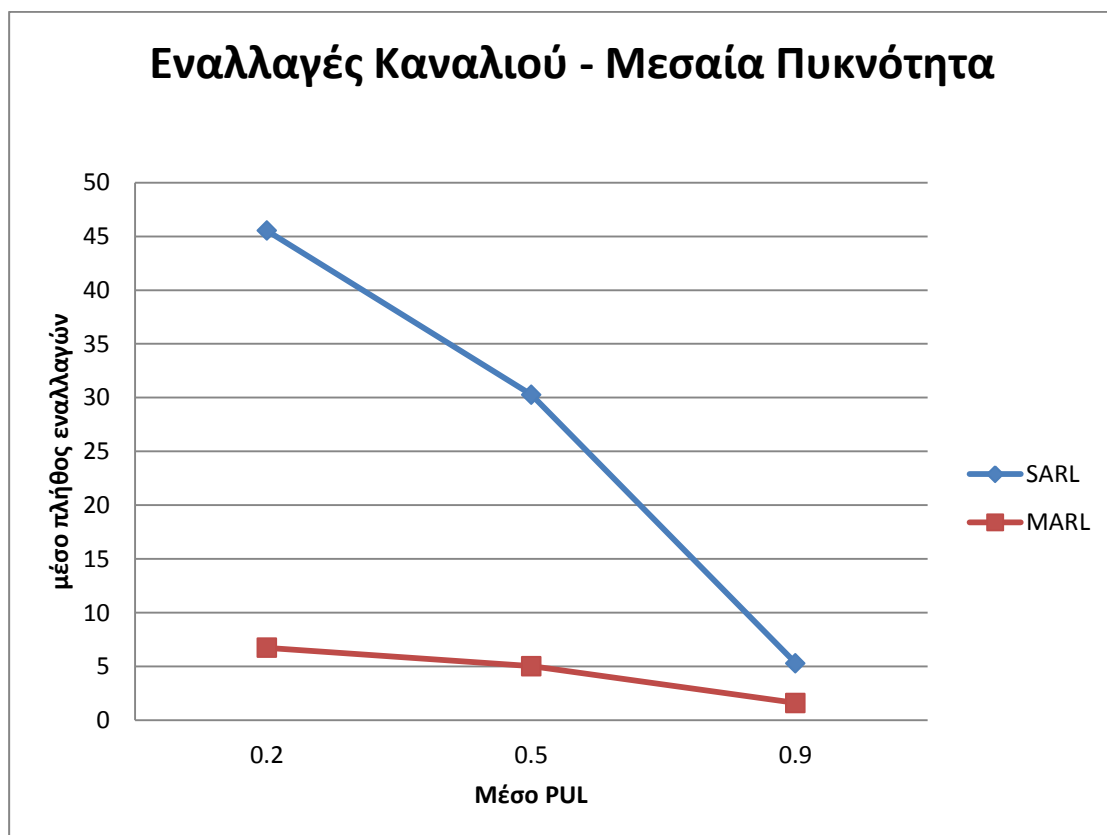
Σχήμα 4-2 - Μέσο πλήθος εναλλαγών καναλιού για πράκτορες SARL και MARL, για κανάλια με PUL=0.2,0.5 και 0.9 και τρεις δευτερεύοντες χρήστες.

Πακέτα που στάλθηκαν επιτυχώς - Χαμηλή πυκνότητα

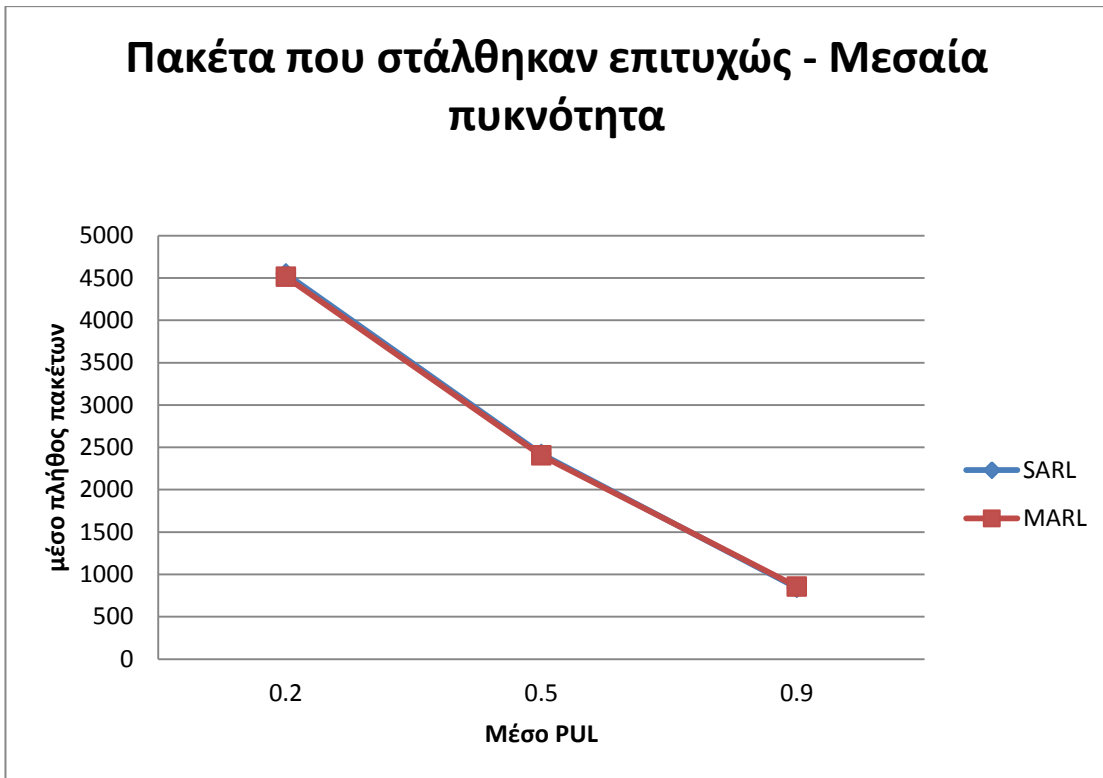


Σχήμα 4-3 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για πράκτορες SARL και MARL, για κανάλια με PUL=0.2,0.5 και 0.9 και τρεις δευτερεύοντες χρήστες.

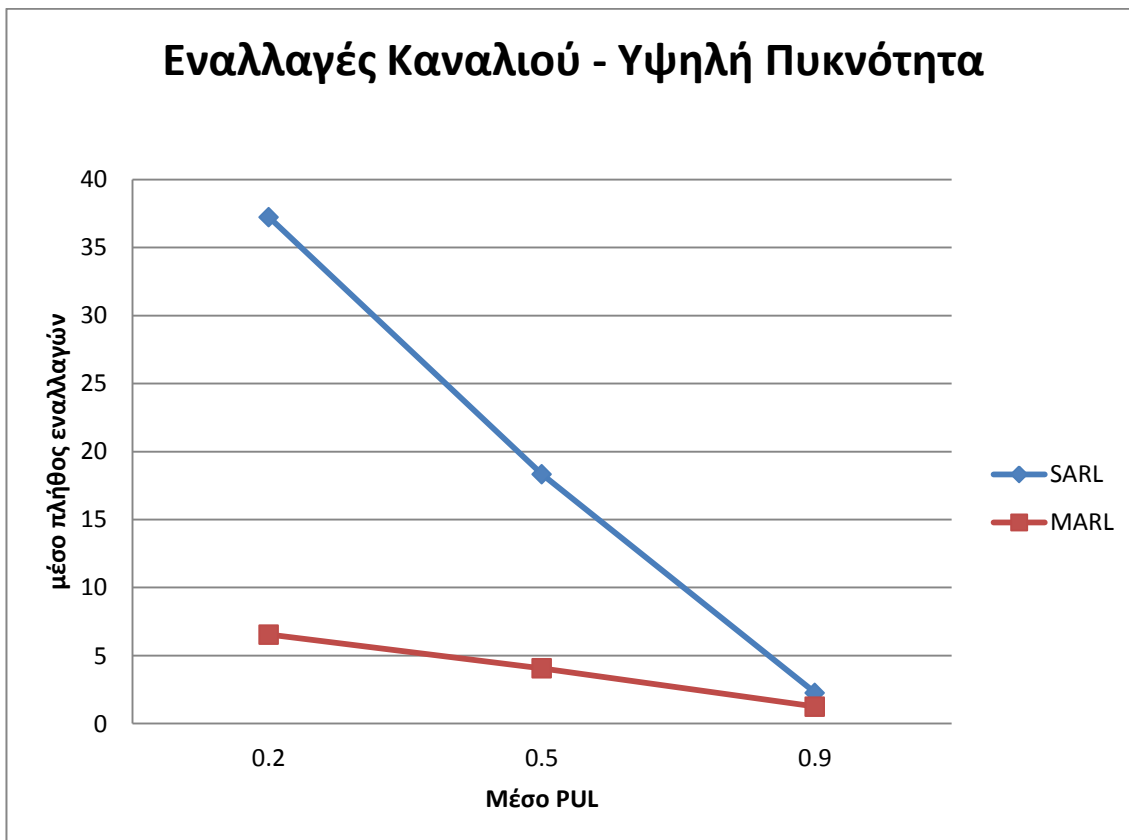
Εναλλαγές Καναλιού - Μεσαία Πυκνότητα



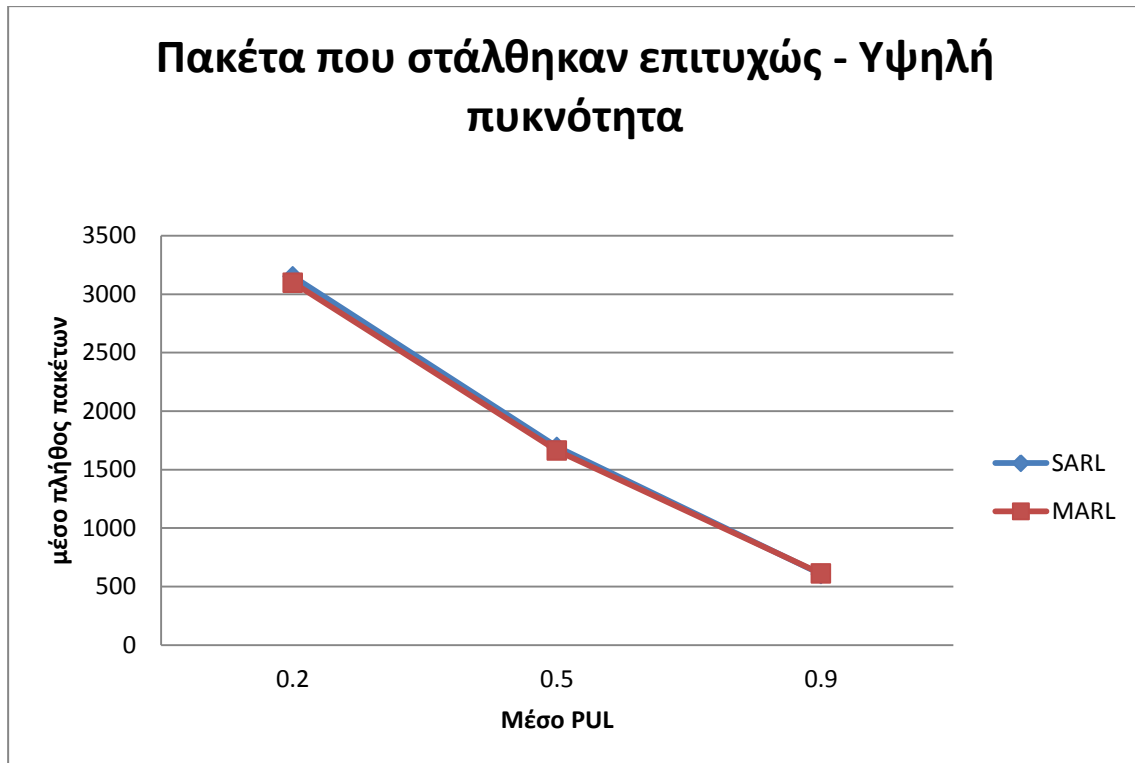
Σχήμα 4-4 - Μέσο πλήθος εναλλαγών καναλιού για πράκτορες SARL και MARL, για κανάλια με PUL=0.2,0.5 και 0.9 και έξι δευτερεύοντες χρήστες.



