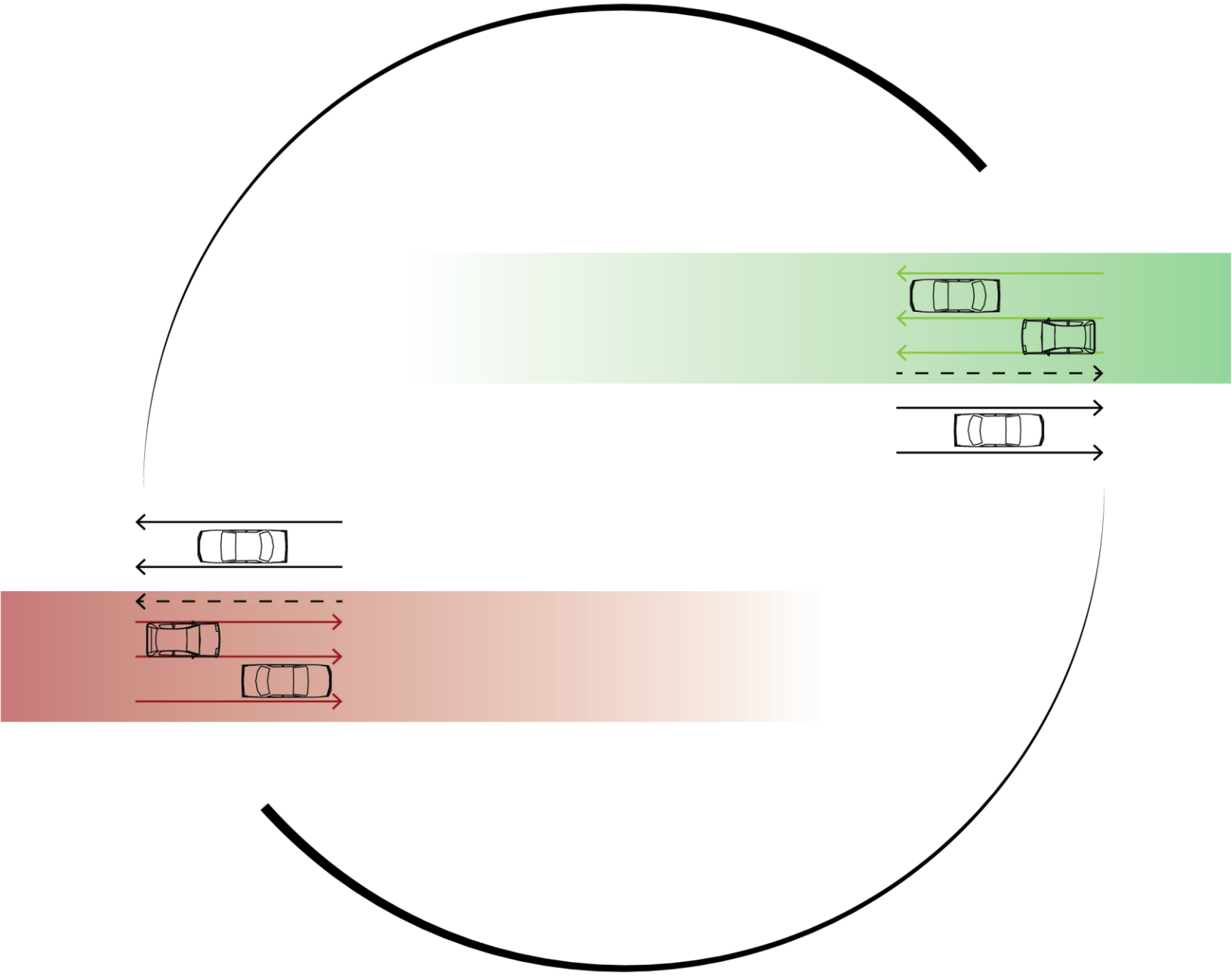
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Δυναμικός διαμοιρασμός λωρίδων κυκλοφορίας σε αστικές αρτηρίες με χρήση μεθόδων ενισχυτικής μάθησης



Διπλωματική Εργασία:
Κατερίνα Βακρινού
Επιβλέπουσα καθηγήτρια:
Ελένη Ι. Βλαχογιάννη,
Καθηγήτρια Σχολής Πολιτικών Μηχανικών Ε.Μ.Π.

Αθήνα, Ιούλιος 2023

Ευχαριστίες

Η ολοκλήρωση της παρούσας διπλωματικής εργασίας σηματοδοτεί το τέλος του προπτυχιακού κύκλου σπουδών μου στην Σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Έναν κύκλο που παρά τις δυσκολίες, με έκανε να αγαπήσω το αντικείμενο του Συγκοινωνιολόγου Μηχανικού και μου χάρισε στόχους για το μέλλον.

Αρχικά, θα ήθελα να ευχαριστήσω θερμά την κα. Ελένη Ι. Βλαχογιάννη, Καθηγήτρια της Σχολής Πολιτικών Μηχανικών Ε.Μ.Π., για την εμπιστοσύνη που μου έδειξε με την ανάθεση ενός τόσο ενδιαφέροντος και απαιτητικού θέματος διπλωματικής εργασίας αλλά και για τα γνωστικά εφόδια που μου παρείχε κατά τη διάρκεια των σπουδών μου.

Ένα μεγάλο ευχαριστώ οφείλω στον κ. Κωνσταντίνο Κατζηλιέρη, υποψήφιο Διδάκτορα του Ε.Μ.Π., για την καθοριστική συμβολή και την καθοδήγηση σε όλα τα στάδια εκπόνησης της παρούσας εργασίας. Δεν μπορώ παρά να αναγνωρίσω την βοήθεια που μου προσέφερε μέσω της γνώσης του, των πολύτιμων συμβουλών του, της υπομονής του και της αγάπης του για το αντικείμενο.

Επίσης, θα ήθελα να ευχαριστήσω τον κ. Εμμανουήλ Καμπιτάκη, υποψήφιο Διδάκτορα του Ε.Μ.Π για τη γνωστική του βοήθεια στην διεκπεραίωση της εργασίας αλλά και για τις συμβουλές του για την μετέπειτα ακαδημαϊκή και επαγγελματική μου πορεία.

Σε αυτό το σημείο θα ήθελα να απευθύνω ένα μεγάλο ευχαριστώ στους ανθρώπους εκείνους που στάθηκαν δίπλα μου κατά τη διάρκεια της εκπόνησης της διπλωματικής εργασίας αλλά και καθ' όλη τη διάρκεια των σπουδών μου. Υπήρξαν πάντα στήριγμά μου σε όλες τις δυσκολίες που αντιμετώπισα. Αυτοί δεν είναι άλλοι από την οικογένεια μου και τους φίλους μου. Ευχαριστώ πολύ, λοιπόν, τους γονείς μου, Όλγα και Γιάννη, που ήταν πάντα στο πλάι μου, όσο δύσκολο και αν ήταν αυτό. Ευχαριστώ πολύ τους φίλους μου, Γιώργο, Ασημίνα, Αντιγόνη, Ιλεάννα, Ηλία, Χρήστο, Λυδία, Πάνο, Βέρα, Ευγενία και Ευθύμη για όλες τις όμορφες στιγμές που μου έχουν προσφέρει και για την κατανόηση τους. Τέλος, ένα μεγάλο ευχαριστώ οφείλω στον Φίλιππο που με ενθάρρυνε σε κάθε μου βήμα και που διάβασε ολόκληρη την εργασία και με στήριξε με εύστοχες παρατηρήσεις και στην Μαίρη, που μου έδειξε έμπρακτα τη στήριξή της αναλαμβάνοντας την επιμέλεια του εξωφύλλου.

Αθήνα, Ιούλιος 2023

Κατερίνα Βακρινού

Δυναμικός διαμοιρασμός λωρίδων κυκλοφορίας σε αστικές αρτηρίες με τη χρήση μεθόδων ενισχυτικής μάθησης.

Συγγραφέας: Κατερίνα Βακρινού

Επιβλέπουσα Καθηγήτρια: Ελένη Ι. Βλαχογιάννη

ΣΥΝΟΨΗ

Τα τελευταία χρόνια, οι εξελίξεις στους τομείς της τεχνολογίας των δεδομένων και των συνδεδεμένων οχημάτων σε συνδυασμό με την ολοένα αυξανόμενη κυκλοφορική ζήτηση στα αστικά οδικά δίκτυα έχουν συμβάλλει στην ανάπτυξη νέων μεθόδων διαχείρισης της κυκλοφορίας, όπως είναι οι εναλλασσόμενες λωρίδες κυκλοφορίας. Οι εναλλασσόμενες λωρίδες κυκλοφορίας χρησιμοποιούνται παγκοσμίως, ωστόσο μόνο στην στατική τους μορφή. Γι' αυτό το λόγο αποτελούν συχνά ένα μη αποδοτικό μέτρο διαχείρισης της κυκλοφορίας. Στόχος της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός μοντέλου δυναμικού διαμοιρασμού λωρίδων κυκλοφορίας σε μία οδική αρτηρία, χρησιμοποιώντας τον αλγόριθμο Proximal Policy Optimization (PPO), που ανήκει στην κατηγορία των μεθόδων ενισχυτικής μάθησης. Το ανεπτυγμένο μοντέλο εκπαιδεύεται και αξιολογείται σε σενάρια πραγματικών περιπτώσεων κυκλοφοριακής ζήτησης και σε διαφορετικά γεωμετρικά χαρακτηριστικά του οδικού δικτύου, μέσω προσομοίωσης στο προγραμματιστικό περιβάλλον Simulation of Urban Mobility (SUMO). Τα αποτελέσματα δείχνουν ότι η στρατηγική διαχείρισης της κυκλοφορίας με δυναμικά εναλλασσόμενες λωρίδες οδηγεί σε αύξηση της μέσης ταχύτητας των οχημάτων έως και 10% και μείωση του συνολικού χρόνου ταξιδιού και καθυστερήσεων έως και 57%. Τέλος, η αξιολόγηση σε διαφορετικές γεωμετρίες του δικτύου έδειξε ότι η αποδοτικότητα του μέτρου μειώνεται όσο μικρότερη είναι η απόσταση μεταξύ των κόμβων του δικτύου.

Λέξεις κλειδιά: Δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας, δυναμική διαχείριση αστικού χώρου, αστικά οδικά δίκτυα, ενισχυτική μάθηση, νευρωνικά δίκτυα, αλγόριθμος Proximal Policy Optimization (PPO), Simulation of Urban Mobility (SUMO).

Dynamic traffic lane allocation on urban arterials using reinforcement learning methods.

Author: Katerina Vakrinou

Supervising Professor: Eleni I. Vlahogianni

ABSTRACT

In recent years, developments in the fields of data technology and connected vehicles combined with the ever-increasing traffic demand on urban road networks have contributed to the development of new traffic management methods such as reversible lanes. Reversible traffic lanes have been implemented worldwide, but only in their static form. For this reason, they are often an inefficient traffic management measure. The aim of this thesis is to develop a dynamic lane allocation model on a road artery, using reinforcement learning methods and specifically the Proximal Policy Optimization (PPO) algorithm. This model is trained and evaluated in real traffic demand scenarios and different geometric characteristics of the road network through simulation in the Simulation of Urban Mobility (SUMO) programming environment. Findings showed that implementing dynamically reversible lanes may lead to an increase in average vehicle speed of up to 10% and a reduction in total travel time and delays of up to 57%. Finally, the evaluation in different geometric characteristics of the network showed that the efficiency of the measure improves the greater the distance between the junctions of the network is.

Keywords: Dynamic reversible traffic lanes, dynamic urban space allocation, urban road networks, reinforcement learning, neural networks, Proximal Policy Optimization (PPO) algorithm, Simulation of Urban Mobility (SUMO).

Περίληψη

Η ραγδαία ανάπτυξη των σύγχρονων πόλεων καθιστά πλέον σαφές ότι η αποδοτική και αποτελεσματική χρήση του χώρου στις αστικές περιοχές αποκτά ολοένα και μεγαλύτερη σημασία. Ο αστικός χώρος είναι περιορισμένος και εξυπηρετεί ένα ευρύ φάσμα δραστηριοτήτων. Προς αντιμετώπιση αυτών των φαινομένων, η νέα τάση στον σύγχρονο πολεοδομικό σχεδιασμό αφορά τη δυναμική διαχείριση του χώρου. Η συνεχής αύξηση της κυκλοφοριακής ζήτησης, σε συνδυασμό με την ανάπτυξη των συνδεδεμένων οχημάτων και της τεχνολογίας των δεδομένων δημιουργούν τη βάση για την εφαρμογή της δυναμικής διαχείρισης του αστικού οδικού χώρου, κυρίως μέσω της εφαρμογής δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας.

Οι δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας αναφέρονται σε ένα σύστημα που επιτρέπει την αλλαγή της κατεύθυνσης ή της χρήσης των λωρίδων κυκλοφορίας, ανάλογα με τις συνθήκες κυκλοφορίας και τη ζήτηση. Οι περισσότερες έρευνες προσεγγίζουν την ανάπτυξη μοντέλων εναλλαγής των λωρίδων, τα οποία βασίζονται σε μεθόδους βελτιστοποίησης. Βασικό μειονέκτημα αυτών των μεθόδων είναι ότι η γραμμικότητα και η βεβαιότητα, χαρακτηριστικά των προβλημάτων για την επίλυση των οποίων ενδείκνυται η χρήση γραμμικού προγραμματισμού και βελτιστοποίησης, δεν αποτελούν ιδιότητες της κυκλοφορίας.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη προτύπου Ενισχυτικής Μάθησης για τη βελτιστοποίηση των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας σε αστικά σηματοδοτούμενα δίκτυα. Επιλέχθηκε η εργασία με αλγορίθμους ενισχυτικής μάθησης γιατί το γεγονός ότι μπορούν να λαμβάνουν αποφάσεις βάσει παρατηρήσεων και να προβλέπουν άγνωστες καταστάσεις τους καθιστά ιδανικούς για την επίλυση κυκλοφοριακών ζητημάτων.

Πρώτο βήμα για την ανάπτυξη του μοντέλου εναλλαγής των λωρίδων κυκλοφορίας αποτέλεσε η δομή ενός δικτύου, συγκεκριμένα η δημιουργία ενός οδικού τμήματος τριών κόμβων. Οι λωρίδες της κύριας αρτηρίας είναι εφτά εκ των οποίων οι δύο κεντρικές επιλέχθηκε να είναι εναλλασσόμενες. Η σηματοδότηση του δικτύου επιλέχθηκε για διευκόλυνση της κυκλοφορίας να γίνει με αυτόματους σηματοδότες. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκαν πολλοί και διαφορετικοί κυκλοφοριακοί φόρτοι.

Για την επίτευξη του σκοπού, έπειτα καταστρώνεται το πρόβλημα της εναλλαγής λωρίδας σε περιβάλλον ενισχυτικής μάθησης. Ως κατάσταση του περιβάλλοντος λαμβάνονται τα κυκλοφοριακά χαρακτηριστικά που παρατηρούνται στο δίκτυο ανά 300 δευτερόλεπτα. Ο πράκτορας μετά από κάθε κατάσταση καλείται να επιλέξει μία ενέργεια, δηλαδή μία από τις τρεις διαμορφώσεις των λωρίδων κυκλοφορίας οι οποίες έχουν οριστεί.

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκαν δύο αλγόριθμοι ενισχυτικής μάθησης ο αλγόριθμος Deep Q-Learning (DQN) με μη ικανοποιητικά αποτελέσματα σύγκλισης και μετά ο αλγόριθμος Proximal Policy Optimization, με καλύτερα αποτελέσματα, ωστόσο όχι πλήρως ικανοποιητικά. Για να βελτιωθεί η σύγκλισης επιλέχθηκε η εκπαίδευση σε δύο στάδια, ένα στάδιο προεκπαίδευσης και ένα στάδιο εκπαίδευσης. Η διαφορά ανάμεσα στα δύο στάδια είναι στην ανταμοιβή που λαμβάνει ο πράκτορας μετά από κάθε ενέργεια. Στην προ-εκπαίδευση η ανταμοιβή έχει οριστεί με τέτοιο τρόπο ώστε να θεωρεί ως ιδανική τη διαμόρφωση που δίνει μία επιπλέον λωρίδα

στην κατεύθυνση με το μεγαλύτερο φόρτο. Στη φάση της εκπαίδευσης ως ανταμοιβή λαμβάνεται ο μέγιστος μέσος χρόνος διαδρομής που παρατηρείται στις δύο κατευθύνσεις.

Το τελικό εκπαιδευμένο μοντέλο αξιολογείται μέσω προσομοίωσης σε τρία διαφορετικά σενάρια ζήτησης που προσομοιάζουν πραγματικές κυκλοφοριακές συνθήκες. Ένα σενάριο πρωινής και απογευματινής αιχμής, ένα σενάριο κυκλοφοριακών συνθηκών κατά τη διάρκεια μεγάλων εκδηλώσεων και σε ένα σενάριο αναγκαστικής διακοπής της κυκλοφορίας σε τμήμα 100 μέτρων μιας λωρίδας του δικτύου. Η σύγκριση αφορά σε τρεις διαφορετικούς τρόπους διαχείρισης της κυκλοφορίας, δηλαδή στο εκπαιδευμένο μοντέλο, σε ένα μοντέλο που αποδίδει την επιπλέον λωρίδα στην κατεύθυνση με τον μεγαλύτερο φόρτο και σε μία στατική διαμόρφωση του οδικού χώρου. Επιπλέον, για να διερευνηθεί και η επιρροή των κόμβων και της σηματοδότησης στην αποδοτικότητα του μέτρου, τα παραπάνω κυκλοφοριακά σενάρια αξιολογούνται και σε οδικά τμήματα με διαφορετικές αποστάσεις μεταξύ των κόμβων.

Τα αποτελέσματα αυτής της αξιολόγησης δείχνουν καλύτερες συνθήκες κυκλοφορίας με την εφαρμογή του μοντέλου διαμοιρασμού των λωρίδων που αναπτύχθηκε. Συγκεκριμένα παρατηρήθηκε αύξηση της μέσης ταχύτητας των οχημάτων έως και 10% και μείωση του συνολικού χρόνου ταξιδιού και των καθυστερήσεων έως και 57%. Σημαντική είναι και η μείωση του παρατηρήθηκε στις εκπομπές καυσαερίων και στην κατανάλωση καυσίμων. Επιπλέον, ενδιαφέρον παρουσιάζει ότι το εκπαιδευμένο μοντέλο εμφανίζει καλύτερη απόδοση από το μοντέλο που γνωρίζει τις κυκλοφοριακές συνθήκες και αποδίδει τη λωρίδα στην κατεύθυνση με τον μεγαλύτερο φόρτο. Όσον αφορά τη γεωμετρία του δικτύου αποδεικνύεται ότι όσο μεγαλύτερη είναι η απόσταση μεταξύ των κόμβων, τόσο καλύτερες είναι οι κυκλοφοριακές συνθήκες του δικτύου από την εφαρμογή του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας.

Ολοκληρώνοντας την παρούσα εργασία προέκυψαν σημεία για περαιτέρω έρευνα. Η διερεύνηση του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας σε επίπεδο πλέον δικτύου μπορεί να δώσει πιο ρεαλιστικά αποτελέσματα. Επίσης, ενδιαφέρον θα παρουσίαζε μία έρευνα για τον βέλτιστο αριθμό δυναμικά εναλλασσόμενων λωρίδων σε μία οδική αρτηρία. Τέλος, προτείνεται και η διερεύνηση δυναμικά εναλλασσόμενων λωρίδων αυτή τη φορά όμως ως προς τα μέσα που επιτρέπουν να κυκλοφορούν επάνω τους, κυρίως για την ενίσχυση των μέσων μαζικής μεταφοράς.

Περιεχόμενα

Ευρετήριο Εικόνων

Ευρετήριο Διαγραμμάτων

Ευρετήριο Πινάκων

Κεφάλαιο 1: Εισαγωγή

1.1Γενική Ανασκόπηση

Στον σημερινό ραγδαία αστικοποιούμενο κόσμο, η αποδοτική και αποτελεσματική χρήση του χώρου στις αστικές περιοχές αποκτά ολοένα και μεγαλύτερη σημασία. Ο αστικός χώρος είναι περιορισμένος και εξυπηρετεί ένα ευρύ φάσμα δραστηριοτήτων. Οδικό δίκτυο, μέσα μαζικής μεταφοράς, χώροι στάθμευσης, φορτοεκφόρτωση εμπορευμάτων, ποδηλατόδρομοι, υπαίθριες αγορές αλλά και χώροι αναψυχής πρέπει να μπορούν να δρουν παράλληλα. Ωστόσο, ο έως τώρα συμβατικός πολεοδομικός σχεδιασμός ευνοεί πολύ συχνά μέσα και δραστηριότητες που καταναλώνουν πολύ χώρο, όπως τα ιδιωτικά αυτοκίνητα, παρόλο που ο χώρος που καταναλώνει ένας οδηγός σε ένα αυτοκίνητο ανά χιλιόμετρο είναι αρκετές τάξεις μεγέθους μεγαλύτερος από οποιονδήποτε άλλο μέσο (*ITF,2022*).

Για την αντιμετώπιση των παραπάνω ζητημάτων, μία νέα τάση στον πολεοδομικό σχεδιασμό των σύγχρονων πόλεων που προτείνεται τα τελευταία χρόνια σε επίπεδο έρευνας, είναι η δυναμική κατανομή του χώρου, δηλαδή η διάθεση του χώρου στη χρήση-δραστηριότητα με τη περισσότερη ζήτηση για κάθε χρονική στιγμή. Αυτό μπορεί να γίνει, με την εφαρμογή πολυχρηστικών χώρων, οι οποίοι για παράδειγμα τις πρωινές ώρες μπορούν να χρησιμοποιούνται για την φορτοεκφόρτωση εμπορευμάτων, στη συνέχεια της ημέρας να δίνονται στην κυκλοφορία ώστε να διευκολύνουν την μετακίνηση των εργαζομένων από και προς τους χώρους εργασίας τους και στο ενδιάμεσο και στο τέλος της ημέρας να διατίθενται σε καταστήματα για εστίαση και αγορές. Μία τέτοια πολιτική δυναμικής διαχείρισης του χώρου πρέπει να διέπεται από πέντε βασικούς πυλώνες (*ITF,2022*):

- Υιοθέτηση δεικτών καλής χρήσης του αστικού χώρου
- Συμπεριληπτική ανακατανομή του χώρου, η οποία λαμβάνει υπόψη διαφορετικούς χρήστες και χρήσεις
- Προτεραιότητα στα άτομα και όχι στα οχήματα κατά την κατανομή του χώρου
- Διερεύνηση δυναμικών εναλλακτικών στην ανακατανομή του χώρου και
- Υιοθέτηση αρχών του Ασφαλούς Συστήματος (Safe System)

Γενικά, η έννοια της δυναμικής χρήσης του χώρου ενσωματώνει την ιδέα ότι τα αστικά περιβάλλοντα πρέπει να είναι προσαρμόσιμα, ευέλικτα και να ανταποκρίνονται στις μεταβαλλόμενες ανάγκες και απαιτήσεις των κατοίκων τους. Ξεπερνά τις παραδοσιακές προσεγγίσεις του αστικού πολεοδομικού σχεδιασμού και υιοθετεί την ιδέα ότι οι χώροι πρέπει να είναι πολυλειτουργικοί, εύκολα μεταμορφώσιμοι και ικανοί να φιλοξενούν ποικίλες δραστηριότητες και χρήστες.

Η σημασία της δυναμικής χρήσης του χώρου έγκειται στην ικανότητά της να βελτιστοποιεί τη χρήση της γης και να μεγιστοποιεί τα οφέλη που προκύπτουν από τους περιορισμένους διατιθέμενους πόρους. Με την υιοθέτηση μιας δυναμικής προσέγγισης, οι πόλεις μπορούν να αντιμετωπίσουν μια πληθώρα προβλημάτων, όπως η αύξηση του πληθυσμού, η περιβαλλοντική βιωσιμότητα, η κοινωνική ένταξη και η οικονομική ευρωστία.

Στις πυκνοκατοικημένες αστικές περιοχές, όπου η γη είναι λιγοστή και πολύτιμη, η αποδοτική χρήση του χώρου είναι υψίστης σημασίας. Οι δυναμικοί χώροι επιτρέπουν την ενσωμάτωση

διαφόρων λειτουργιών σε μια ενιαία περιοχή, επιτρέποντας στους κατοίκους των πόλεων να ζουν, να εργάζονται και να κοινωνικοποιούνται σε κοντινή απόσταση. Αυτή η ενσωμάτωση προωθεί την προσβασιμότητα, μειώνει τους χρόνους μετακίνησης και βελτιώνει τη συνολική ποιότητα ζωής.

Επιπλέον, η δυναμική χρήση του χώρου προάγει την περιβαλλοντική βιωσιμότητα με την ελαχιστοποίηση της αστικής εξάπλωσης και τη μείωση της κατανάλωσης ενέργειας που συνδέεται με τις μεταφορές. Βελτιστοποιώντας τη χρήση του χώρου, οι πόλεις μπορούν να διατηρήσουν τις φυσικές περιοχές, να προωθήσουν τις πράσινες υποδομές και να ενισχύσουν την ανθεκτικότητα των αστικών οικοσυστημάτων.