Σχήμα 4-5 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για πράκτορες SARL και MARL, για κανάλια με PUL=0.2,0.5 και 0.9 και έξι δευτερεύοντες χρήστες.

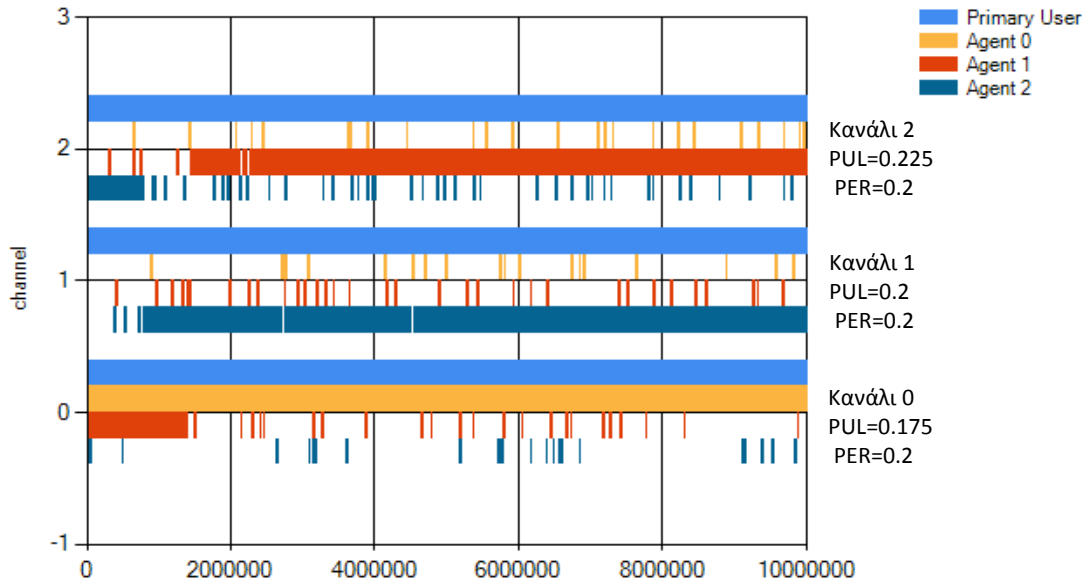


Σχήμα 4-6 - Μέσο πλήθος εναλλαγών καναλιού για πράκτορες SARL και MARL, για κανάλια με PUL=0.2,0.5 και 0.9 και εννιά δευτερεύοντες χρήστες.

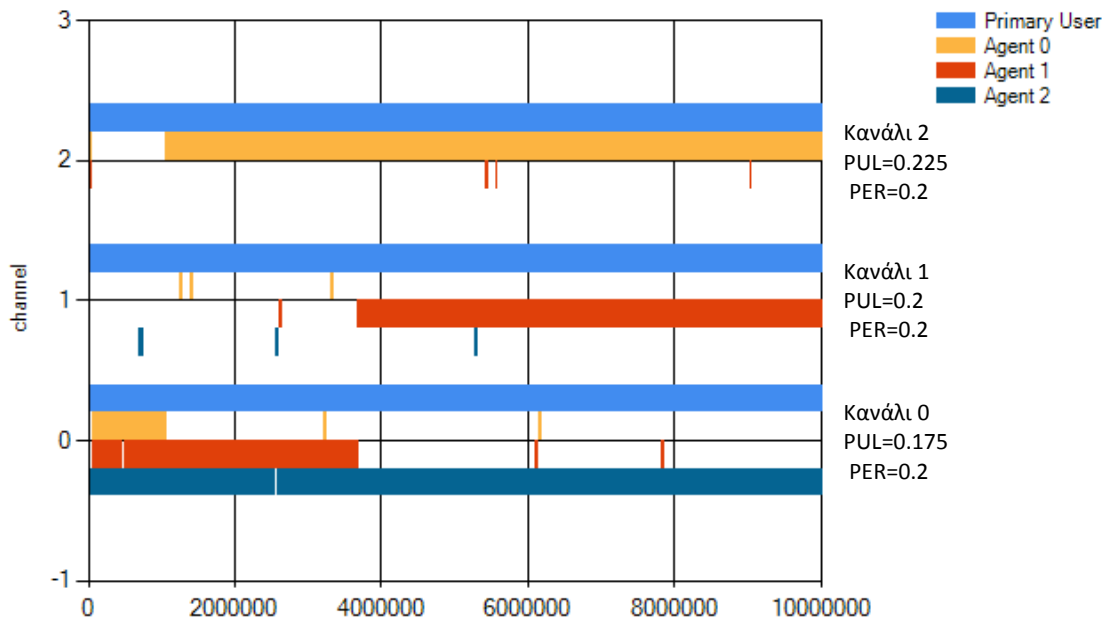


Σχήμα 4-7 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για πράκτορες SARL και MARL, για κανάλια με PUL=0.2,0.5 και 0.9 και εννιά δευτερεύοντες χρήστες.

Οι εναλλαγές καναλιού, τόσο για ενισχυτική μάθηση απλού πράκτορα όσο για ενισχυτική μάθηση πολλαπλών πρακτόρων, φθίνουν όσο αυξάνεται η χρήση του καναλιού από τους πρωτεύοντες χρήστες, για όλες τις πυκνότητες δευτερευόντων χρηστών. Αυτό συμβαίνει διότι, σε υψηλότερα επίπεδα χρησιμοποίησης καναλιού από τους πρωτεύοντες χρήστες, η χαμηλότερη ρυθμαπόδοση των δευτερευόντων (όπως φαίνεται στα σχήματα 4-3, 4-5 και 4-7) οδηγεί σε μικρότερες ανταμοιβές, σύμφωνα με τον τύπο 3.3. Αυτές οι μικρότερες ανταμοιβές με τη σειρά τους οδηγούν σε μικρότερες αλλαγές των τιμών Q και, επομένως, μικρότερες διαφορές μεταξύ των τιμών Q που διατηρεί ένας SU για κάθε κανάλι. Στην περίπτωση της ενισχυτικής μάθησης πολλαπλών πρακτόρων, η γνώση των ανταμοιβών των υπολοίπων SU για κάθε κανάλι, επιτρέπει σε κάθε δευτερεύοντα χρήστη να διαλέγει με περισσότερο επιτυχή τρόπο το κανάλι του, μειώνοντας έτσι το πλήθος των εναλλαγών καναλιού σε σχέση με την ενισχυτική μάθηση απλού πράκτορα (σχήματα 4-2, 4-4 και 4-6). Αυτή η μείωση του πλήθους εναλλαγών καναλιού οδηγεί σε μεγαλύτερη ευστάθεια του συστήματος. Παρά την επιτυχεστέρα επιλογή καναλιών από τους SU που χρησιμοποιούν ενισχυτική μάθηση πολλαπλών πρακτόρων, η ρυθμαπόδοση παραμένει στα ίδια επίπεδα με τους SU που χρησιμοποιούν ενισχυτική μάθηση απλού πράκτορα. Αυτό συμβαίνει διότι στην ενισχυτική μάθηση πολλαπλών πρακτόρων ένα σημαντικό τμήμα κάθε εποχής χρησιμοποιείται για τη μετάδοση των μηνυμάτων ανταμοιβής στους υπόλοιπους SU. Επιπλέον, η προσέγγιση των πολλαπλών πρακτόρων έχει το αυξημένο κόστος της επικοινωνίας μεταξύ πρακτόρων. Η ατομική συμπεριφορά διαφορετικών πρακτόρων φαίνεται στα σχήματα 4-8 και 4-9.



Σχήμα 4-8 - Διάγραμμα χρήσης καναλιών για δευτερεύοντες χρήστες που χρησιμοποιούν ενισχυτική μάθηση μονού πράκτορα (SARL). Μέσο PUL 0.2, Μέσο PER 0.2.



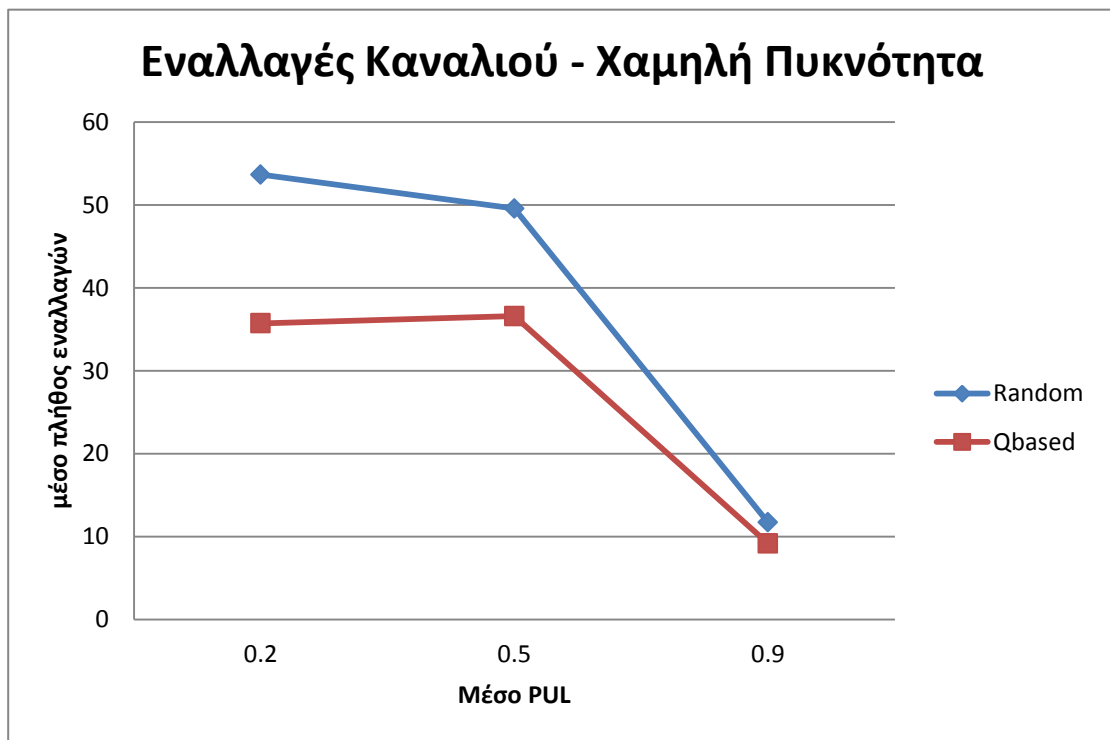
Σχήμα 4-9 - Διάγραμμα χρήσης καναλιών από δευτερεύοντες χρήστες που χρησιμοποιούν ενισχυτική μάθηση πολλών πρακτόρων (MARL). Μέσο PUL 0.2, Μέσο PER 0.2.