Από οικονομική άποψη, η δυναμική χρήση του χώρου μπορεί να καταλύσει την καινοτομία και την επιχειρηματικότητα. Με τη δημιουργία ευέλικτων χώρων που μπορούν να προσαρμόζονται στις εξελισσόμενες απαιτήσεις της αγοράς, οι πόλεις προωθούν την ανάπτυξη αναδυόμενων βιομηχανιών, υποστηρίζουν τις μικρές επιχειρήσεις και προσελκύουν επενδύσεις.

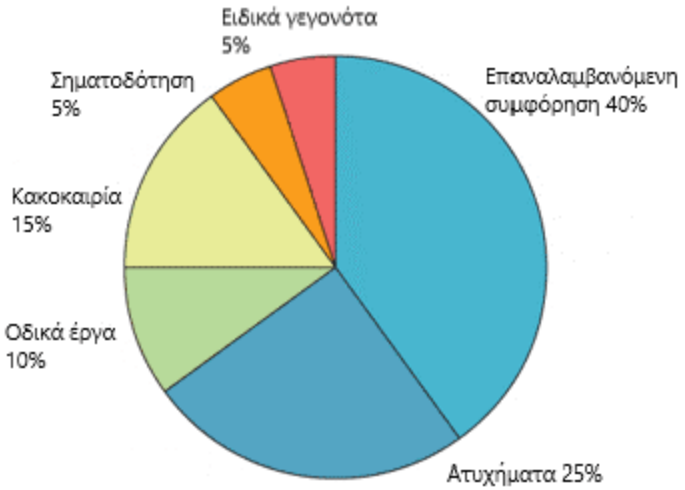
Εν κατακλείδι, η δυναμική χρήση του χώρου είναι υψίστης σημασίας στις αστικές περιοχές. Αγκαλιάζοντας την προσαρμοστικότητα, την ευελιξία και την πολυλειτουργικότητα, οι πόλεις μπορούν να απελευθερώσουν το πλήρες δυναμικό των περιορισμένων πόρων τους, να προωθήσουν τη βιωσιμότητα, τη συμμετοχικότητα και την οικονομική ανάπτυξη και τελικά να δημιουργήσουν βιώσιμα, ανθεκτικά και ακμαία αστικά περιβάλλοντα για τους κατοίκους τους.

Οι πιο σημαντικές υποδομές των πόλεων, λαμβάνοντας υπόψιν τον χώρο που αυτές καταλαμβάνουν, συνήθως σχετίζονται με τις μεταφορές. Αστικοί αυτοκινητόδρομοι, λεωφόροι, συνοικιακές οδοί, θέσεις και χώροι στάθμευσης, ισόπεδοι και ανισόπεδοι κόμβοι αποτελούν απαραίτητα θεμέλια του αστικού ιστού. Παράλληλα ωστόσο, καταναλώνουν και μια τεράστια ποσότητα πολύτιμου χώρου η οποία είναι αναγκαία για την εξυπηρέτηση και άλλων χρήσεων στα πλαίσια μίας πόλης. Ένα ακόμη πρόβλημα που καλούνται να αντιμετωπίσουν οι σύγχρονες πόλεις είναι ότι οι υποδομές τους για το μεγαλύτερο χρονικό διάστημα της ημέρας υπολειτουργούν, ενώ κατά τις ώρες αιχμής κρίνονται ανεπαρκείς για την ορθή εξυπηρέτηση της ζήτησης.

Σε αυτή την έλλειψη χώρου έρχεται να προστεθεί και η συνεχώς αυξανόμενη κυκλοφοριακή συμφόρηση, η οποία μπορεί να προκαλείται από ποικίλα αίτια που διαφέρουν από περιοχή σε περιοχή, όπως για παράδειγμα:

- Απροσδόκητες αυξήσεις στη ζήτηση κυκλοφορίας
- Οδικά ατυχήματα που εμποδίζουν τις λωρίδες κυκλοφορίας
- Οδικά έργα
- Μη συγχρονισμένοι χρόνοι σηματοδοτήσεων
- Άλλοι τυχαίοι παράγοντες

Στο Διάγραμμα 1 φαίνεται η ποσοστιαία κατανομή της κυκλοφοριακής συμφόρησης ανάλογα με το αίτιο της.



Διάγραμμα 1: Κατανομή της κυκλοφοριακής συμφόρησης ανάλογα με το αίτιό της
(Πηγή: FHWA- U.S. Department of Transportation)

Βασικό χαρακτηριστικό του αστικού οδικού χώρου είναι ο διαμοιρασμός του σε λωρίδες, οι οποίες διαχωρίζουν τις κατευθύνσεις κυκλοφορίας αλλά και τη ροή των οχημάτων στην ίδια κατεύθυνση. Ως εκ τούτου αποτελούν έναν χωρικό διαχωρισμό, του οποίου η δυναμική διαχείριση θα μπορούσε να αποτελέσει λύση στα παραπάνω προβλήματα. Το μέτρο αυτό ονομάζεται δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας και μπορεί να εφαρμοστεί σε κύριες οδικές αρτηρίες του αστικού οδικού δικτύου. Προς το παρόν, οι εναλλασσόμενες λωρίδες κυκλοφορίας έχουν εφαρμοστεί σε διάφορες πόλεις παγκοσμίως, ωστόσο μόνο στη στατική τους μορφή. Δηλαδή, η κατεύθυνση συγκεκριμένων λωρίδων κυκλοφορίας αλλάζει σε σταθερά και προκαθορισμένα χρονικά διαστήματα μέσα σε μία μέρα. Παραδείγματα τέτοιων μέτρων φαίνονται στην Εικόνα 1.

Σε περιπτώσεις όμως που η κυκλοφορία δεν συμπεριφέρεται έτσι όπως έχει υπολογισθεί το μέτρο αυτό δημιουργεί χειρότερες συνθήκες στην κυκλοφορία.

Εικόνα 1: Παραδείγματα στατικά εναλλασσόμενων λωρίδων κυκλοφορίας
(Πηγή: Wolshon,2006)

Η ανάπτυξη της τεχνολογίας και ιδιαίτερα της επιστήμης των δεδομένων και των συνδεδεμένων οχημάτων δίνουν πλέον τη δυνατότητα για την ανάπτυξη δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας. Δηλαδή, λωρίδων που μπορούν να αλλάξουν την κατεύθυνση τους κάθε χρονική στιγμή ανάλογα με τη ζήτηση. Το κύριο πρόβλημα που προκύπτει για την εφαρμογή αυτού του μέτρου είναι η ανάπτυξη μοντέλων τα οποία μπορούν να ανταποκριθούν στις ανάγκες ενός τόσο πολύπλοκου και συνεχώς μεταβαλλόμενου συστήματος όπως είναι η κυκλοφοριακή ζήτηση μέσα στον αστικό ιστό.

Ως ένα ανερχόμενο θέμα έρευνας, η παρούσα βιβλιογραφία για τις δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας είναι περιορισμένη. Βασίζεται κυρίως στην ανάπτυξη μοντέλων εναλλαγής των λωρίδων με μεθόδους γραμμικού προγραμματισμού και βελτιστοποίησης. Ωστόσο, η γραμμικότητα και η βεβαιότητα, χαρακτηριστικά των προβλημάτων για την επίλυση των οποίων ενδείκνυται η χρήση γραμμικού προγραμματισμού και βελτιστοποίησης, δεν αποτελούν ιδιότητες της κυκλοφορίας.

1.2. Σκοπός διπλωματικής εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη προτύπου Ενισχυτικής Μάθησης για τη βελτιστοποίηση των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας σε αστικά σηματοδοτούμενα δίκτυα, με στόχο τη βελτίωση των κυκλοφοριακών χαρακτηριστικών του δικτύου. Βασική ικανότητα των πρακτόρων ενισχυτικής μάθησης είναι να δρουν σε αβέβαια περιβάλλοντα και χωρίς να υπάρχει απαραίτητα ένα πλήρες μοντέλο. Το γεγονός ότι μπορούν να λαμβάνουν αποφάσεις βάσει παρατηρήσεων και να προβλέπουν άγνωστες καταστάσεις τους καθιστά ιδανικούς για την επίλυση κυκλοφοριακών ζητημάτων.

Για την ανάπτυξη του μοντέλου βελτιστοποίησης επιλέγεται έπειτα από δοκιμές ο αλγόριθμος Proximal Policy Optimization (PPO). Η μεθοδολογία εφαρμόζεται σε οδικό τμήμα τριών κόμβων και η αποτελεσματικότητά της αξιολογείται μέσω προσομοίωσης σε κυκλοφοριακά σενάρια πραγματικών κυκλοφοριακών ζητήσεων, όπως παραδείγματος χάριν συνθήκες πρωινής-απογευματινής αιχμής, μεγάλων εκδηλώσεων και περιπτώσεων αναγκαστικής διακοπής της λειτουργίας μιας λωρίδας. Επίσης αξιολογείται και σε διαφορετικές γεωμετρικές συνθήκες του δικτύου, για να εξεταστεί η επιρροή των κόμβων στην αποδοτικότητα του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας.

1.3. Διάρθρωση διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία διαρθρώνεται σε έξι κεφάλαια. Στο πρώτο και παρόν κεφάλαιο γίνεται μία σύντομη αναφορά στο θέμα και τον στόχο της διπλωματικής εργασίας, καθώς και στις μεθόδους που θα χρησιμοποιηθούν.

Στο δεύτερο κεφάλαιο γίνεται εκτενής αναφορά στην υπάρχουσα βιβλιογραφία γύρω από τις εναλλασσόμενες λωρίδες κυκλοφορίας και επιπλέον αναλύονται βασικές έννοιες, τρόποι και υποδομές για την εφαρμογή τους σε πραγματικές συνθήκες.

Το τρίτο κεφάλαιο αφιερώνεται στην μεθοδολογική προσέγγιση του θέματος και για λόγους πληρότητας στην θεωρητική περιγραφή των μαθηματικών εργαλείων και αλγορίθμων ενισχυτικής μάθησης που θα χρησιμοποιηθούν για την επίλυση του προβλήματος της εργασίας, καθώς και των βασικών αρχών της προσομοίωσης κυκλοφορίας.

Στο τέταρτο κεφάλαιο παρουσιάζεται αναλυτικά η δομή του προβλήματος και η ανάπτυξη του μοντέλου εναλλαγής των λωρίδων.

Το πέμπτο κεφάλαιο αφορά την αξιολόγηση του μοντέλου που αναπτύχθηκε στο τέταρτο κεφάλαιο και η εξέταση της επιρροής του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας στο οδικό δίκτυο.

Τέλος, στο έκτο κεφάλαιο συνοψίζεται η πορεία της διπλωματικής εργασίας, παρουσιάζονται τα βασικά συμπεράσματα της και δίνονται προτάσεις για περαιτέρω έρευνα.

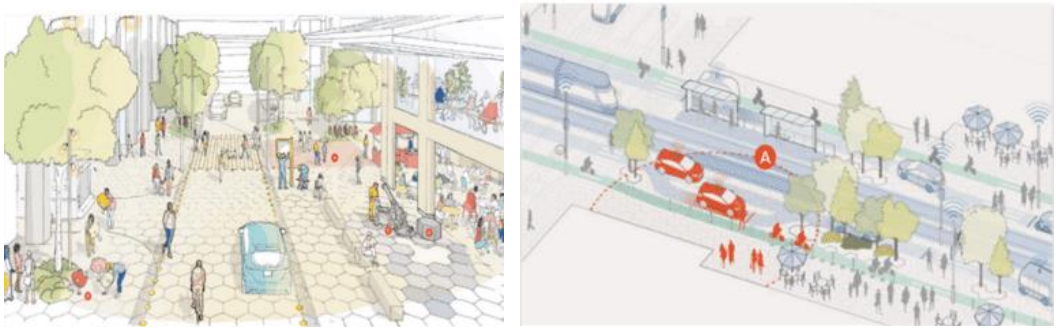
Κεφάλαιο 2: Βιβλιογραφική Ανασκόπηση

2.1. Εισαγωγή

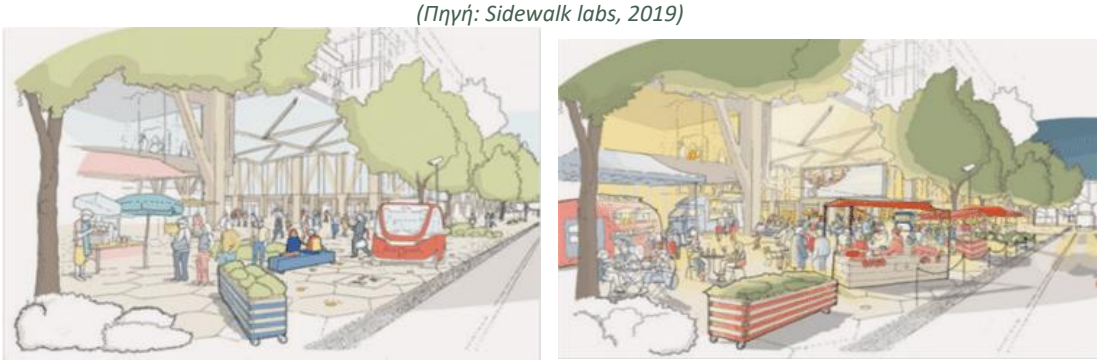
Στο παρόν κεφάλαιο παρουσιάζεται η βιβλιογραφία πάνω στην οποία βασίστηκε η έρευνα της παρούσας διπλωματικής εργασίας. Συγκεκριμένα, πέρα από την υπάρχουσα έρευνα για τις δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας, οι οποίες είναι και το κύριο θέμα της εργασίας, γίνεται αναφορά και στην γενικότερη δυναμική διαχείριση του αστικού χώρου, στα συνδεδεμένα οχήματα και στις υποδομές για την εφαρμογή των εναλλασσόμενων λωρίδων και την οδική ασφάλεια.