Από τα δύο προηγούμενα διαγράμματα γίνεται φανερό ότι οι ανεξάρτητοι πράκτορες SARL αν και αποφασίζουν ταχύτερα σε ποιο κανάλι θα εκπέμψουν (κατά κύριο λόγο) σε σχέση με τους πράκτορες MARL (που έχουν ανάγκη για περισσότερη ανταλλαγή πληροφοριών μεταξύ τους), εμφανίζονται λιγότερο σταθεροί στις επιλογές τους και πραγματοποιούν περισσότερες εναλλαγές και δοκιμές εκπομπής και στα άλλα κανάλια. Οι

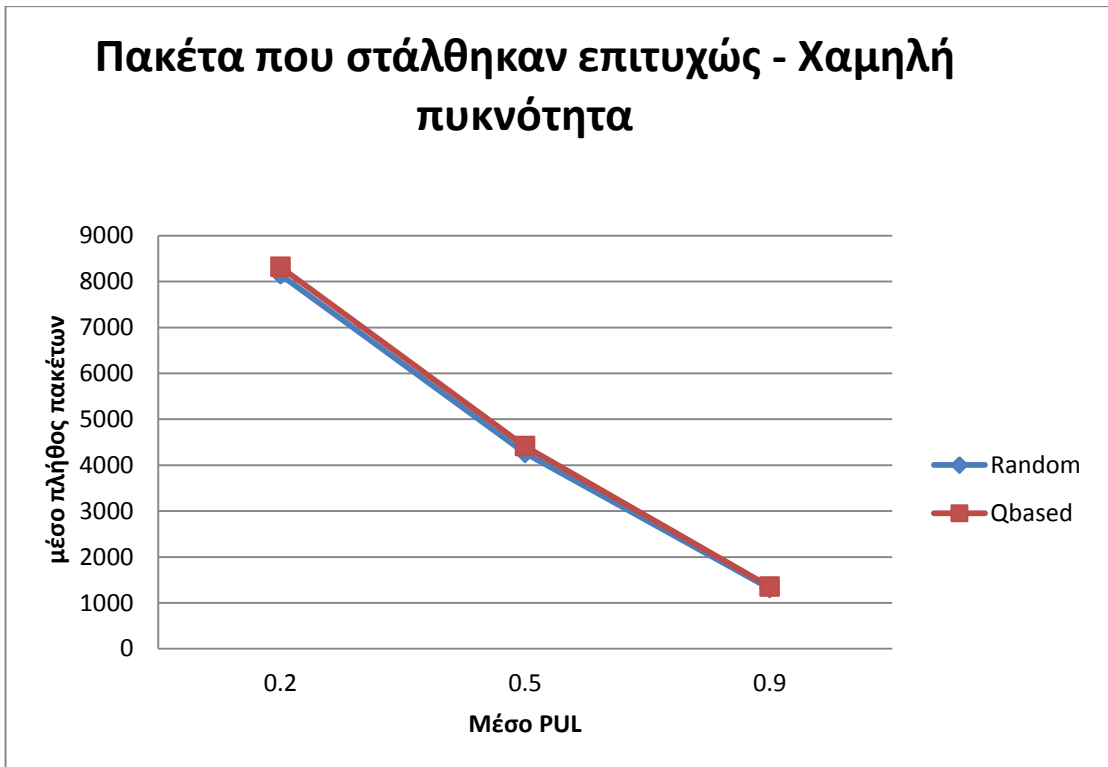
συχνές εναλλαγές οφείλονται στις αυξομειώσεις των ανταμοιβών και τις αντίστοιχες αυξομειώσεις στις τιμές Q των καναλιών. Οι πράκτορες MARL δεν είναι τόσο ευαίσθητοι στις αυξομειώσεις αυτές καθώς διαθέτουν γνώση των ανταμοιβών και των υπολοίπων πρακτόρων, αποφεύγοντας έτσι άσκοπες εναλλαγές καναλιού.

4.2.3 Σύγκριση διαφορετικών μεθόδων εξερεύνησης(exploration)

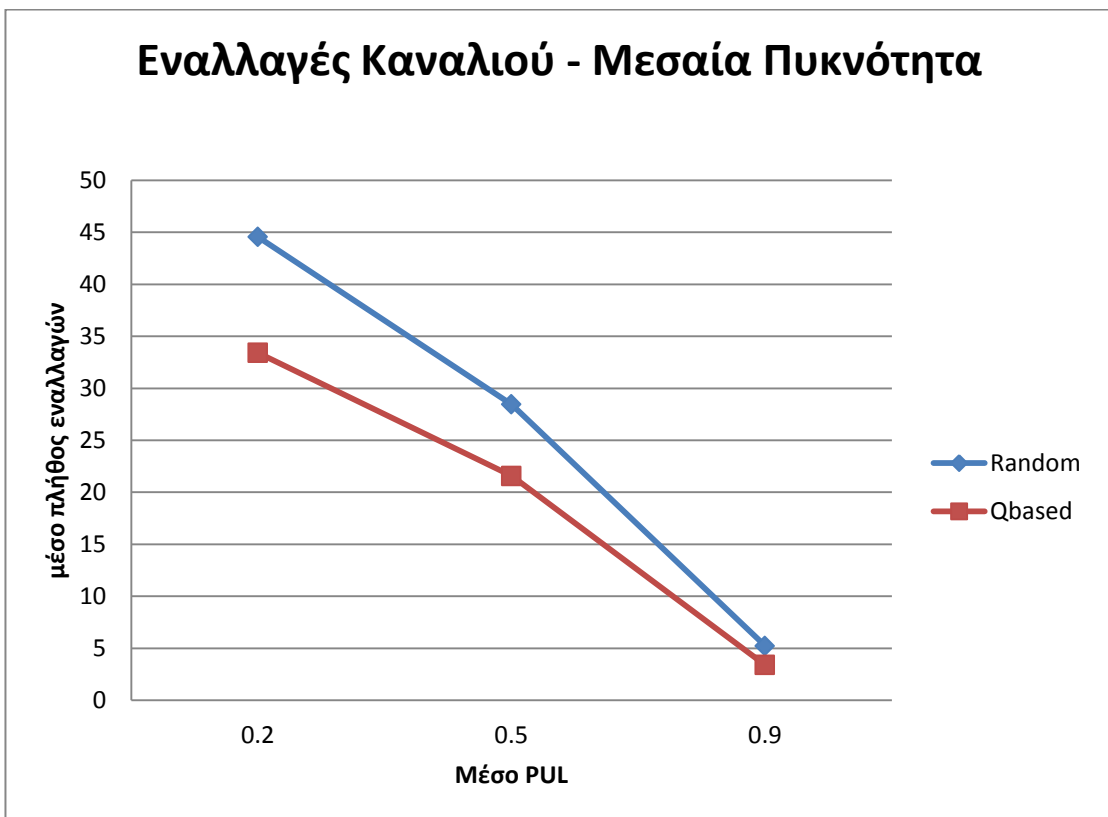
Συγκρίνονται οι δύο μέθοδοι εξερεύνησης που παρουσιάστηκαν στο εδάφιο 3.1.5 για ανεξάρτητους πράκτορες SARL με σταθερό μέσο PER = 0.2. Η τυχαία εξερεύνηση (που προτείνεται στην εργασία [37]) απεικονίζεται με τη μπλε γραμμή, ενώ η εξερεύνηση με πιθανότητες που βασίζονται στις τιμές Q σύμφωνα με τη σχέση 3.8 (που προτείνεται στην παρούσα εργασία) με την κόκκινη.



Σχήμα 4-10 – Μέσο πλήθος εναλλαγών καναλιού για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασισμένη στις τιμές Q, για κανάλια με PUL=0.2,0.5 και 0.9 και τρεις δευτερεύοντες χρήστες.



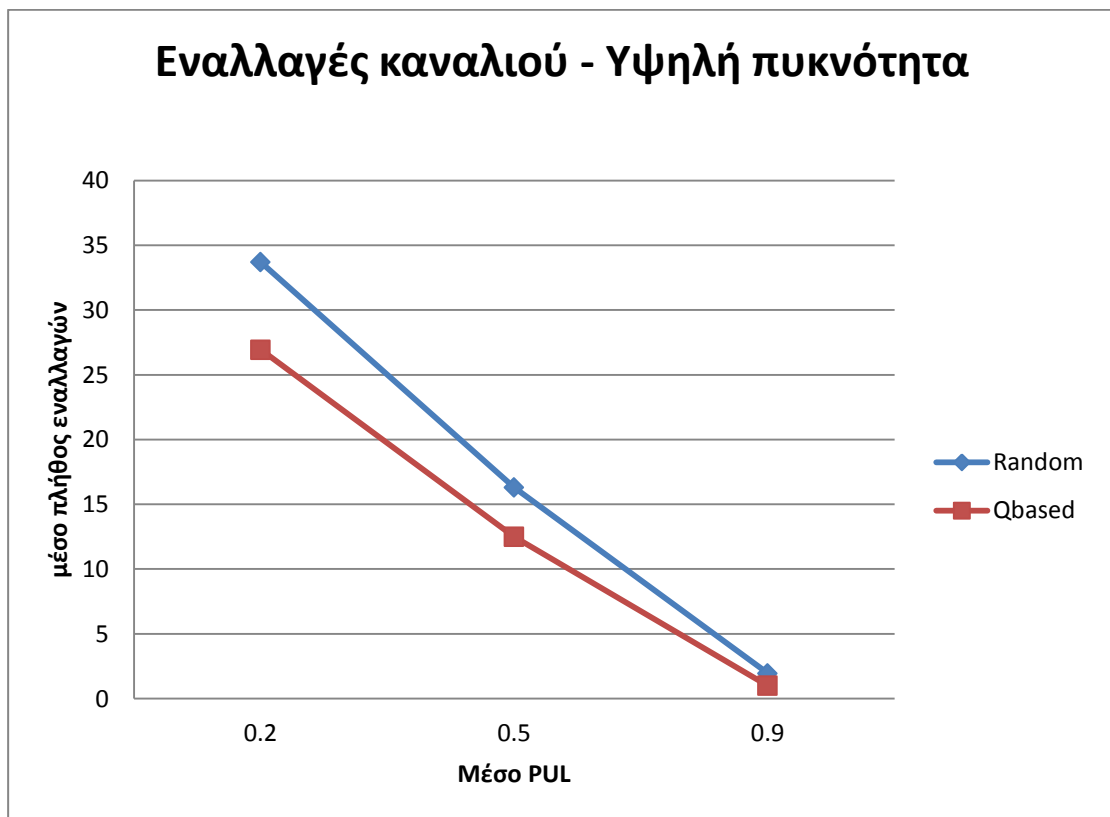
Σχήμα 4-11 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασισόμενη στις τιμές Q, για κανάλια με PUL=0.2,0.5 και 0.9 και τρεις δευτερεύοντες χρήστες.



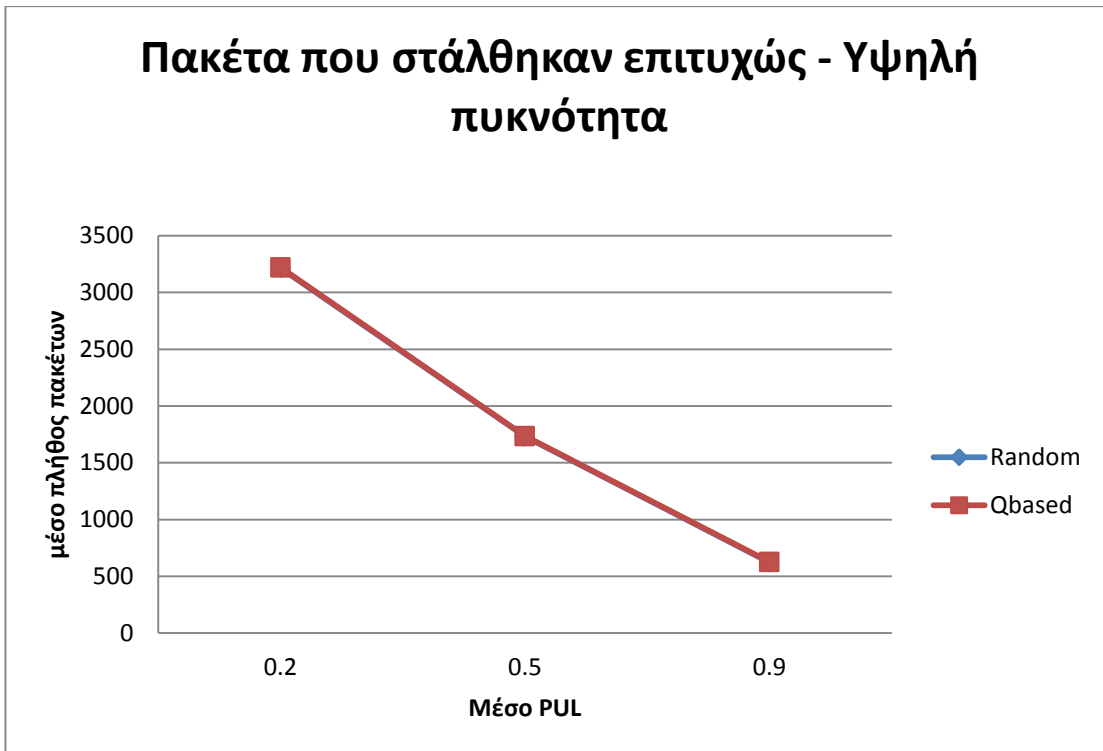
Σχήμα 4-12 - Μέσο πλήθος εναλλαγών καναλιού για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασισόμενη στις τιμές Q, για κανάλια με PUL=0.2,0.5 και 0.9 και έξι δευτερεύοντες χρήστες.