2.2. Δυναμική διαχείριση αστικού χώρου

Ο όρος δυναμική διαχείριση του αστικού χώρου αναφέρεται στην πρακτική της αναδιαμόρφωσης και του επαναπροσδιορισμού των αστικών χώρων ως απάντηση στις μεταβαλλόμενες ανάγκες και απαιτήσεις. Αυτή η προσέγγιση αναγνωρίζει ότι οι ανάγκες μιας πόλης και των κατοίκων της μπορούν να αλλάξουν γρήγορα και απρόβλεπτα και ότι ο παραδοσιακός πολεοδομικός σχεδιασμός και η χωροθέτηση μπορεί να μην είναι πάντα σε θέση να συμβαδίσουν με αυτές. Η δυναμική κατανομή αστικών χώρων μπορεί να περιλαμβάνει την προσωρινή ή μόνιμη αναδιαμόρφωση δρόμων, πεζοδρομίων και δημόσιων χώρων για την υποστήριξη διαφορετικών χρήσεων, όπως για παράδειγμα υπαίθριων αγορών, στάθμευσης και ποδηλατοδρόμων, όπως φαίνεται στην Εικόνα 2 και Εικόνα 3. Αυτή η προσέγγιση μπορεί να είναι ιδιαίτερα σημαντική σε πυκνοκατοικημένες αστικές περιοχές όπου ο χώρος είναι σε έλλειψη και απαιτείται ευελιξία για να ικανοποιηθούν οι διαφορετικές ανάγκες όλων των κοινοτήτων και ενδιαφερόμενων. *(ITF, 2022)*



Εικόνα 2: Εφαρμογή δυναμικής διαχείρισης σε χώρο πολλαπλών χρήσεων και σε συγκοινωνιακό δίκτυο (shared space) (Πηγή: Sidewalk labs, 2019)



Εικόνα 3: Παράδειγμα ώρας αιχμής που χρησιμοποιείται ο χώρος για στάθμευση οχημάτων και έπειτα ως αγορά (Πηγή: Sidewalk labs, 2019)

2.3. Εναλλασσόμενες λωρίδες κυκλοφορίας

Η ολοένα αυξανόμενη κυκλοφορία οδηγεί τις σύγχρονες πόλεις σε ορθότερη διαχείριση του περιορισμένου χώρου του αστικού ιστού. Η βελτίωση των δημόσιων συγκοινωνιών και η προώθηση της χρήσης εναλλακτικών οχημάτων όπως τα ποδήλατα και τα ηλεκτρικά πατίνια είναι ζωτικής σημασίας για τη μείωση του αριθμού των αυτοκινήτων στους δρόμους. Ωστόσο, παρά τη συμφόρηση οι συγκοινωνιακές υποδομές υπολειτουργούν για μεγάλα διαστήματα κατά τη διάρκεια μιας ημέρας λόγω της μεγάλης και ασταθούς διακύμανσης του φόρτου στις δύο κατευθύνσεις των οδικών αρτηριών. Για αυτό τον λόγο, πέρα από τα παραπάνω μέτρα, σημαντική στη μείωση της συμφόρησης είναι και η συμβολή μέτρων όπως η εφαρμογή έξυπνης δρομολόγησης των μέσων μαζικής μεταφοράς, ο έξυπνος προγραμματισμός της σηματοδότησης και οι αλλαγές στις υποδομές, όπως η εφαρμογή αναστρέψιμων λωρίδων κυκλοφορίας.

Η τεχνική των αναστρέψιμων λωρίδων κυκλοφορίας ξεκίνησε στις Η.Π.Α. τη δεκαετία του 1920 και έκτοτε εφαρμόστηκε και σε άλλες χώρες για τη διαχείριση της κυκλοφορίας. Αποτελούν μία σημαντική καινοτομία επειδή μπορούν να αυξήσουν σημαντικά την κυκλοφοριακή ικανότητα των οδών, ενώ συχνά απαιτούν μικρή επένδυση σε υποδομές οδοστρώματος και ελέγχου. Ο θεμελιώδης στόχος τους είναι η εκμετάλλευση των ανεπαρκώς χρησιμοποιούμενων λωρίδων επαναπροσανατολίζοντας την κατεύθυνση της ροής της κυκλοφορίας προς την κατεύθυνση με τον μεγαλύτερο φόρτο, αυξάνοντας έτσι τη συνολική χωρητικότητα του οδοστρώματος (*Wolshon et al., 2006*). Έχουν χρησιμοποιηθεί κυρίως για την αύξηση της κυκλοφοριακής ικανότητας στις παρακάτω περιπτώσεις:

- Σε περιόδους ώρας αιχμής
- Σε προγραμματισμένες μεγάλες εκδηλώσεις
- Σε περιόδους προσωρινών οδικών και κατασκευαστικών έργων
- Σε περιπτώσεις έκτακτων συμβάντων

Στην Ελλάδα το μέτρο των αναστρέψιμων λωρίδων εφαρμόζεται πολύ συχνά στις περιοχές των διοδίων.

Παρά τα οφέλη των αναστρέψιμων λωρίδων κυκλοφορίας η εφαρμογή τους αυτή τη στιγμή βασίζεται στην παρατήρηση ότι στις περισσότερες περιπτώσεις, υπάρχουν δύο διακριτές αιχμές στη ζήτηση της κυκλοφορίας, μία το πρωί και μία το απόγευμα. Με βάση αυτά τα μοτίβα, οι διαχειριστές κυκλοφορίας αποφασίζουν να αλλάξουν την κατεύθυνση των αναστρέψιμων λωρίδων για ένα σταθερό χρονικό διάστημα. Μπορεί να είναι μία ή πολλές ώρες, ανάλογα με τη διάρκεια της ώρας αιχμής. Όταν η κυκλοφορία συμπεριφέρεται όπως αναμένεται, οι αναστρέψιμες λωρίδες λειτουργούν αποτελεσματικά. Ωστόσο, η κυκλοφορία σε μια μεγάλη πόλη είναι κομμάτι ενός μεγαλύτερου πολύπλοκου συστήματος. Πολλοί παράμετροι και οι αλληλεπιδράσεις τους καθορίζουν τα μοτίβα ροής της κυκλοφορίας. Κατά συνέπεια, η κυκλοφοριακή συμφόρηση είναι ένα εξαιρετικά απρόβλεπτο φαινόμενο. Είναι δύσκολο να προβλεφθούν οι διακυμάνσεις της ροής της κυκλοφορίας σε μικρές χρονικές κλίμακες, επειδή θα μπορούσαν να επηρεαστούν από πολλά φαινόμενα όπως ένα ατύχημα, έναν κλειστό δρόμο ή ελαττωματικούς σηματοδότες χιλιόμετρα μακριά *(Pérez-Méndez et al., 2021)*. Άρα, γίνεται αντιληπτό ότι οι συμβατικές στατικά εναλλασσόμενες λωρίδες κυκλοφορίας δεν μπορούν να λειτουργήσουν με τον βέλτιστο τρόπο. Ιδιαίτερα σε ακραία

σενάρια όπου η ζήτηση κυκλοφορίας είναι αντίθετη από την αναμενόμενη, το αποτέλεσμα είναι να επικρατούν χειρότερες επιδόσεις από την κατάσταση χωρίς καμία εναλλασσόμενη λωρίδα.

Η ανάπτυξη του τομέα των δεδομένων και της επεξεργασίας δεδομένων σε πραγματικό χρόνο, καθώς και οι καινοτομίες στον κλάδο των συνδεδεμένων οχημάτων αποτελούν απάντηση στο παραπάνω πρόβλημα, καθώς πλέον οι στατικά εναλλασσόμενες λωρίδες μπορούν να μετατραπούν σε δυναμικά εναλλασσόμενες, με την επεξεργασία δεδομένων πραγματικού χρόνου για την επιλογή του καταλληλότερου διαμοιρασμού των λωρίδων.

2.4. Συνδεδεμένα οχήματα

Ο όρος συνδεδεμένα οχήματα αναφέρεται σε εφαρμογές, υπηρεσίες και τεχνολογίες που συνδέουν ένα όχημα με το περιβάλλον του*(Uhlemann, 2015)*. Τα συνδεδεμένα οχήματα είναι συμβατικά οχήματα με οδηγό τα οποία όμως είναι ενισχυμένα με συστήματα τηλεματικής, τα οποία τους δίνουν την δυνατότητα να επικοινωνούν με κοντινά τους οχήματα και με τις οδικές υποδομές. Οι πιο σημαντικές μορφές συνδεσιμότητας αφορούν σε όχημα με όχημα (vehicle to vehicle, V2V), όχημα με υποδομές (vehicle to infrastructure, V2I) και σε μία τελική μορφή συνδεσιμότητας του οχήματος με όλο το κοντινό του περιβάλλον (vehicle to all, V2X), όπως φαίνεται στην Εικόνα 4.

Εικόνα 4: Οδικό δίκτυο συνδεδεμένων οχημάτων
(Πηγή: The Future of Transportation Part 1)

Στην συνδεσιμότητα από όχημα σε όχημα (V2V), τα οχήματα εκπέμπουν ένα βασικό μήνυμα ασφαλείας που περιλαμβάνει πληροφορίες όπως η ταχύτητα, η κατεύθυνση και η τοποθεσία του οχήματος. Άλλα παρόμοια εξοπλισμένα οχήματα θα μπορούσαν να λάβουν αυτές τις εκπομπές, έτσι ώστε, συνεργατικά, να αποφεύγονται ατυχήματα. Οι εφαρμογές περιλαμβάνουν προειδοποίηση πριν από πιθανή σύγκρουση, προειδοποίηση για αλλαγή λωρίδας, συνεργατική προσαρμογή διαδρομής οχημάτων, βελτίωση ορατότητας, προειδοποίηση τυφλού σημείου, προειδοποίηση οδηγού που κινείται σε λάθος κατεύθυνση, υποβοήθηση κίνησης σε διασταυρώσεις, προειδοποίηση

οδικής κατάστασης βάσει οχημάτων και αναμετάδοση επικοινωνίας σε περίπτωση έκτακτης ανάγκης (*Abdelkader et al., 2021)*

Στην συνδεσιμότητα από όχημα σε οδική υποδομή (V2I) η ασφάλεια ενισχύεται μέσω της επικοινωνίας με αισθητήρες και άλλους εξοπλισμούς που είναι εγκατεστημένοι στο ίδιο το οδόστρωμα, καθώς και στους σηματοδότες, σε σήματα κυκλοφορίας όπως πινακίδες υποχρεωτικής στάσης, ζώνες εργασίας ή σχολείων και διαβάσεις πεζών και σιδηροδρομικές διαβάσεις.

Ωστόσο, πέρα από τα οφέλη που παρουσιάζουν τα συνδεδεμένα οχήματα στον τομέα της οδικής ασφάλειας μπορούν να λειτουργήσουν καταλυτικά και σε θέματα διαχείρισης της κυκλοφορίας. Αυτό γίνεται με την κοινή χρήση πληροφοριών παρακολούθησης της κυκλοφορίας, που βοηθούν τους οδηγούς να επαναδρομολογήσουν τον προορισμό τους και τους συγκοινωνιολόγους μηχανικούς να βελτιστοποιούν τον προγραμματισμό των φωτεινών σηματοδοτών, μειώνοντας έτσι την κυκλοφοριακή συμφόρηση.