Σχήμα 4-13 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασιζόμενη στις τιμές Q, για κανάλια με PUL=0.2,0.5 και 0.9 και έξι δευτερεύοντες χρήστες.

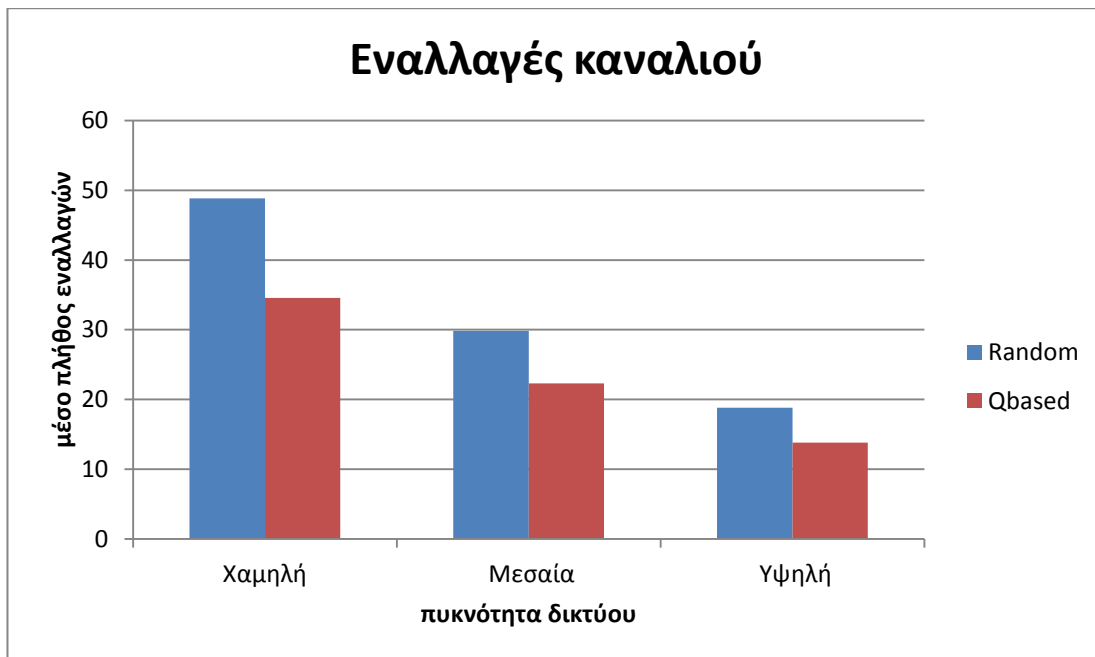


Σχήμα 4-14 - Μέσο πλήθος εναλλαγών καναλιού για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασιζόμενη στις τιμές Q, για κανάλια με PUL=0.2,0.5 και 0.9 και εννιά δευτερεύοντες χρήστες.

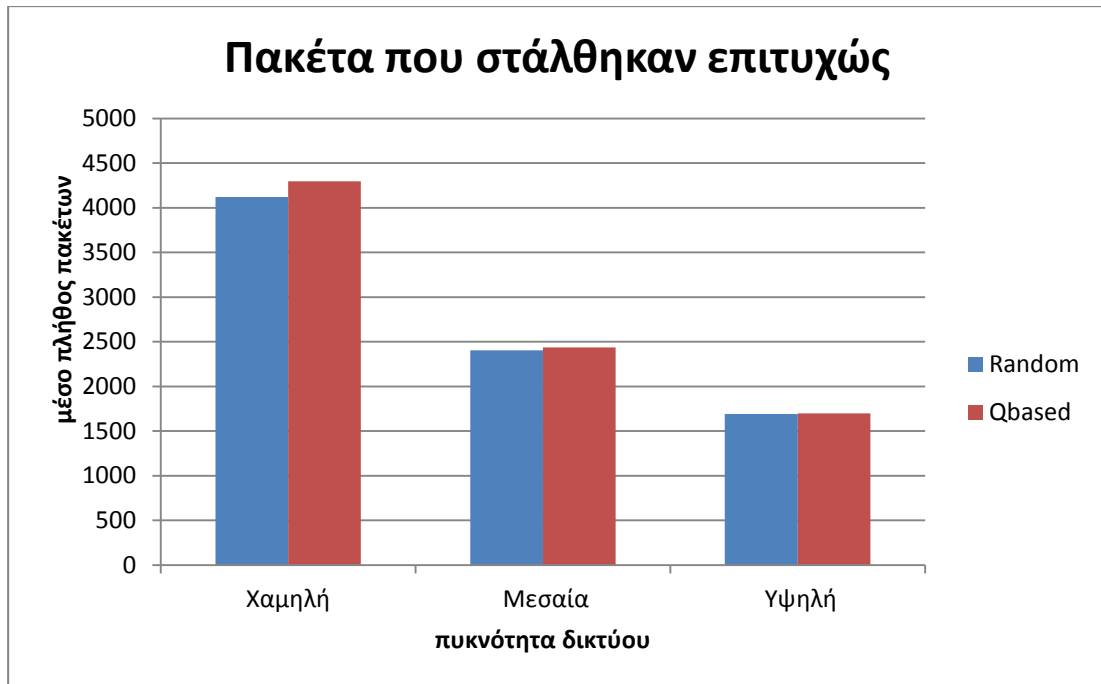


Σχήμα 4-15 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασισμένη στις τιμές Q, για κανάλια με PUL=0.2,0.5 και 0.9 και εννιά δευτερεύοντες χρήστες.

Στις προσομοιώσεις που ακολουθούν (σχήματα 4-16 και 4-17), το PUL κάθε πρωτεύοντος χρήστη διατηρείται ίσο με 0.5 ενώ κάθε κανάλι εμφανίζει διαφορετικό PER στους δευτερεύοντες χρήστες. Το πρώτο εμφανίζει PER ίσο με 0.1, το δεύτερο ίσο με 0.2 και το τρίτο ίσο με 0.3



Σχήμα 4-16 - Μέσο πλήθος εναλλαγών καναλιού για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασισμένη στις τιμές Q, για κανάλια με PUL=0.1,0.2 και 0.3 και διαφορετικές πυκνότητες δικτύου.

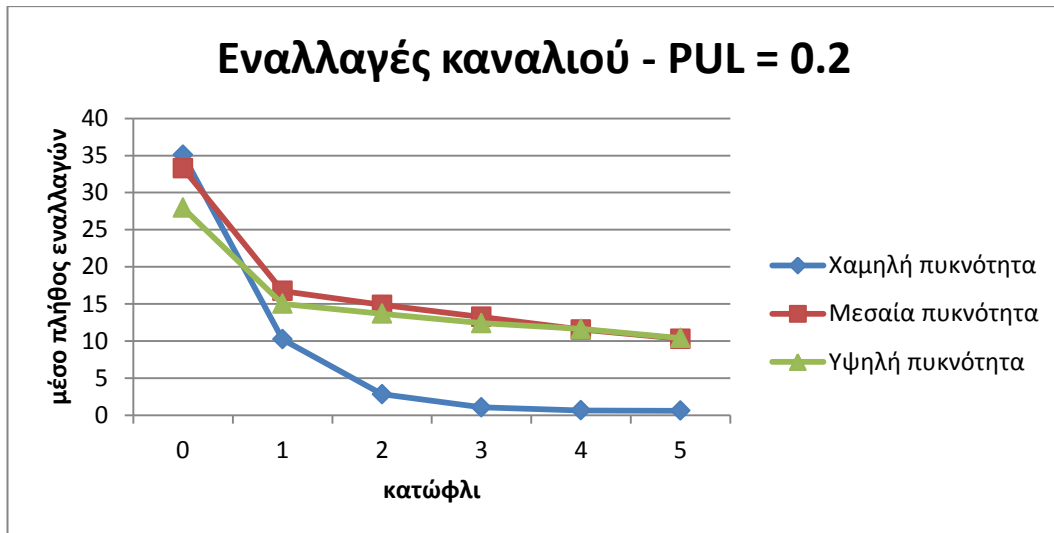


Σχήμα 4-17 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για πράκτορες SARL με τυχαία εξερεύνηση και εξερεύνηση βασισμένη στις τιμές Q, για κανάλια με PUL=0.1,0.2 και 0.3 και διαφορετικές πυκνότητες δικτύου.

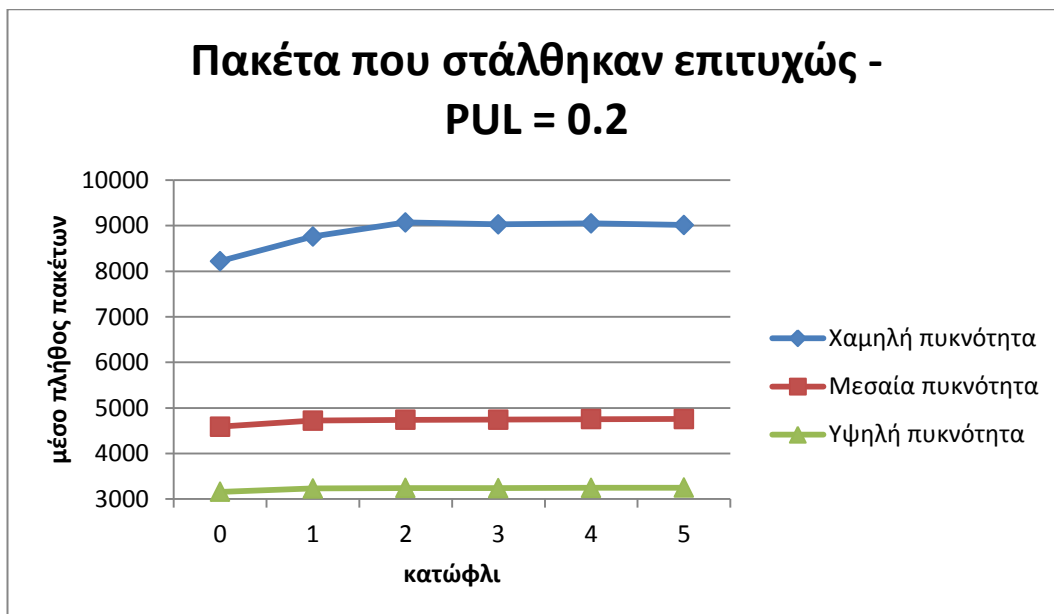
Από όλα τα προηγούμενα σχήματα (4-10 έως 4-17) προκύπτει ότι η χρήση εξερεύνησης με πιθανότητες οι οποίες βασίζονται στις τιμές Q επιτυγχάνει μικρότερο μέσο πλήθος εναλλαγών καναλιού ενώ η μέση ρυθμαπόδοση του συστήματος διατηρείται στα ίδια επίπεδα με την τυχαία εξερεύνηση. Στις επόμενες προσομοιώσεις χρησιμοποιείται η εξερεύνηση με πιθανότητες οι οποίες βασίζονται στις τιμές Q.