Άρα μπορεί να γίνει εύκολα αντιληπτό ότι σε ένα τέτοιο συνδεδεμένο περιβάλλον η εφαρμογή δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας ευνοείται, καθώς η συνδεσιμότητα δίνει τη δυνατότητα στο έξυπνο σύστημα μεταφορών να γνωρίζει την κατάσταση που επικρατεί στο δίκτυο και μετριάζει τη συμφόρηση μέσω της κατάλληλης στρατηγικής κατανομής χώρου και της κοινοποίησης των αλλαγών στη διαμόρφωση του δρόμου ή του δικτύου στα οχήματα που είναι συνδεδεμένα σε αυτό. Με αυτόν τον τρόπο, ο αστικός χώρος μπορεί να αξιοποιηθεί στο μέγιστο των δυνατοτήτων του.

2.5. Δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας

Λόγω των τεχνολογικών εξελίξεων, που παρουσιάστηκαν στα προηγούμενα κεφαλαία, τη δεκαετία του 2010 ξεκίνησαν και τα πρώτα ερευνητικά βήματα στον τομέα των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας. Οι *Hausknecht et al. (2011)* χρησιμοποίησαν για πρώτη φόρα μεθόδους γραμμικού προγραμματισμού (linear programming) και δι-επίπεδης βελτιστοποίησης (bi-level optimization) για να υπολογίσουν τη βέλτιστη διαμόρφωση εναλλασσόμενων λωρίδων κυκλοφορίας, ώστε να πετύχουν τη μέγιστη κυκλοφοριακή ροή. Τα πειράματα και για τις δύο μεθόδους αξιολογήθηκαν σε δίκτυο 10x10 τύπου πλέγματος (Manhattan grid). Από τα αποτελέσματα της έρευνας προέκυψε ότι ενώ και τα δύο μοντέλα βρίσκουν έγκυρες λύσεις, η λύση της δι-επίπεδης βελτιστοποίησης είναι πιο ρεαλιστική σε ένα πραγματικό δίκτυο κυκλοφορίας, επειδή αυτό το μοντέλο ενσωματώνει πολλές πτυχές της πραγματικής κυκλοφορίας, όπως η συμφόρηση, η ισορροπία του δικτύου και τα όρια ταχύτητας. Παρόλα αυτά, το μοντέλο της δι-επίπεδης βελτιστοποίησης μπορεί να εγγυηθεί μόνο κατά προσέγγιση λύσεις, ενώ με τον γραμμικό προγραμματισμό υπολογίζεται η βέλτιστη λύση. Τελικά, παρατηρήθηκε μία αύξηση 72% της αποδοτικότητας του δικτύου από την εφαρμογή των εναλλασσόμενων λωρίδων.

Οι *Alhajyaseen et al. (2017)* χρησιμοποίησαν και πάλι μεθόδους βελτιστοποίησης για να επιλύσουν αυτή τη φορά το πρόβλημα του διαμοιρασμού των εναλλασσόμενων λωρίδων κυκλοφορίας σε όλες τις προσβάσεις ενός μεμονωμένου κόμβου, καθώς και το πρόβλημα της σηματοδότησης του. Η ανάλυση αυτή ανέδειξε σημαντικές βελτιώσεις στην απόδοση των διασταυρώσεων με κατά μέσο όρο μειώσεις των καθυστερήσεων που φτάνουν έως και το 82%.

Οι *Lu et al. (2018)* θεώρησαν ότι το πρόβλημα του βέλτιστου διαμοιρασμού των εναλλασσόμενων λωρίδων σε έναν σηματοδοτούμενο κόμβο είναι πρόβλημα ισορροπίας Nash ανάμεσα στους χρήστες της οδού και τους ελεγκτές κυκλοφορίας και πάνω σε αυτή τη θεώρηση ανέπτυξαν ένα μοντέλο δι-επίπεδης βελτιστοποίησης. Το μοντέλο αυτό έπειτα εξετάστηκε σε δύο σενάρια. Το πρώτο χρησιμοποιεί ένα δίκτυο 4 διαδρομών που ενώνουν δύο σημεία και η αξιολόγηση έδειξε ότι παρά τη χρήση διαφορετικών κυκλοφοριακών φόρτων και προς τις δύο κατευθύνσεις, επιτυγχάνεται μόνο ένας βέλτιστος διαμοιρασμός λωρίδων. Στο δεύτερο σενάριο, το οποίο αφορά οδικό δίκτυο 25 κόμβων, φαίνεται ότι η βέλτιστη διαμόρφωση του δικτύου εξαρτάται από μία μετρική, η οποία αφορά την αντίληψη του χρήστη για την δυσκολία του δικτύου να αναλάβει την κυκλοφορία.

Μία άλλη προσέγγιση εξέτασαν οι *Mao et al.(2020)*, με δεδομένο ένα περιβάλλον συνδεδεμένων οχημάτων ανέπτυξαν μοντέλο, το οποίο δίνει τη βέλτιστη διαμόρφωση του δικτύου βασιζόμενο στον λόγο φόρτου κυκλοφορίας προς τη χωρητικότητα σε οχήματα των λωρίδων μιας κατεύθυνσης από την επεξεργασία δεδομένων σε πραγματικό χρόνο.

Τέλος, οι *Pérez-Méndez et al.(2021)* ανέπτυξαν ένα μοντέλο βέλτιστου διαμοιρασμού των εναλλασσόμενων λωρίδων κυκλοφορίας βασιζόμενοι στο πρότυπο ροής Cellular Automata και στην επεξεργασία δεδομένων πραγματικού χρόνου.

2.6. Τρόποι υλοποίησης και υποδομές

Η ανάπτυξη της τεχνολογίας και τα έξυπνα συστήματα οδικών υποδομών επιτρέπουν τη σταδιακή εισαγωγή των δυναμικής διαχείρισης χώρου και των δυναμικά εναλλασσόμενων λωρίδων στους σύγχρονους αστικούς ιστούς. Υπάρχουν ήδη ορισμένα παραδείγματα διασφάλισης δυναμικών αστικών χώρων στο σχεδιασμό του οδοστρώματος και στη διαχείριση της κυκλοφορίας. Μερικές από αυτές τις εφαρμογές αναλύονται σε αυτήν την ενότητα και εξηγείται πώς θα μπορούσαν να χρησιμοποιηθούν για δυναμική διαχείριση του οδικού χώρου.

Στην Εικόνα 5 φαίνονται αυτόματες ανυψωτικές κολώνες, οι οποίες χρησιμοποιούνται για να περιορίσουν την πρόσβαση των αυτοκινήτων σε τοπικούς δρόμους με υψηλή ζήτηση για μετακίνηση πεζών και ποδηλάτων ή για να επιτρέπουν μόνο στους κατοίκους να εισέλθουν σε μια συγκεκριμένη ζώνη. Αυτό το εργαλείο μπορεί να είναι ζωτικής σημασίας για την εφαρμογή προσαρμόσιμων αστικών χώρων, όχι μόνο σε τοπικές και οικιστικές περιοχές. Επιπλέον, μπορεί να συμβάλλει και στη διαχείριση χώρου για διαφορετικούς τρόπους μεταφοράς σε διαφορετικές περιόδους της ημέρας, της εβδομάδας ή του χρόνου. Για παράδειγμα, αυτό θα μπορούσε να επεκταθεί σε πιο πολυσύχναστα περιβάλλοντα (κύριες οδικές αρτηρίες), όπου η κυκλοφορία επιτρέπεται κατά τη διάρκεια της ημέρας και περιορίζεται το βράδυ για να παρέχεται περισσότερος χώρος για δραστηριότητες όπως αγορές ή εκδηλώσεις.

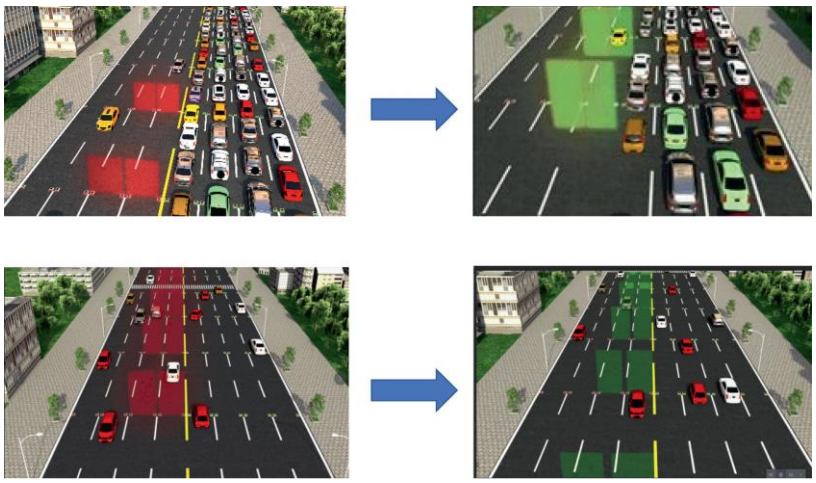
Εικόνα 5:Αυτόματες ανυψωτικές κολώνες στο Πόρτο και στη Λισαβόνα
(Πηγή: Valenca et al.)

Επιπλέον, οι πινακίδες ελέγχου λωρίδας ή οι πινακίδες μεταβλητών μηνυμάτων μπορούν να ενημερώσουν τον χρήστη για τις τρέχουσες και μελλοντικές δυναμικές αλλαγές του χώρου του δρόμου. Ο φωτισμός με LED στο οδόστρωμα μπορεί επίσης να χρησιμοποιηθεί για να υποδείξει τη χρήση του δημόσιου χώρου με αλλαγές στο χρώμα του. Για παράδειγμα, λεωφορειολωρίδες μπορούν να χρησιμοποιήσουν τον φωτισμό του οδοστρώματος για να ενημερώσουν πότε τα οχήματα μπορούν να χρησιμοποιήσουν (πράσινο χρώμα) ή όχι (κόκκινο χρώμα) τη λωρίδα λεωφορείων. Το πλεονέκτημα αυτών των εργαλείων είναι ότι είναι εύκολης εφαρμογής και χαμηλού κόστους και χρησιμοποιούνται ήδη σε πολλές πόλεις. Έτσι, η δυναμική κατανομή οδικού χώρου μπορεί να χρησιμοποιήσει υπάρχοντες και απλές υποδομές για την υλοποίησή της σε αρχικό επίπεδο. Παραδείγματα αυτών των μέτρων φαίνονται στην Εικόνα 6.

Εικόνα 6:Πινακίδες μεταβλητών μηνυμάτων και φώτα LED σε οδόστρωμα
(Πηγή: Pinterest, Smart City Streets)

Συγκεκριμένα για τις δυναμικά εναλλασσόμενες λωρίδες μπορεί να εφαρμοστεί μία τεχνική σηματοδότησης με έξυπνο φωτισμό LED του οδοστρώματος, ο οποίος να αποτελείται από τρία φωτεινά χρώματα, κόκκινο, κίτρινο και πράσινο και να έχει τη δυνατότητα να αναβοσβήνει. Το κόκκινο χρώμα που αναβοσβήνει σημαίνει υπενθύμιση για έξοδο από τη λωρίδα ενώ το στατικά κόκκινο απαγόρευση κυκλοφορίας, το πράσινο που αναβοσβήνει σημαίνει ότι η λωρίδα πρόκειται να ανοίξει και το στατικά πράσινο επιτρέπει την κυκλοφορία, η κίτρινη σηματοδότηση υποδεικνύει τη ζώνη ασφαλείας, πράγμα που σημαίνει ότι τα οχήματα πρέπει να απομακρυνθούν το συντομότερο δυνατό και προειδοποιεί τους οδηγούς ότι υπάρχει απαγορευμένη ζώνη μπροστά.

Παράλληλα με τον φωτισμό LED του οδοστρώματος, προτείνεται και μια ιδέα εικονικής κουρτίνας, που σχηματίζεται από προβολή φωτός (light curtain wall), όπως φαίνεται στην Εικόνα 7. Το φωτεινό παραπέτασμα μπορεί να συνεργαστεί με τα φώτα δρόμου για να παρέχει οδηγίες και υπενθυμίσεις και για να διασφαλίσει ότι τα οχήματα μπορούν να αλλάξουν με ασφάλεια λωρίδα και να οδηγήσουν. Προς το παρόν, η τεχνολογία φωτοκουρτίνας δεν πληροί τις απαιτήσεις για πρακτική εφαρμογή σε δρόμους. Σύμφωνα με την τεχνική έρευνα και την πειραματική ανάλυση, η τεχνολογία αυτή θα μπορούσε να εφαρμοστεί στους δρόμους ως ένδειξη οδικής σηματοδότησης στο μέλλον.



Εικόνα 7: Εφαρμογή φωτοκουρτίνας κατά την εναλλαγή λωρίδας
(Πηγή:Mao et al.)

2.7. Συμπεράσματα βιβλιογραφικής ανασκόπησης

Από την ανάλυση της βιβλιογραφίας, η πλειονότητα των προσεγγίσεων για τις εναλλασσόμενες λωρίδες χρησιμοποιούν μαθηματικούς αλγόριθμους βελτιστοποίησης για να βρουν την καλύτερη στρατηγική εναλλαγής. Με αυτόν τον τρόπο όμως αγνοείται ο αντίκτυπος της εναλλαγής στην κυκλοφορία. Αυτός ο αντίκτυπος προκύπτει από τους ελιγμούς για αλλαγή λωρίδας από τα οχήματα που ακολουθούν την αλλαγή της κατεύθυνσης μιας λωρίδας, καθώς και από την προσωρινή διακοπή της κυκλοφορίας στη λωρίδα για αποφυγή συγκρούσεων. Επιπλέον, οι αλγόριθμοι βελτιστοποίησης παρουσιάζουν μειονεκτήματα στην προσαρμογή τους σε αλλαγές στις συνθήκες του δικτύου. Αυτό συμβαίνει γιατί η γραμμικότητα και η βεβαιότητα, χαρακτηριστικά των προβλημάτων για την επίλυση των οποίων ενδείκνυται η χρήση γραμμικού προγραμματισμού και βελτιστοποίησης, δεν αποτελούν ιδιότητες της κυκλοφορίας και σημαίνει επαναϋπολογισμός του μοντέλου για την προσαρμογή σε διαφορετικά δίκτυα.

Κεφάλαιο 3: Μεθοδολογική προσέγγιση

3.1. Διάγραμμα ροής εργασιών

Στο παρόν κεφάλαιο παρουσιάζεται η μεθοδολογία που χρησιμοποιήθηκε όσον αφορά στη δομή και στην επίλυση του προβλήματος των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας. Στο Διάγραμμα 2 φαίνεται η ροή των εργασιών που ακολουθήθηκε για την επίτευξη της εκπαίδευσης ενός μοντέλου ενισχυτικής μάθησης ικανό να διαχειρίζεται αποτελεσματικά τη διαμόρφωση των λωρίδων κυκλοφορίας του οδικού δικτύου που δημιουργήθηκε.

Διάγραμμα 2: Γενικό διάγραμμα ροής εργασιών

Αρχικά, ορίζεται το πρόβλημα το οποίο επιχειρεί να λύσει η παρούσα εργασία καθώς και τα χαρακτηριστικά της κυκλοφορίας τα οποία προσπαθεί να βελτιώσει μέσα από την εφαρμογή του μοντέλου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας, όπως είναι η ταχύτητα των οχημάτων του δικτύου, ο χρόνος διαδρομής και οι εκπομπές καυσαερίων.

Έπειτα, παρουσιάζονται βασικές έννοιες και αλγόριθμοι ενισχυτικής μάθησης. Πρώτο βήμα, για την ανάπτυξη του μοντέλου εναλλαγής των λωρίδων κυκλοφορίας, αποτέλεσε η δομή ενός δικτύου, συγκεκριμένα η δημιουργία ενός οδικού τμήματος τριών κόμβων. Η κύρια αρτηρία αποτελείται από εφτά λωρίδες κυκλοφορίας εκ των οποίων οι δύο κεντρικές επιλέγεται να είναι

εναλλασσόμενες. Με αυτό τον τρόπο, οι πιθανές διαμορφώσεις του οδικού δικτύου είναι τρεις. Η σηματοδότηση του δικτύου επιλέγεται, για διευκόλυνση της κυκλοφορίας, να γίνει με αυτόματους σηματοδότες. Για την εκπαίδευση του μοντέλου δημιουργούνται εκατό διαφορετικά σενάρια ζήτησης, τα οποία διαφέρουν στον συνολικό φόρτο του δικτύου και στο ποσοστό του συνολικού φόρτου που κατανέμεται στις δύο κατευθύνσεις.

Στη συνέχεια, αναπτύσσεται ένας αλγόριθμος εναλλαγής των λωρίδων, ο οποίος λειτουργεί με την λογική ότι μία εναλλασσόμενη λωρίδα κυκλοφορίας πρέπει να εκκενωθεί από όλα τα οχήματα που την καταλαμβάνουν προτού δοθεί στην αντίθετη κατεύθυνση.

Για την επίτευξη του τελικού σκοπού, έπειτα καταστρώνεται το πρόβλημα της εναλλαγής λωρίδας σε περιβάλλον ενισχυτικής μάθησης. Ως κατάσταση του περιβάλλοντος λαμβάνονται τα κυκλοφοριακά χαρακτηριστικά που παρατηρούνται στο δίκτυο ανά 300 δευτερόλεπτα. Ο πράκτορας μετά από κάθε κατάσταση καλείται να επιλέξει μία ενέργεια, δηλαδή μία από τις τρεις διαμορφώσεις των λωρίδων κυκλοφορίας, οι οποίες έχουν οριστεί. Για την ανταμοιβή του πράκτορα χρησιμοποιήθηκαν διαφορετικές μετρικές, με καλύτερα αποτελέσματα να δίνει ο μέγιστος μέσος χρόνος διαδρομής που παρατηρείται στις δύο κατευθύνσεις

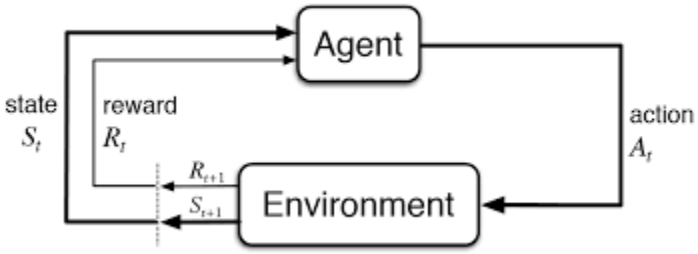
Για την εκπαίδευση του μοντέλου δοκιμάζονται δύο αλγόριθμοι, ο αλγόριθμος Deep Q-Learning (DQN) και ο αλγόριθμος Proximal Policy Optimization (PPO). Εξετάζεται επιπλέον η εκπαίδευση σε δύο στάδια, ένα στάδιο προ-εκπαίδευσης και ένα στάδιο εκπαίδευσης. Η διαφορά ανάμεσα στα δύο στάδια είναι στην ανταμοιβή που λαμβάνει ο πράκτορας μετά από κάθε ενέργεια. Στην προεκπαίδευση η ανταμοιβή έχει οριστεί με τέτοιο τρόπο ώστε να θεωρεί ως ιδανική τη διαμόρφωση που δίνει μία επιπλέον λωρίδα στην κατεύθυνση με το μεγαλύτερο φόρτο. Στη φάση της εκπαίδευσης ως ανταμοιβή λαμβάνεται ο μέγιστος μέσος χρόνος διαδρομής που παρατηρείται στις δύο κατευθύνσεις.

Τέλος, το τελικό μοντέλο αξιολογείται με βάση την αποδοτικότητα του στην βελτίωση των βασικών δεικτών απόδοσης, δηλαδή των κυκλοφοριακών χαρακτηριστικών που έχουν ορισθεί, σε φαινόμενα πραγματικής κυκλοφοριακής ζήτησης και διαφορετικές γεωμετρικές διαμορφώσεις του δικτύου. Στα επόμενα κεφάλαια και παραγράφους παρατίθεται αναλυτικά η μεθοδολογία που ακολουθήθηκε.

3.2. Θεωρητικό Υπόβαθρο

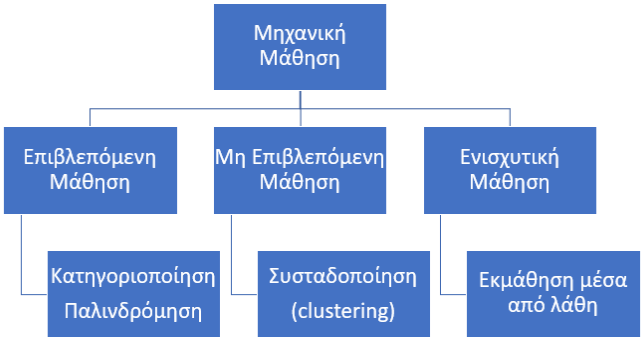
3.2.1 Βασικές έννοιες ενισχυτικής μάθησης

Η ενισχυτική μάθηση (reinforcement learning) είναι μέθοδος μηχανικής μάθησης κατά την οποία ο αλγόριθμος μαθαίνει να εκτελεί βέλτιστες ενέργειες μέσω αλληλεπίδρασης με το περιβάλλον στο οποίο δρα και τη βοήθεια ενός συστήματος ανταμοιβής (*Sutton & Barto, 2018*). Η ενισχυτική μάθηση αναφέρεται σε μια διαδικασία μάθησης, η οποία έχει επηρεαστεί από θεωρίες της νευρολογίας και της ψυχολογίας, οι οποίες εξηγούν τον τρόπο με τον οποίο οι οργανισμοί μαθαίνουν και αναπτύσσουν συμπεριφορές. Η συμπεριφοριστική ψυχολογία περιγράφει τον ανθρώπινο τρόπο μάθησης ως μια διαδικασία συλλογής εμπειριών μέσω αλληλεπίδρασης με το περιβάλλον. Κατά τη διάρκεια αυτής της διαδικασίας, οι εμπειρίες αξιολογούνται από το εσωτερικό σύστημα επιβράβευσης του εγκεφάλου και αποθηκεύονται στο βιολογικό νευρωνικό του σύστημα. Με παρόμοιο τρόπο στην ενισχυτική μάθηση, ένας πράκτορας αλληλεπιδρά με ένα περιβάλλον και προσπαθεί να προσαρμοστεί σε αυτό προκειμένου να λάβει τη μέγιστη ανταμοιβή. Αυτή η διαδικασία επεξηγείται και στην Εικόνα 8. Ο πράκτορας πρέπει να μάθει πώς να αντιδρά σε διάφορες καταστάσεις χρησιμοποιώντας ελάχιστα δεδομένα και πληροφορίες, με το μοναδικό ερέθισμα που λαμβάνει να είναι η ανταμοιβή από το περιβάλλον.



Εικόνα 8:Βασική ιδέα ενισχυτικής μάθησης
(Πηγή: Sutton & Barto, 2018)

Η ενισχυτική μάθηση διαφέρει από την επιβλεπόμενη μάθηση λόγω της απουσίας δεδομένων εκπαίδευσης. Στη θέση αυτών υπάρχει το σύστημα ανταμοιβής. Από την άλλη μεριά δεν μπορεί να χαρακτηρισθεί ως μη επιβλεπόμενη μάθηση, καθώς ο στόχος της δεν είναι η εύρεση κοινών δομών σε ένα σύνολο δεδομένων. Ο διαχωρισμός στις κατηγορίες της μηχανικής μάθησης φαίνεται στην Εικόνα 9. Συνεπώς, η ενισχυτική μάθηση αποτελεί μία ξεχωριστή κατηγορία μηχανικής μάθησης.



Εικόνα 9: Κατηγορίες μηχανικής μάθησης

Στην ενισχυτική μάθηση, η υποθετική οντότητα που εκτελεί ενέργειες σε ένα περιβάλλον για να κερδίσει κάποια ανταμοιβή (reward) ονομάζεται πράκτορας (agent). Το σενάριο, το οποίο πρέπει να αντιμετωπίσει ο πράκτορας, δηλαδή οτιδήποτε πέρα είναι πέρα από τον έλεγχό του ονομάζεται περιβάλλον (environment). Ο πράκτορας καλείται σε κάθε χρονική στιγμή t να προβεί σε μία ενέργεια (action) με στόχο να μπορέσει να αυξήσει την αμοιβή του. Κατάσταση (state) είναι η θέση του πράκτορα μέσα στο περιβάλλον σε ένα συγκεκριμένο χρονικό βήμα. Έτσι, κάθε φορά που ένας πράκτορας εκτελεί μια ενέργεια, το περιβάλλον δίνει στον πράκτορα αριθμητική ανταμοιβή (reward) και μια νέα κατάσταση, στην οποία έφτασε ο πράκτορας εκτελώντας αυτή την ενέργεια.