4.2.4 Σύγκριση SMAC και SMAC με έλεγχο υπαναχωρήσεων

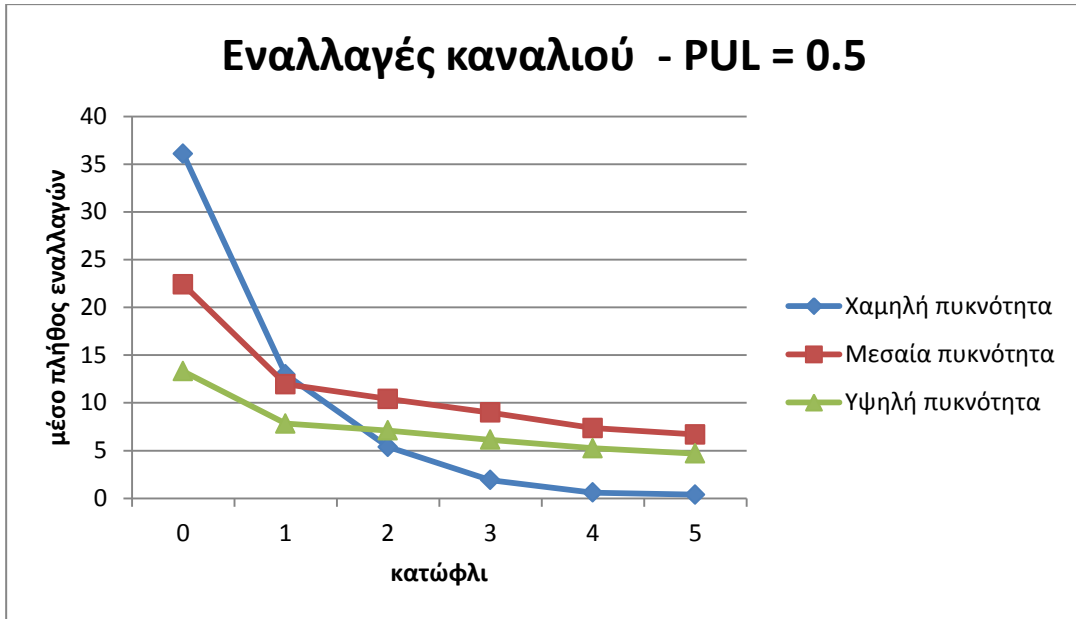
Για πράκτορες SARL συγκρίνονται τα δύο πρωτόκολλα πρόσβασης στο μέσο (MAC) που παρουσιάστηκαν στο εδάφιο 3.2 Στο πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων δοκιμάζονται διαφορετικές τιμές κατωφλίου του αριθμού των υπαναχωρήσεων. Στις ακόλουθες προσομοιώσεις το μέσο PER παραμένει σταθερό και ίσο με 0.2 ενώ οι πρωτεύοντες χρήστες έχουν ίδιο μέσο PUL (0.2, 0.5 και 0.9). Το κατώφλι 0 αναφέρεται στο κανονικό πρωτόκολλο SMAC, ενώ στις υπόλοιπες τιμές χρησιμοποιείται το πρωτόκολλο SMAC με έλεγχο για το πλήθος των υπαναχωρήσεων.



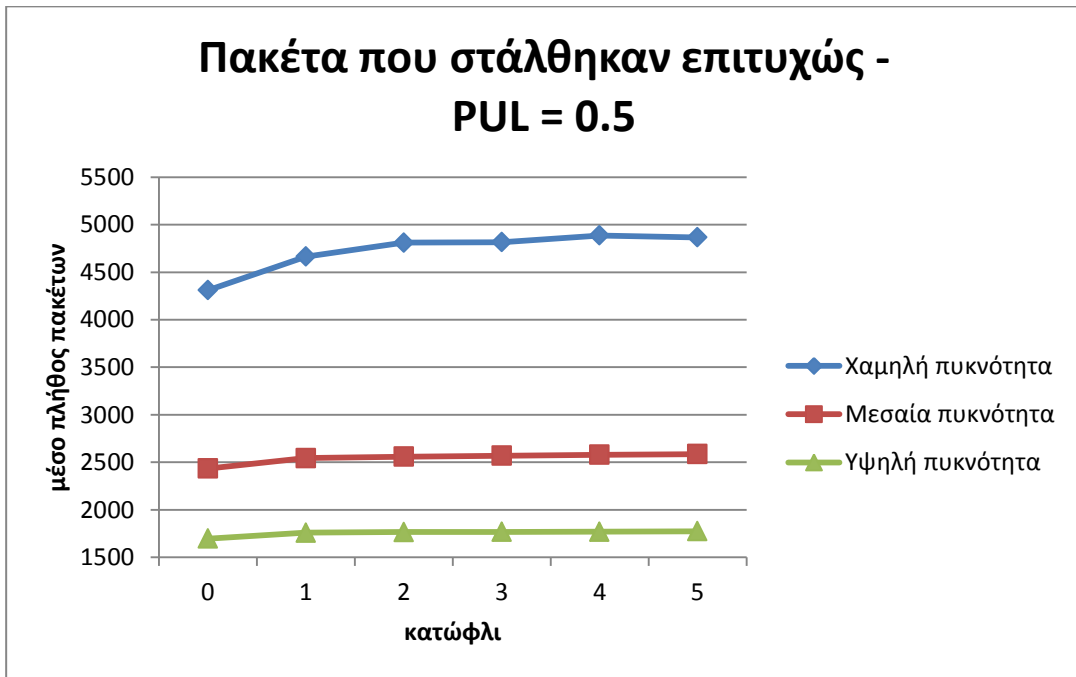
Σχήμα 4-18 - Μέσο πλήθος εναλλαγών καναλιού για χρήστες που χρησιμοποιούν SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με μέσο PUL = 0.2, PER = 0.2, για διάφορες τιμές του κατωφλίου και διάφορες πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.



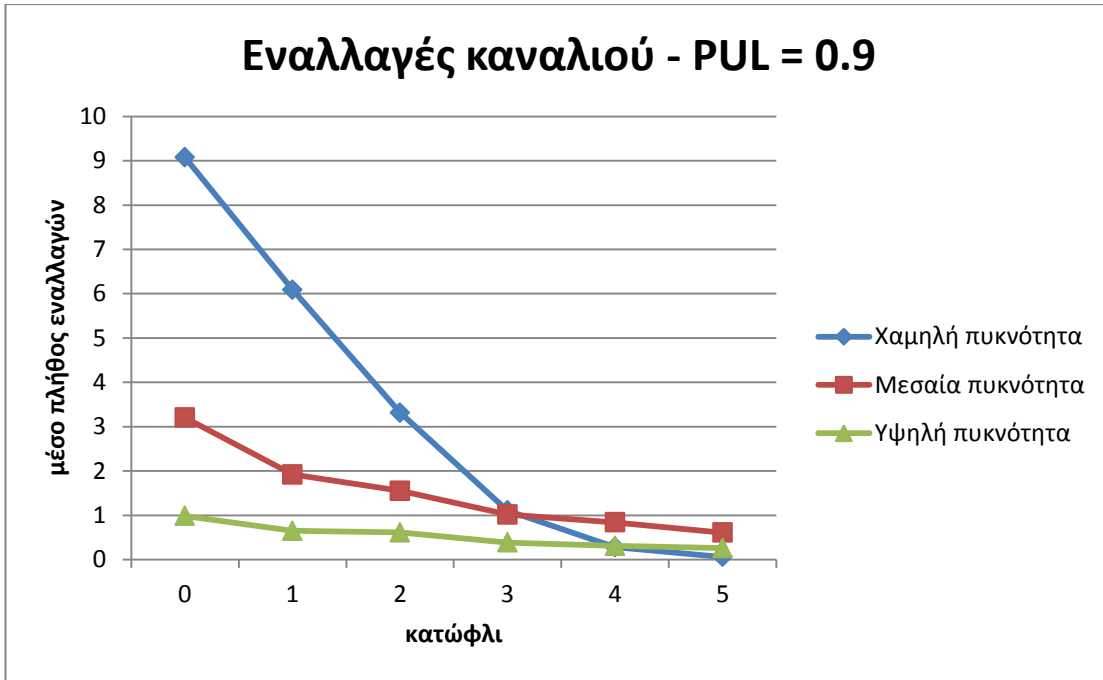
Σχήμα 4-19 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για χρήστες που χρησιμοποιούν SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με μέσο PUL = 0.2, PER = 0.2, για διάφορες τιμές του κατωφλίου και διάφορες πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.



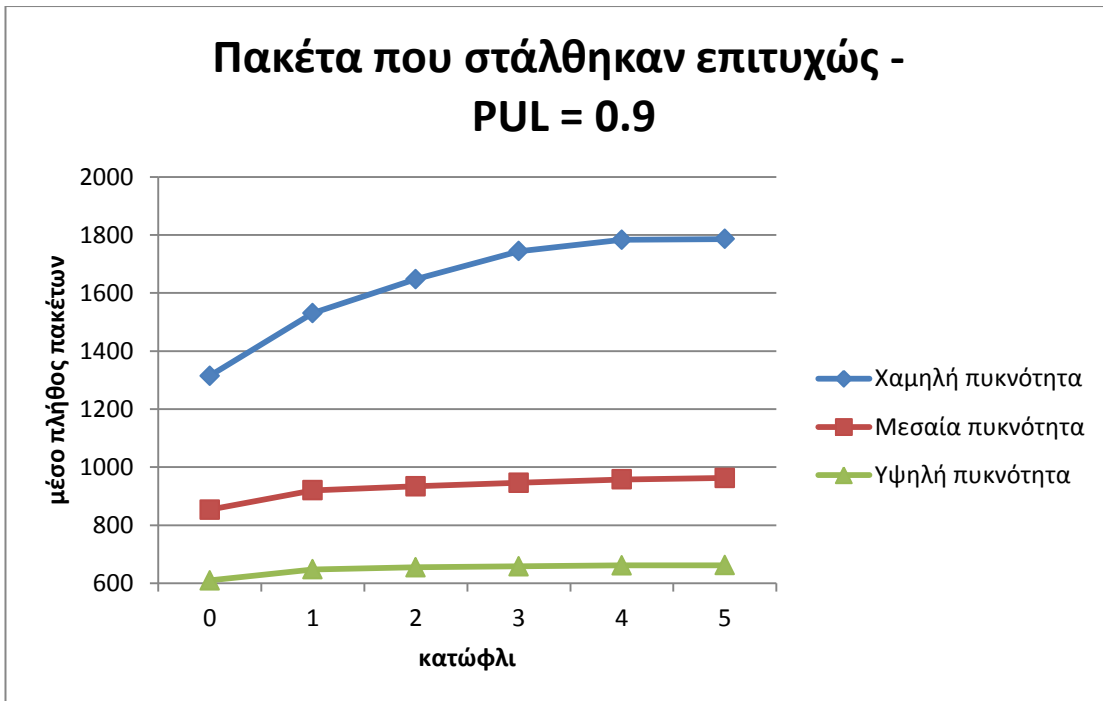
Σχήμα 4-20 - Μέσο πλήθος εναλλαγών καναλιού για χρήστες που χρησιμοποιούν SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με μέσο PUL = 0.5, PER = 0.2, για διάφορες τιμές του κατωφλίου και διάφορες πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.



Σχήμα 4-21 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για χρήστες που χρησιμοποιούν SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με μέσο PUL = 0.5, PER = 0.2, για διάφορες τιμές του κατωφλίου και διάφορες πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.

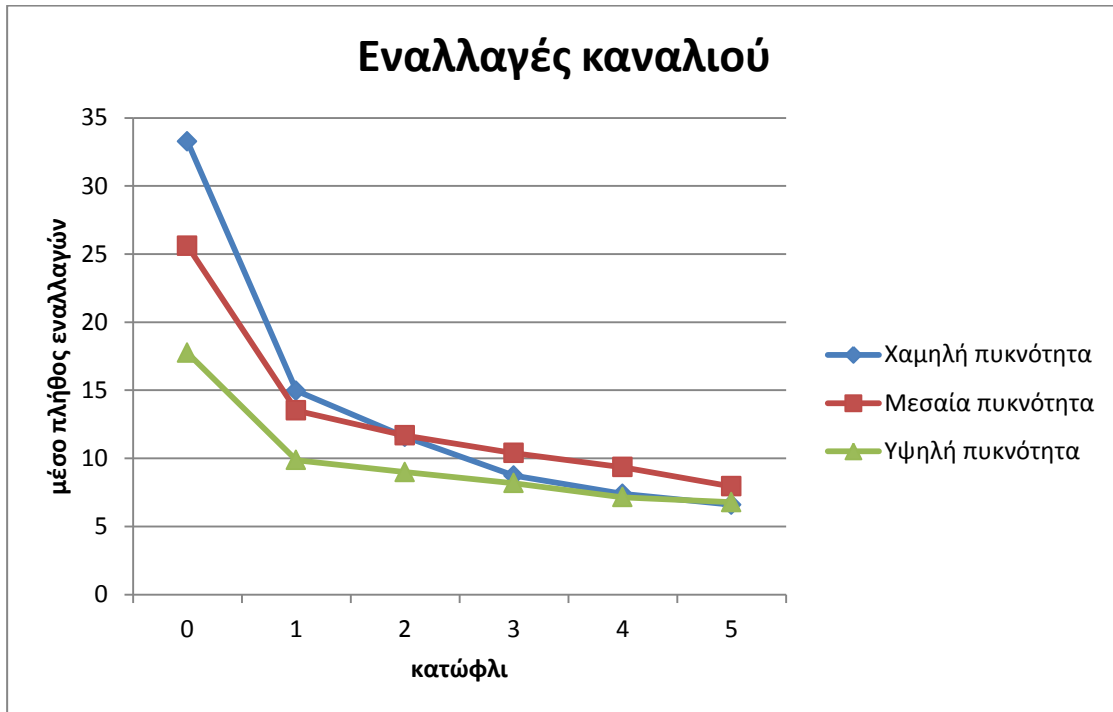


Σχήμα 4-22 - Μέσο πλήθος εναλλαγών καναλιού για χρήστες που χρησιμοποιούν SMAC με έλεγχο υπαναχωρήσεων , για κανάλια με μέσο PUL = 0.9, PER = 0.2, για διάφορες τιμές του κατωφλίου και διάφορες πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.

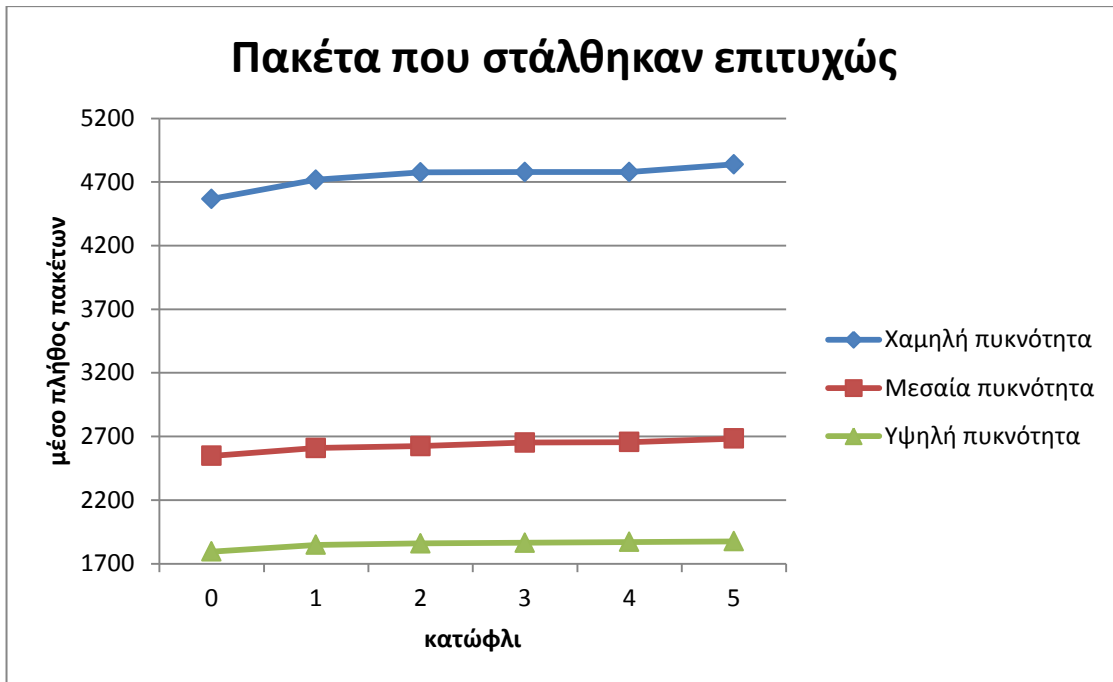


Σχήμα 4-23 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για χρήστες που χρησιμοποιούν SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με μέσο PUL = 0.9, PER = 0.2, για διάφορες τιμές του κατωφλίου και διάφορες πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.

Στις προσομοιώσεις που ακολουθούν (σχήματα 4-24 και 4-25) το μέσο PER παραμένει σταθερό και ίσο με 0.2 αλλά κάθε κανάλι έχει διαφορετικό επίπεδο χρησιμοποίησης από τον αντίστοιχο πρωτεύοντα χρήστη, με το πρώτο να έχει PUL =0.2, το δεύτερο 0.5 και το τρίτο 0.9:

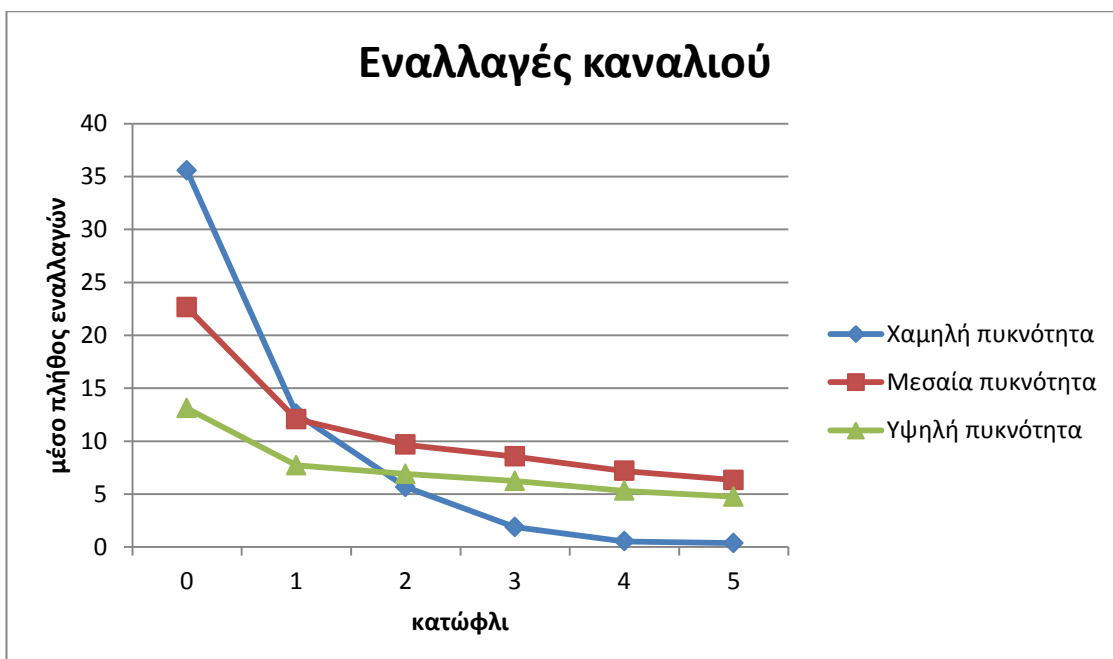


Σχήμα 4-24 - Μέσο πλήθος εναλλαγών καναλιού για χρήστες με πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με διαφορετικά PUL και μέσο PER=0.2, για διαφορετικές τιμές κατωφλίου και διαφορετικές πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.

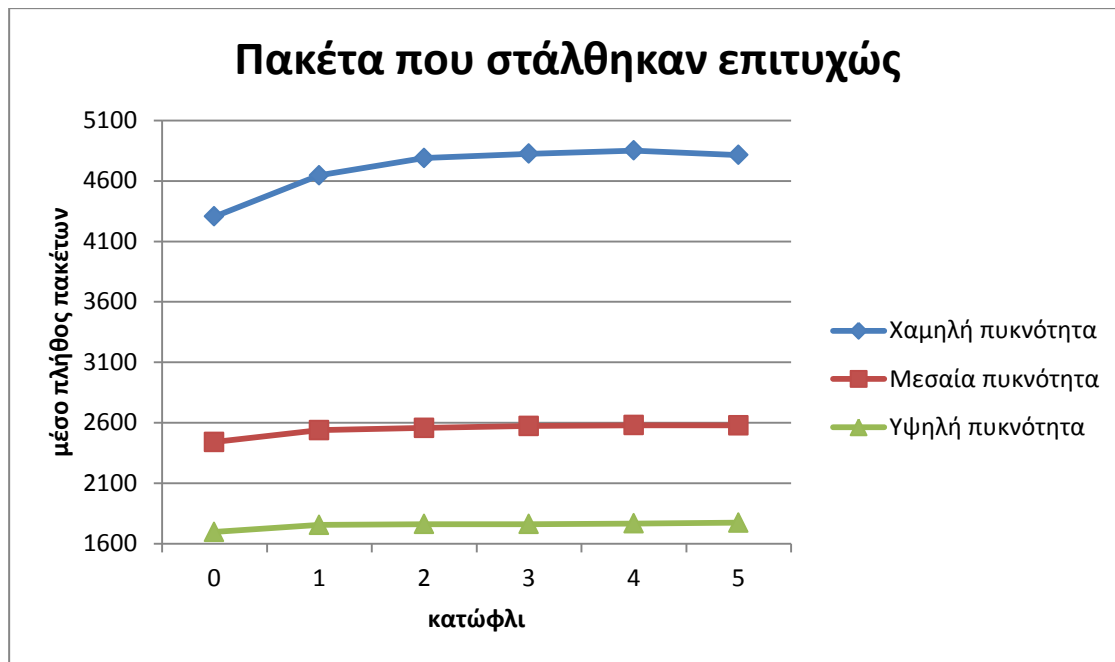


Σχήμα 4-25 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για χρήστες με πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με διαφορετικά PUL και μέσο PER=0.2, για διαφορετικές τιμές κατωφλίου και διαφορετικές πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.