Συγκεντρωτικά, σε κάθε χρονικό βήμα t ο πράκτορας:

- Καλείται να επιλέξει μία ενέργεια A_t, τυχαία ή ελεγχόμενα
- Λαμβάνει μία ανταμοιβή R_t και
- Λαμβάνει μία παρατήρηση από το περιβάλλον O_t

Ως ιστορικό H_t ορίζεται μία τέτοια αλληλουχία ενεργειών A_t, ανταμοιβών R_t και παρατηρήσεων O_t, δηλαδή $H_t = A_1O_1R_1A_2O_2R_2...A_tO_tR_t$. Κάθε επόμενη κατάσταση του πράκτορα εξαρτάται από αυτό, δηλαδή ισχύει $S_t=f(H_t)$. Στην ενισχυτική μάθηση η αλληλεπίδραση του πράκτορα με το περιβάλλον του αποτελεί μία Μαρκοβιανή Διαδικασία Απόφασης (Markov Decision Process-MDP), δηλαδή για το μέλλον του συστήματος είναι αρκετή η γνώση πληροφοριών του προηγούμενου βήματος και όχι ολόκληρού του ιστορικού.

Οι ανταμοιβές είναι οι αριθμητικές τιμές που λαμβάνει ο πράκτορας, κατά την εκτέλεση κάποιας ενέργειας, σε μία κατάσταση στο περιβάλλον. Η αριθμητική τιμή μπορεί να είναι θετική ή αρνητική με βάση τις ενέργειες του πράκτορα. Στην ενισχυτική μάθηση, στόχος είναι η μεγιστοποίηση της σωρευτικής ανταμοιβής, όλες οι ανταμοιβές που λαμβάνει ο πράκτορας από το περιβάλλον. Αυτό το συνολικό άθροισμα της ανταμοιβής, που λαμβάνει ο πράκτορας από το περιβάλλον, ονομάζεται επιστροφή G_t (return). Η επιστροφή ξεκινώντας από βήμα t ορίζεται ως εξής:

$$G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \quad (3.1)$$

Όπου, γ ένας συντελεστής που ονομάζεται συντελεστής έκπτωσης (discount factor). Αυτός ο συντελεστής καθορίζει πόση σημασία πρέπει να δοθεί στην άμεση ανταμοιβή και τις μελλοντικές ανταμοιβές. Ένας συντελεστής έκπτωσης μπορεί να θεωρηθεί μια παράμετρος μάθησης, που ποικίλει από $0\leq\gamma\leq1$. Κάθε μελλοντική ανταμοιβή κατά τη χρονική στιγμή t μειώνεται κατά γ^{k-1}. Όσο το γ πλησιάζει το μηδέν, ο πράκτορας θεωρεί τις κοντινές, στην παρούσα κατάσταση, ανταμοιβές πολύ πιο πολύτιμες. Αντίθετα, αν το γ προσεγγίζει τη μονάδα, ο πράκτορας μεταβάλλεται, ώστε να μην συμπεριφέρεται άπληστα και να θεωρεί τις μελλοντικές ανταμοιβές εξίσου σημαντικές, ωστόσο γ ίσο με τη μονάδα μπορεί να οδηγήσει στο άπειρο. Συνήθως λαμβάνει τιμές ανάμεσα στο 0,2 και το 0,8.

Η συνάρτηση αξίας (value function) v(s) είναι μια συνάρτηση πρόβλεψης της μελλοντικής ανταμοιβής και χρησιμοποιείται για την εκτίμηση των καταστάσεων ως καλές ή κακές και την επιλογή

επόμενων ενεργειών. Η αξία μίας κατάστασης είναι η μέση τιμή του συνόλου των ανταμοιβών που αναμένουμε όπως φαίνεται στην παρακάτω εξίσωση, ξεκινώντας από την κατάσταση του χρονικού βήματος t:

$$v(s) = E[R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots | S_t = s], 0 < \gamma < 1 \quad (3.2),$$

όπου γ ο συντελεστή έκπτωσης.

Η διαφορά της συνάρτησης αξίας με την ανταμοιβή είναι ότι οι ανταμοιβές είναι άμεσες ενώ οι αξίες δείχνουν την μακροπρόθεσμη επίδραση του συνδυασμού κατάστασης.

Ως πολιτική (policy) ορίζεται η αντίληψη του πράκτορα για την κατάσταση στην οποία βρίσκεται και την ενέργεια που θα λάβει, δηλαδή είναι μία απεικόνιση από καταστάσεις σε ενέργειες. Μια πολιτική μπορεί να είναι ντετερμινιστική, να ακολουθεί δηλαδή τη λογική ενός πίνακα αναζήτησης όταν διασφαλίζεται η εκτέλεση μιας ενέργειας, ή στοχαστική, όταν εξετάζεται μια συγκεκριμένη πιθανότητα. Ωστόσο, σε κάθε πρόβλημα ενισχυτικής μάθησης υπάρχει μία τουλάχιστον βέλτιστη πολιτική, η οποία είναι καλύτερη από όλες τις άλλες, ή τουλάχιστον ίση με όλες, και συμβολίζεται με π*. Μία πολιτική π είναι καλύτερη ή ίση με μία πολιτική π′ αν και μόνο αν η μέση τιμή της αναμενόμενης ανταμοιβής τής είναι μεγαλύτερη ή ίση από της π′ για κάθε κατάσταση, δηλαδή αν $v_\pi(s) \geq v_{\pi'}(s)$.

Με την ίδια λογική ορίζεται η συνάρτηση ενέργειας- αξίας (action-value function) q(s,α) ως μία συνάρτηση, η οποία εκφράζει πόσο καλή είναι μία ενέργεια α στην κατάσταση s.

$$q(s,a) = E[G_t | S_t = s, A_t = a] \quad (3.3)$$

Και η βέλτιστη συνάρτηση αξίας q*, η οποία είναι η συνάρτηση, που επιφέρει τη μέγιστη ανταμοιβή ακολουθώντας οποιαδήποτε πολιτική και κάνοντας ενέργεια α.

$$q^*(s,a) = max_\pi v_\pi(s,a) \quad (3.4)$$

Ως μοντέλο (model) ορίζεται ο μηχανισμός που προβλέπει πως θα αποκριθεί το περιβάλλον στα επόμενα χρονικά βήματα, για παράδειγμα μπορεί να προβλέψει την ανταμοιβή και την επόμενη κατάσταση δεδομένης της κατάστασης του πράκτορα και μίας ενέργειας που θα επιλέξει. Ουσιαστικά, αποτελούν ένα τρόπο επίλυσης του περιβάλλοντος και οι πράκτορες που τα χρησιμοποιούν καλούνται model-based.

Τα περισσότερα προβλήματα ενισχυτικής μάθησης μπορούν να αναλυθούν σε ακολουθίες στις οποίες ένας πράκτορας αλληλεπιδρά με το περιβάλλον του για ένα πεπερασμένο αριθμό χρονικών βημάτων t=1,2,3…, μέχρι να φτάσει σε μια ορισμένη τελική κατάσταση, μετά την οποία γίνεται επαναφορά στην αρχική του κατάσταση. Κάθε τέτοια ακολουθία ονομάζεται επεισόδιο (episode). Οι ενέργειες που εκτελεί ο πράκτορας σε κάθε επεισόδιο δίνουν μία διαφορετική ακολουθία καταστάσεων και τελικώς μία διαφορετική συνολική ανταμοιβή. Στόχος κάθε επεισοδίου είναι ο αλγόριθμος να αυξάνει την ανταμοιβή του, δηλαδή να μαθαίνει όσο το δυνατόν καλύτερα με το πέρας των επεισοδίων.

3.2.2. Μαρκοβιανή Διαδικασία Απόφασης

Η μαρκοβιανή ιδιότητα αναφέρει ότι το μέλλον ενός συστήματος δεν εξαρτάται από τις παρελθοντικές του καταστάσεις αλλά μονάχα από την τρέχουσα κάθε φορά κατάσταση. Μαθηματικά αυτό εκφράζεται με την παρακάτω εξίσωση:

$$P[S_{t+1}|S_t] = P[S_{t+1}|S_1, \dots, S_t] \quad (3.5)\,,$$

Όπου S_t η τωρινή κατάσταση του συστήματος και S_{t+1} η επόμενη κατάσταση.

Όπως αναφέρθηκε και στην παράγραφο Υ.1.2. η ενισχυτική μάθηση υπακούει στην μαρκοβιανή ιδιότητα. Η μαρκοβιανή διαδικασία απόφασης είναι ένα μαθηματικό πλαίσιο που χρησιμοποιείται για τη μοντελοποίηση τέτοιων προβλημάτων, δηλαδή προβλημάτων λήψης αποφάσεων, όπου τα αποτελέσματα είναι εν μέρει τυχαία και εν μέρει ελεγχόμενα. Προέρχεται από τη θεωρία των διακριτών δυναμικών συστημάτων και είναι μία κλασσική μέθοδος αναπαράστασης προβλημάτων βελτιστοποίησης που περιλαμβάνουν ακολουθιακή λήψη αποφάσεων. Περιέχουν την έννοια της κατάστασης, των ενεργειών, της ανταμοιβής και του συμβιβασμού ανάμεσα στην άμεση και την μακροπρόθεσμη ανταμοιβή. Είναι επέκταση των διακριτών μαρκοβιανών αλυσίδων (Markov Chains) οι οποίες είναι στοχαστικές ακολουθίες καταστάσεων που χαρακτηρίζονται από πιθανότητες μετάβασης μεταξύ αυτών και από έλλειψη μνήμης.

Μία μαρκοβιανή διαδικασία απόφασης μπορεί να χαρακτηριστεί από μία αλυσίδα πέντε στοιχείων < S,A,P,R,γ>, όπου *(Sutton & Barto, 2018)*:

- S ένα διακριτό σύνολο καταστάσεων
- A ένα διακριτό σύνολο ενεργειών
- P ένας πίνακας καταστάσεων-μεταβάσεων (state transition matrix), $P^{\alpha}_{ss'}=P[S_{t+1}=s'|S_t=s]$, s, s'∈ S, που δείχνει την πιθανότητα του συστήματος σε κατάσταση s στον χρόνο t να φτάσει σε κατάσταση s' στον χρόνο t+1 όταν πραγματοποιεί ενέργεια α
- R μία συνάρτηση ανταμοιβής $R^{\alpha}_{s}=E[R_{t+1}|S_t=s, A_t=\alpha]$, s ∈ S, a ∈ A
- γ ένας συντελεστής έκπτωσης, $\gamma \in [0,1]$

Στις μαρκοβιανές διαδικασίες απόφασης η πολιτική είναι ο μηχανισμός, ο οποίος λαμβάνει αποφάσεις και παίρνει ενέργειες. Άρα, μία πολιτική π ορίζεται ως μία απεικόνιση από καταστάσεις σε πιθανότητα επιλογής ενέργειας και όχι απλά ενέργειας. Είναι δηλαδή μία συνάρτηση πυκνότητας πιθανότητας ενεργειών, δεδομένων των καταστάσεων:

$$\pi(a|s) = P(A_t = \alpha|S_t = s) \quad (3.6)$$

Επεκτείνοντας τα παραπάνω ορίζεται η συνάρτηση κατάστασης αξίας (state-value function):

$$v_\pi = E_\pi[G_t|S_t = s], \quad (3.7)$$

ακολουθώντας πολιτική π. Ως συνέπεια του ορισμού τους τα παρατηρήσιμα περιβάλλοντα αποτελούν μία μαρκοβιανή διαδικασία απόφασης. Οι πιθανότητες του πίνακα καταστάσεων-μεταβάσεων χαρακτηρίζουν πλήρως τις δυναμικές του περιβάλλοντος. Τα μη παρατηρήσιμα περιβάλλοντα αποτελούν Μερικώς Παρατηρήσιμη Μαρκοβιανή Διαδικασία Απόφασης (Partially Observable Markov Decision Process - POMDP). Ακόμα και τέτοια προβλήματα όμως μπορούν να

μετατραπούν σε μαρκοβιανές διαδικασίες απόφασης. Σχεδόν όλα τα προβλήματα ενισχυτικής μάθησης περιγράφονται από τη θεωρία των μαρκοβιανών διαδικασιών απόφασης.

3.2.3. Αλγόριθμοι ενισχυτικής μάθησης-Η περίπτωση του Q-learning

Σχεδόν όλοι οι αλγόριθμοι ενισχυτικής μάθησης προσπαθούν να προσεγγίσουν συναρτήσεις αξίας. Αυτές οι συναρτήσεις αξίας είναι συνήθως συνδεδεμένες με τις πολιτικές που ακολουθεί ο πράκτορας. Ο τρόπος που θα αλλάξει η πολιτική του πράκτορα ως αποτέλεσμα της εμπειρίας που αποκτά, καθορίζεται από τον εκάστοτε αλγόριθμο. Ο αλγόριθμος Q-learning είναι ένας τέτοιος αλγόριθμος.