Παρόμοια είναι τα αποτελέσματα όταν όλοι οι πρωτεύοντες χρήστες έχουν $PUL = 0.5$ και κάθε κανάλι παρουσιάζει διαφορετικό PER στους δευτερεύοντες χρήστες. Το πρώτο κανάλι παρουσιάζει $PER = 0.1$ το δεύτερο ίσο με 0.2 και το τρίτο ίσο με 0.3

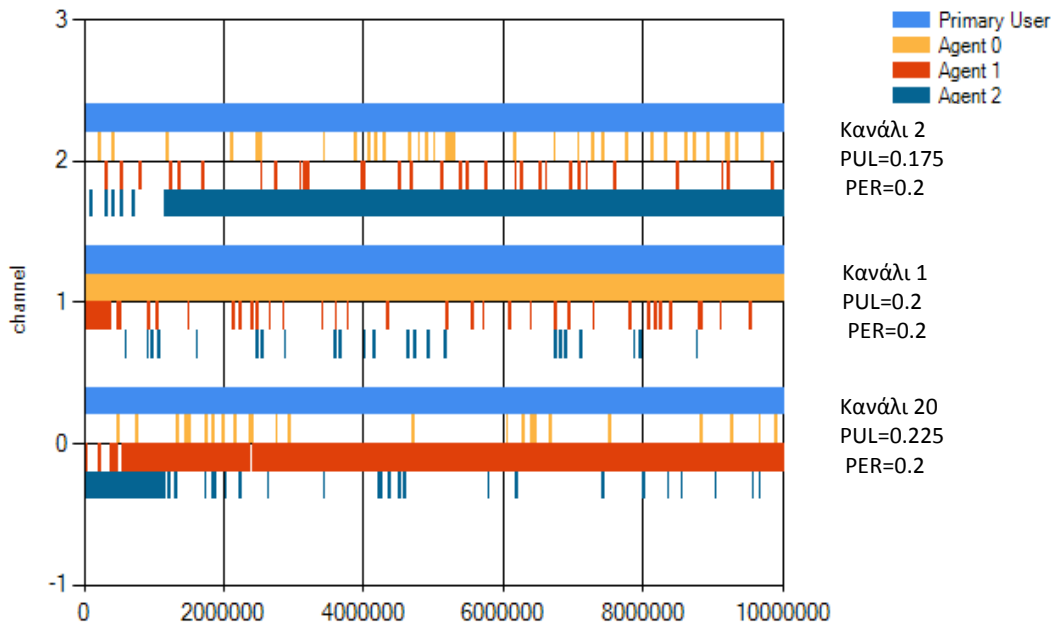


Σχήμα 4-26 - Μέσο πλήθος εναλλαγών καναλιού για χρήστες με πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με διαφορετικά PER και $PUL = 0.5$, για διαφορετικές τιμές κατωφλίου και διαφορετικές πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.

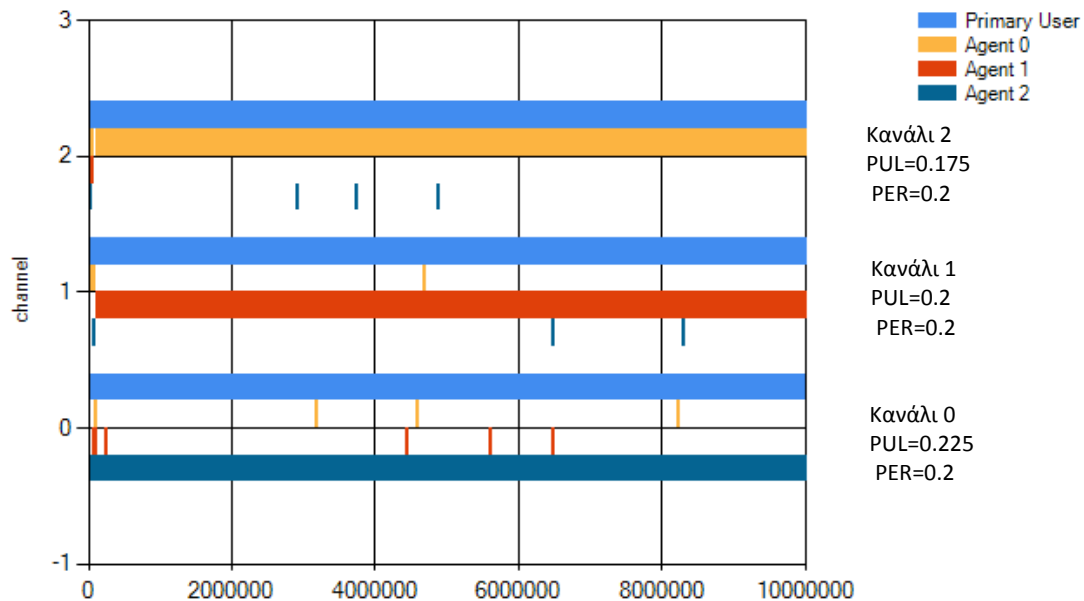


Σχήμα 4-27 - Μέσο πλήθος πακέτων που στάλθηκαν με επιτυχία για χρήστες με πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων, για κανάλια με διαφορετικά PER και PUL = 0.5, για διαφορετικές τιμές κατώφλιου και διαφορετικές πυκνότητες δικτύου. Το κατώφλι 0 αναφέρεται σε πρωτόκολλο SMAC χωρίς έλεγχο υπαναχωρήσεων.

Όπως φαίνεται και τα προηγούμενα σχήματα (4-18 έως 4-27), η χρήση του πρωτοκόλλου SMAC με έλεγχο υπαναχωρήσεων, οδηγεί σε χαμηλότερο μέσο πλήθος εναλλαγών καναλιού σε σχέση με το πρωτόκολλο SMAC, γεγονός που οδηγεί σε μεγαλύτερη ευστάθεια του συστήματος. Επίσης, η ρυθμαπόδοση του συστήματος αυξάνεται, όπως στην περίπτωση χαμηλής πυκνότητας, ή παραμένει στα ίδια επίπεδα με το πρωτόκολλο SMAC, όπως στις περιπτώσεις μεσαίας και υψηλής πυκνότητας. Με τη χρήση του ελέγχου υπαναχωρήσεων, η άφιξη νέων δευτερευόντων χρηστών, και κατά συνέπεια η αύξηση συμφόρησης του καναλιού, γίνεται ταχύτερα αντιληπτή σε σχέση με το απλό πρωτόκολλο SMAC και έτσι ο πράκτορας αντιδρά ανάλογα, χωρίς να υπάρχει απώλεια πολλών πακέτων. Επιπλέον, αν δεν εμφανιστούν καινούργιοι δευτερεύοντες χρήστες στο κανάλι, αποφεύγονται οι άσκοπες εναλλαγές καναλιού. Η συμπεριφορά των δύο πρωτοκόλλων μπορεί να φανεί και στα σχήματα 4-28 και 4-29. Το σχήμα 4-28 αφορά χρήστες που χρησιμοποιούν το απλό πρωτόκολλο SMAC, όπου φαίνεται το μεγάλο πλήθος εναλλαγών καναλιού που πραγματοποιούνται κατά τη διάρκεια των 100s. Το σχήμα 4-29 δείχνει το αντίστοιχο διάγραμμα χρήσης καναλιών για χρήστες με πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων, όπου ο αριθμός των εναλλαγών καναλιού είναι φανερά μειωμένος, σχεδόν στα επίπεδα του MARL (βλ. Σχήμα 4-9).



Σχήμα 4-28 - Διάγραμμα χρήσης καναλιών για χρήστες με πρωτόκολλο SMAC. Μέσο PUL 0.2, Μέσο PER 0.2.



Σχήμα 4-29 - Διάγραμμα χρήσης καναλιών για χρήστες με πρωτόκολλο SMAC με έλεγχο υπαναχωρήσεων. Μέσο PUL 0.2, Μέσο PER 0.2.

4.3 Συμπεράσματα – Παρατηρήσεις

Από τα αποτελέσματα των προσομοιώσεων αποδείχθηκε αλλά και θεωρητικά ότι η παρουσία πολλών δευτερευόντων χρηστών αυξάνει τη συνολική χρησιμοποίηση των

λευκών διαστημάτων ενός καναλιού αλλά, ταυτόχρονα, μειώνει την ατομική ρυθμαπόδοση κάθε χρήστη.

Επιπλέον, σχεδιάστηκαν, αξιολογήθηκαν και επιβεβαιώθηκαν τρόποι μείωσης των εναλλαγών καναλιού (όπως η εξερεύνηση βασιζόμενη στις τιμές Q), οι οποίοι δεν απαιτούν επικοινωνία μεταξύ των χρηστών όπως συμβαίνει στην ενισχυτική μάθηση πολλών πρακτόρων και το πρωτόκολλο MMAC. Επίσης, επιτεύχθηκε αύξηση της ρυθμαπόδοσης και παράλληλη μείωση των εναλλαγών καναλιού, χωρίς τη χρήση επικοινωνίας μεταξύ χρηστών, μέσω του πρωτοκόλλου SMAC με έλεγχο του πλήθους των υπαναχωρήσεων. Ο συνδυασμός εξερεύνησης βασιζόμενης στις τιμές Q και του πρωτοκόλλου SMAC με έλεγχο υπαναχωρήσεων δίδει αρκετά ικανοποιητικά αποτελέσματα πολύ κοντά στις επιδόσεις του πρωτοκόλλου MMAC, χωρίς να απαιτείται σηματοδοσία μεταξύ των χρηστών.

Καθώς τα ανωτέρω αφορούν δίκτυα τα οποία θεωρούνται στατικά, λόγω της υποθεθείσας σχετικά σταθερής χρήσης από τους πρωτεύοντες χρήστες, η εφαρμογή τους σε μη-στατικά δίκτυα πιθανότατα δεν θα εμφανίζει τα οφέλη που εμφανίζονται σε στατικά. Η χρήση σε μη-στατικά δίκτυα θα απαιτούσε καλύτερες μεθόδους εξερεύνησης, για την εξεύρεση κατάλληλων καναλιών, οι οποίες να υποστηρίζουν την επικοινωνία των δευτερευόντων χρηστών, καθώς και κατάλληλες αλλαγές στα πρωτόκολλα, ώστε να αναγνωρίζουν και να αντιμετωπίζουν τη δυναμική συμπεριφορά. Σε κάθε περίπτωση, οι προτεινόμενες παραλλαγές του SMAC αναμένεται να εξασφαλίζουν μεγαλύτερη ταχύτητα προσαρμογής σε μεταβαλλόμενα περιβάλλοντα, με μειωμένες απαιτήσεις επικοινωνίας μεταξύ των χρηστών χωρίς σοβαρή απώλεια επιδόσεων.

Παράρτημα Α - Μέση τιμή του ελαχίστου n ανεξάρτητων ομοιόμορφα κατανεμημένων μεταβλητών.

Έστω X_1, X_2, \dots, X_n ανεξάρτητες τυχαίες μεταβλητές ομοιόμορφα κατανεμημένες με μέση τιμή M . Η συνάρτηση πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής X_i ομοιόμορφα κατανεμημένης στο διάστημα $[a, b]$ είναι

$$f_{X_i}(x) = \begin{cases} \frac{1}{b-a}, & \text{αν } a \leq x \leq b \\ 0, & \text{αλλιώς} \end{cases} \quad (\text{A.1})$$

η μέση τιμή της είναι

$$E[X_i] = \int_{-\infty}^{\infty} x f_{X_i}(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2} = M \quad (\text{A.2})$$

και η συνάρτηση κατανομής της είναι

$$F_{X_i}(x) = Pr(X \leq x) = \int_{-\infty}^x f_{X_i}(t) dt = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} \quad (\text{A.3})$$

Εφόσον όλα τα X_i έχουν την ίδια κατανομή ισχύει $Pr(X_i > x) = Pr(X > x)$ για $i = 1, 2, \dots, n$.

Έστω $W = \min \{X_1, X_2, \dots, X_n\}$. Τότε ισχύει

$$\begin{aligned} \{W \leq x\} &= \{\min\{X_1, X_2, \dots, X_n\} > x\}' = \left\{ \bigcap_{i=1}^n \{X_i > x\} \right\}' \Rightarrow \\ Pr(W \leq x) &= Pr\left(\left(\bigcap_{i=1}^n \{X_i > x\}\right)'\right) = 1 - Pr\left(\bigcap_{i=1}^n \{X_i > x\}\right) = \\ &= 1 - Pr(X_1 > x) Pr(X_2 > x) \dots Pr(X_n > x) = 1 - [Pr(X > x)]^n = \\ &= 1 - [1 - Pr(X \leq x)]^n = 1 - [1 - F_X(x)]^n = \\ &= \begin{cases} 0, & x < a \\ 1 - \left(1 - \frac{x-a}{b-a}\right)^n, & a \leq x < b \\ 1, & x \geq b \end{cases} \quad (\text{A.4}) \end{aligned}$$

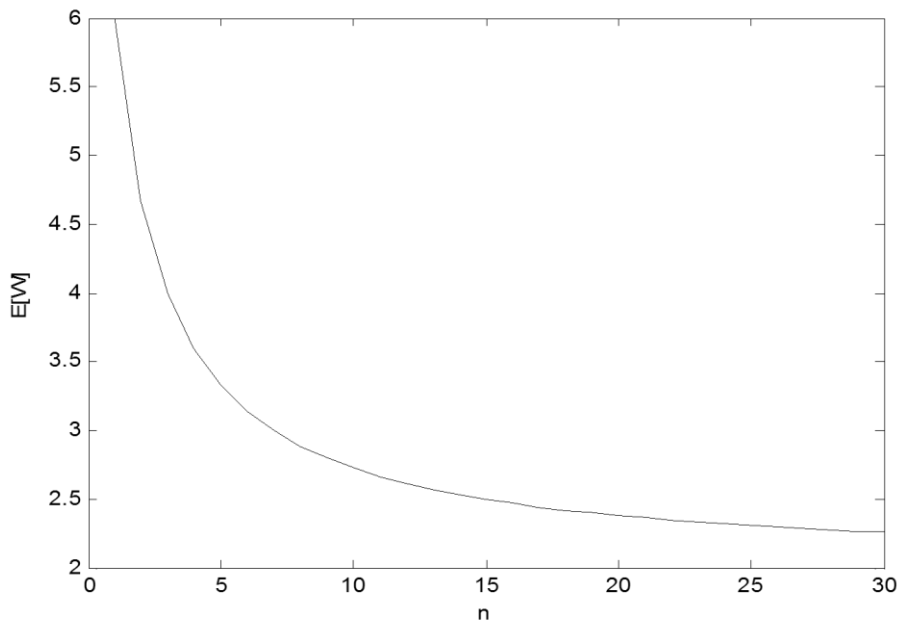
και επειδή $Pr(W \leq x) = F_W(x)$ και $f_W(x) = F'_W(x)$

$$f_W(x) = \begin{cases} 0, & x < a \\ \frac{n}{b-a} \left[1 - \frac{x-a}{b-a}\right]^{n-1}, & a \leq x < b \\ 0, & x \geq b \end{cases} \quad (A.5)$$

Η μέση τιμή της W είναι

$$\begin{aligned} E[W] &= \int_{-\infty}^{\infty} x f_W(x) dx = \frac{n}{b-a} \int_a^b x \left[1 - \frac{x-a}{b-a}\right]^{n-1} dx = \\ &= n \int_0^1 [b - (b-a)u] u^{n-1} du = \frac{b+na}{n+1} = \alpha + \frac{b-a}{n+1} \end{aligned} \quad (A.6)$$

Για $n=1$ η μέση τιμή $E[W]$ γίνεται $\frac{\alpha+b}{2} = M$ που είναι η μέση τιμή μίας τυχαίας μεταβλητής ομοιόμορφα κατανομημένης στο διάστημα $[a,b]$. Επίσης η μέση τιμή $E[W]$ είναι φθίνουσα για $n > 1$ και για $b > a$ όπως φαίνεται και στο σχήμα A-1 και επομένως $E[W] \leq M$. Επομένως όσο αυξάνεται ο αριθμός n των τυχαίων μεταβλητών τόσο μειώνεται η μέση τιμή $E[W]$ του ελαχίστου τους.



Σχήμα A-1 – Η μέση τιμή $E[W]$ συναρτήσει του αριθμού των τυχαίων μεταβλητών n , για $a=2$ και $b=10$.

Πολύ γνωστή είναι και η ανάλογη ιδιότητα για n τυχαίες μεταβλητές με εκθετική κατανομή και ίδιο λ , για τις οποίες το ελάχιστο τους είναι επίσης εκθετικά κατανομημένο με $\lambda' = n\lambda$ (δηλαδή η μέση τιμή του ελαχίστου είναι n φορές μικρότερη).

Βιβλιογραφία

- [1] J. Mitola and G. Q. Maguire, “Cognitive radio: making software radios more personal”, IEEE Personal Communications, 6, pp. 13-18, Aug 1999.
- [2] J. Mitola, “Cognitive Radio: an integrated agent architecture for software defined radio, Ph.D. Thesis, KTH Royal Institute of Technology, 2000.
- [3] S. Haykin, “Cognitive radio: brain-empowered wireless communications”, IEEE Journal on Selected Areas in Communications 23 (2), pp. 201–220, 2005.
- [4] 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 Journal (Elsevier), Sept. 2006.
- [5] T. Yucek and H. Arslan, “A Survey of Spectrum Sensing Algorithms for Cognitive Radio Applications”, IEEE Communications Surveys & Tutorials, Vol. 11, 1, First Quarter 2009.
- [6] R.W. Thomas, L.A. DaSilva, A.B. MacKenzie, Cognitive networks, Proc. IEEE DySPAN 2005, pp. 352–360, November 2005.
- [7] F.K. Jondral, Software-defined radio-basic and evolution to cognitive radio, EURASIP Journal on Wireless Communication and Networking 2005.
- [8] D. Cabric, S.M. Mishra, R.W. Brodersen, Implementation issues in spectrum sensing for cognitive radios, Proc. 38th Asilomar Conference on Signals, Systems and Computers 2004, pp. 772–776, November 2004.
- [9] S. Shankar, C. Cordeiro, and K. Challapali, “Spectrum agile radios: utilization and sensing architectures,” in Proc. IEEE Int. Symposium on New Frontiers in Dynamic Spectrum Access Networks, Baltimore, Maryland, USA, pp. 160–169, Nov. 2005.
- [10] A. Sahai, N. Hoven and R. Tandra, “Some fundamental limits on cognitive radio,” Proc. Allerton Conf. on Commun., Control, and Computing, Monticello, Illinois, Oct. 2004.
- [11] K. Maeda, A. Benjebbour, T. Asai, T. Furuno, and T. Ohya, “Recognition among OFDM-based systems utilizing cyclostationarity-inducing transmission,” Proc. IEEE Int. Symposium on New Frontiers in Dynamic Spectrum Access Networks, Dublin, Ireland, pp. 516–523, Apr. 2007.
- [12] J. G. Proakis, Digital Communications, 4th ed. McGraw-Hill, 2001.

- [13] D. Cabric, A. Tkachenko, and R. Brodersen, "Spectrum sensing measurements of pilot, energy, and collaborative detection," Proc. IEEE Military Commun. Conf., Washington, D.C., USA, pp. 1–7, Oct. 2006.
- [14] X. Liu and S. Shankar, "Sensing-based opportunistic channel access," Mobile Networks and Applications, vol. 11, no. 4, pp. 577–591, 2006.
- [15] FCC Spectrum Policy Task Force, "FCC Report of the Spectrum Efficiency Working Group", Nov. 2002.
- [16] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access: signal processing, networking, and regulatory policy," IEEE Signal Processing Mag., vol. 4, no. 3, pp. 79–89, May 2007.
- [17] Specification of the Bluetooth system, Master Table of Contents & Compliance Requirements, Bluetooth, SIG Std. Bluetooth Standard, Nov. 2004.
- [18] Standard for Wireless Regional Area Networks (WRAN) - Specific requirements - Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and procedures for operation in the TV Bands, The Institute of Electrical and Electronics Engineering, Inc. Std. IEEE 802.22.
- [19] C. R. Stevenson, C. Cordeiro, E. Sofer, and G. Chouinard, "Functional requirements for the 802.22 WRAN standard," IEEE 802.22-05/0007r46, Sept. 2005.
- [20] 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.
- [21] Draft Supplement to STANDARD FOR Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Specification for Radio Resource Measurement, The Institute of Electrical and Electronics Engineering, Inc. Std. IEEE 802.11k/D0.7, Oct. 2003.
- [22] Q. Zhao, L. Tong, A. Swami, Decentralized cognitive MAC for dynamic spectrum access, Proc. IEEE DySPAN 2005, pp. 224–232, November 2005.
- [23] N. Nie, C. Comaniciu, Adaptive channel allocation spectrum etiquette for cognitive radio networks, Proc. IEEE DySPAN 2005, pp. 269–278, November 2005.
- [24] J. Marinho, E. Monteiro, "Cognitive Radio: Technology Survey and Future Research Directions", ANACOM 2012.

- [25] B. Wang and K.J. Ray Liu, "Advances in Cognitive Radio Networks: A Survey", IEEE Journal of Selected Topics in Signal Processing, Vol. 5, No.1, February 2011.
- [26] R.S. Sutton, A. G. Barto, "Reinforcement Learning: An Introduction", MIT Press, 1998.
- [27] L.P. Kaelbling, M.L. Littman and A.W. Moore, "Reinforcement Learning: A survey", Journal of Artificial Intelligence Research 4, 237-285, 1996.
- [28] A. Gosavi, "Reinforcement Learning: A Tutorial Survey and Recent Advances", INFORMS Journal on Computing, vol. 21, no. 2, pp. 178-192, Spring 2009.
- [29] C.J.C.H. Watkins and P. Dayan, "Q-learning", Machine Learning 8 (3), 279-292, 1992.
- [30] R.J. Williams and L.C. Baird III, "Tight performance bounds on greedy policies based on imperfect value functions", Tech. rep. NU-CCS-93-14, Northeastern University, College of Computer Science, Boston, MA, 1993.
- [31] D.P. Bertsekas, "Dynamic Programming: Deterministic and Stochastic Models", Prentice Hall, Englewood Cliffs, NJ, 1987.
- [32] C.J.C.H. Watkins, "Learning from Delayed Rewards", Ph.D. Thesis, King's College, Cambridge, UK, 1989.
- [33] J.N. Tsitsiklis, "Asynchronous stochastic approximation and Q-learning", Machine Learning 16 (3), 1994.
- [34] T. Jaakkola, M.I. Jordan and S.P. Singh, "On the convergence of stochastic iterative dynamic programming algorithms", Neural Computation 6 (6), 1994.
- [35] J. Peng and R.J. Williams, "Incremental multi-step Q-learning", Proceedings of the Eleventh International Conference on Machine Learning, pp. 226-232, San Francisco, CA, Morgan Kaufmann, 1994.
- [36] H. Jaap van den Herik, D. Hennes, M. Kaisers, K. Tuyls and K. Verbeeck, "Multi-agent Learning Dynamics: A Survey", Cooperative Information Agents XI, pp. 36-56, September 2007.
- [37] K.-L. A. Yau, P. Komisarczuk and P.D. Teal, "Enhancing Network Performance in Distributed Cognitive Radio Networks using Single-Agent and Multi-Agent Reinforcement Learning", 35th Annual IEEE Conference on Local Computer Networks, Denver, Colorado, 2010.
- [38] K.-L. A. Yau, "Context Awareness and Intelligence in Cognitive Radio Networks: Design and Applications", Ph.D. Thesis, Victoria University of Wellington, New Zealand, 2010.

- [39] K.-L. A. Yau, P. Komisarczuk and P.D. Teal, "A context-aware and intelligent dynamic channel selection scheme for cognitive radio networks", 4th IEEE International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM), June 2009.
- [40] K.-L. A. Yau, P. Komisarczuk and P.D. Teal, "Context-awareness and intelligence in distributed cognitive radio networks: a reinforcement learning approach", IEEE Australia Communication Theory Workshop, (AUSTCTW), Feb 2010.
- [41] K.-L. A. Yau, P. Komisarczuk and P.D. Teal, "Achieving efficient and optimal joint action in distributed cognitive radio networks using payoff propagation", IEEE International Conference on Communications (ICC), May 2010.
- [42] R.S. Sutton, "Learning to predict by the method of temporal differences", Machine Learning 3 (1), 9-44, 1988.