3.2.3.1. Η εξίσωση Bellman

Επιπλέον, σημαντική για την κατανόηση του αλγορίθμου Q-learning είναι η εξίσωση Bellman, η οποία εκφράζει ένα βασικό χαρακτηριστικό των συναρτήσεων αξίας. Η αξία μίας κατάστασης εκφράζεται ως το άθροισμα της άμεσης ανταμοιβής και της μειωμένης ανταμοιβής της επόμενης κατάστασης, δηλαδή μαθηματικά ισχύει:

$$v(s) = [R_{t+1} + \gamma v(S_{t+1}) | S_t = s] \quad (3.8)$$

Οι βέλτιστες συναρτήσεις αξίας (v*,q*) υπακούν επίσης στην εξίσωση Bellman. Αυτό οδηγεί στον ορισμό της εξίσωσης βελτιστότητας Bellman (Bellman optimality equation), η οποία με λόγια αναφέρει ότι η αξία μιας κατάστασης υπό τη βέλτιστη πολιτική είναι ίση με τη μέση αναμενόμενη επιστροφή από την καλύτερη δυνατή ενέργεια από αυτή την κατάσταση και μαθηματικά εκφράζεται ως εξής:

$$v^* = max_a q^*(s, a) \quad (3.9)$$

3.2.4. Ο αλγόριθμος Q-learning

Ο αλγόριθμος Q-learning προτάθηκε από τον Chris Watkins το 1989. Είναι ένας χωρίς μοντέλο (model-free), με βάση την αξία (value-based) και εκτός πολιτικής (off-policy), ο οποίος βρίσκει την καλύτερη σειρά ενεργειών με βάση την τρέχουσα κατάσταση του πράκτορα. Αναλυτικότερα:

- Είναι αλγόριθμος χωρίς μοντέλο (model-free) που σημαίνει ότι μαθαίνει τις συνέπειες των ενεργειών του πράκτορα από την εμπειρία και όχι από την συνάρτηση ανταμοιβής.
- Είναι αλγόριθμος που βασίζεται στην αξία (value-based) εκπαιδεύει τη συνάρτηση αξίας, ώστε να αναγνωρίζει ποια κατάσταση είναι καλύτερη και να επιλέξει την κατάλληλη ενέργεια.
- Είναι αλγόριθμος εκτός πολιτικής (off-policy) που σημαίνει ότι αξιολογεί και ενημερώνει μια πολιτική που διαφέρει από την πολιτική που χρησιμοποιείται για την εκτέλεση μιας ενέργειας.

Ουσιαστικά, ο αλγόριθμος προσπαθεί να εκπαιδεύσει τη συνάρτηση Q(s,α) μιας άγνωστης μαρκοβιανής διαδικασίας απόφασης προσεγγίζοντας τη βέλτιστη συνάρτηση αξίας q^* απευθείας. Αυτό επιτυγχάνεται κοιτώντας ένα βήμα μπροστά τη φορά (one step look-ahead) και επιλέγοντας την ενέργεια που θα μεγιστοποιήσει την τιμή της συνάρτησης ενέργειας-αξίας q της επόμενης κατάστασης, ανεξάρτητα από την πολιτική που χρησιμοποιείται.

Όλο αυτό επιτυγχάνεται με τη χρήση ενός πίνακα Q[s,α] ως δομή δεδομένων στην οποία αποθηκεύονται οι τιμές q για όλους τους δυνατούς συνδυασμούς καταστάσεων και ενεργειών, οι οποίες ανανεώνονται με βάση την εξίσωση Bellman. Ο πράκτορας θα χρησιμοποιήσει έναν πίνακα Q[s,α] για να κάνει την καλύτερη δυνατή ενέργεια με βάση την αναμενόμενη ανταμοιβή για κάθε κατάσταση του περιβάλλοντός του. Στην αρχή της εκμάθησης ο πίνακας αυτός είναι κενός και στα πρώτα βήματα παίρνει τυχαίες τιμές καθώς ο πράκτορας προβαίνει σε τυχαίες ενέργειες. Κατά τη διάρκεια κάθε επεισοδίου, ο αλγόριθμος επιλέγει την ενέργεια που έχει τη μεγαλύτερη τιμή q για την εκάστοτε κατάσταση. Η ανταμοιβή που λαμβάνει εν συνεχεία χρησιμοποιείται για να ανανεώσει την αντίστοιχη τιμή στον πίνακα σύμφωνα με τη σχέση:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + a[R_{t+1} + \gamma max_{a'} Q(S_{t+1}, a') - Q(S_t, A_t)] \quad (3.10)$$

Η σχέση αυτή ονομάζεται κανόνας ενημέρωσης (update rule) και με βάση αυτή μετά το πέρας πολλών επεισοδίων ο πίνακας προσεγγίζει τη βέλτιστη συνάρτηση αξίας q*.

Στην παραπάνω εξίσωση ο όρος $R_{t+1}+\gamma max_{\alpha'}Q(S_{t+1}, \alpha')$ είναι το άθροισμα της ανταμοιβής που επιστρέφει το περιβάλλον και της μέγιστης τιμής q που δύναται να φέρει η επόμενη κατάσταση απομειωμένη από τον συντελεστή έκπτωσης γ και ονομάζεται στόχος (target) του αλγορίθμου. Αρχικά, αυτός ο όρος παίρνει λάθος τιμές και με το πέρας των καταστάσεων ενημερώνεται. Ο όρος $R_{t+1}+\gamma max_{\alpha'}Q(S_{t+1}, \alpha')- Q(S_t,A_t)$, δηλαδή η αφαίρεση της παλιάς εκτίμησης $Q(S_t,A_t)$ (old estimate) από τον στόχο ονομάζεται σφάλμα χρονικής διαφοράς (temporal difference error) και μειώνεται καθώς η παλιά εκτίμηση πλησιάζει τον στόχο. Ο συντελεστής α ονομάζεται ρυθμός μάθησης (learning rate) του κανόνα ενημέρωσης και βοηθά στο να μη χρησιμοποιούνται εκτιμήσεις οι οποίες πλέον δεν χρειάζονται, ώστε τελικά να μην εμφανίζονται μεγάλες ταλαντώσεις στον αλγόριθμο και να του επιτρέπει να συγκλίνει.

Ένας αλγόριθμος Q-learning συγκλίνει με πιθανότητα 1, αν και μόνο αν συναντήσει όλα τα ζεύγη (s,α). Για να διασφαλισθεί η σύγκλιση ακολουθείται μία πολιτική, η οποία ονομάζεται ε-greedy. Αυτή η πολιτική εξασφαλίζει μία ισορροπία ανάμεσα στην αναζήτηση και την εκμετάλλευση, αφού όλες οι πιθανές ενέργειες μ δοκιμάζονται με μη μηδενική πιθανότητα. Με πιθανότητα $1 - \varepsilon$ επιλέγεται η άπληστη ενέργεια και με πιθανότητα ε επιλέγεται μια τυχαία ενέργεια. Μαθηματικά αυτό εκφράζεται ως εξής:

$$\pi(\alpha \mid s) = \begin{cases} \frac{\varepsilon}{\mu} + 1 - \varepsilon, & \alpha\nu\ \alpha * = argmax_{\alpha \in A} Q(s, \alpha) \\ \frac{\varepsilon}{\mu}, & \alpha\lambda\lambda\iota\acute{\omega}\varsigma \end{cases} \quad (3.11)$$

Ο αλγόριθμός Q-learning είναι αποδοτικός σε μικρό συνδυασμό καταστάσεων-ενεργειών (s,α), όταν ο αριθμός αυτών των ζευγών είναι μεγάλος, η μνήμη και ο υπολογιστικός χρόνος που απαιτείται για την παραγωγή και ενημέρωση του πίνακα Q[s,α] και για να συγκλίνει ο αλγόριθμος καθιστούν τη μέθοδο μη αποδοτική. Επιπλέον, ένα άλλο πρόβλημα του αλγορίθμου αυτού είναι η έλλειψη της ικανότητας γενίκευσης, καθώς δεν δύναται να κάνει εκτιμήσεις για άγνωστες καταστάσεις. Ως λύση σε αυτά έχουν δημιουργηθεί αλγόριθμοι, οι οποίοι χρησιμοποιούν συναρτήσεις προσέγγισης και βασίζονται παραδείγματος χάριν σε νευρωνικά δίκτυα.

Παρακάτω φαίνεται ο ψευδοκώδικας του αλγορίθμου Q-learning:

Algorithm 1: Q-learning

Require: learning rate $\alpha \in (0,1] \vee small\varepsilon > 0$
  Initialize array $Q[s,\alpha]$ $\forall s \in S, \alpha \in A$
  $Q(terminal, \cdot) \leftarrow 0$
  for all episodes **do**
    Reset s
    for t=1 to end of episode **do**
      Choose a_t from s_t using ε −greedy
      Take action α_t , observe r_t+1, s_t+1
      $Q(s_t, \alpha_t) \leftarrow Q(s_t, \alpha_t) + \alpha[r_{t+1} + \gamma max_{\alpha'} Q(s_{t+1}, \alpha') - Q(s_t, \alpha_t)]$
      $s_t \leftarrow s_{t+1}$
    end for
  end for

3.2.5. Νευρωνικά δίκτυα και ενισχυτική μάθηση

Τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANNs) βασίζονται στον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου και νευρώνων και αποτελούνται από πολλούς κόμβους επεξεργασίας που συνδέονται μεταξύ τους. Χρησιμοποιούνται για την προσέγγιση μη γραμμικών συναρτήσεων και στην ενισχυτική μάθηση χρησιμοποιούνται για να προσεγγίσουν συναρτήσεις αξίας όταν ο χώρος καταστάσεων-ενεργειών είναι πολύ μεγάλος.

Οι κόμβοι (nodes) διαθέτουν επίπεδα (layers). Αρχικά υπάρχει ένα επίπεδο εισόδου (input layer), ένα επίπεδο εξόδου (output layer) και ένα ή περισσότερα κρυφά επίπεδα (hidden layers). Όταν σε ένα νευρωνικό δίκτυο υπάρχουν πάνω από ένα κρυφά επίπεδα, τότε ονομάζεται βαθύ νευρωνικό δίκτυο (deeply-layered ANN). Τα επίπεδα αυτά επικοινωνούν μεταξύ τους με συνδέσεις, στις οποίες έχουν αποδοθεί βάρη (weights). Στόχος είναι να υπολογιστούν αυτά τα βάρη μέσω πολλών επαναλήψεων μετάδοσης δεδομένων εκπαίδευσης μέσω του δικτύου τροφοδοσίας προς τα εμπρός και προς τα πίσω.

Οι νευρώνες ενός νευρωνικού δικτύου υπολογίζουν τον τρέχων μέσο της εισόδου τους με βάρη και του εφαρμόζουν μία μη γραμμική συνάρτηση ενεργοποίησης, συνήθως σιγμοειδή, για να παράξουν την έξοδο τους, δηλαδή θεωρούνται ημι-γραμμικοί. Η ενεργοποίηση των εξόδων είναι μία μη γραμμική συνάρτηση των μορφών ενεργοποίησης των εισόδων, η οποία παραμετροποιείται ως προς τα βάρη του δικτύου. Με αυτόν τον τρόπο ένα βαθύ ANN προσεγγίζει μη γραμμικές συναρτήσεις.

3.2.6. Αλγόριθμος Deep Q-Network (DQN)

Ο αλγόριθμος DQN παρουσιάστηκε για πρώτη φορά το 2015 από την DeepMind. Είναι ένας αλγόριθμος χωρίς μοντέλο και εκτός πολιτικής, ο οποίος συνδυάζει την ενισχυτική μάθηση και συγκεκριμένα τον αλγόριθμο Q-learning με βαθιά νευρωνικά δίκτυα, για να προσεγγίσει τη βέλτιστη συνάρτηση $Q^*(s,\alpha)$, που φαίνεται παρακάτω(*Mnih et al., 2015*).

$$Q^*(s,a) = max_\pi E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots | s^t = s, a_t = a, \pi] \quad (3.12)$$

Σε αντίθεση με τον αλγόριθμο Q-learning, ο οποίος χρησιμοποιεί τον πίνακα Q[s,α], ο DQN αλγόριθμός χρησιμοποιεί ένα νευρωνικό δίκτυο. Με βάση τα βάρη θ_i αυτού παραμετροποιεί την βέλτιστη συνάρτηση Q*(s,α), ώστε να την προσεγγίσει. Το νευρωνικό δίκτυο εκπαιδεύεται προσαρμόζοντας την παράμετρο θ_i σε όλα τα μονοπάτια i με βάση τις ενημερώσεις του Q-learning, έτσι ώστε το μέσο τετραγωνικό σφάλμα στην εξίσωση Bellman να μειώνεται. Επιπλέον, αρχιτεκτονικά ο αλγόριθμος έχει ξεχωριστή έξοδο για κάθε πιθανή ενέργεια α και για είσοδο το διάνυσμα κατάστασης s. Οι ενημερώσεις έχουν για στόχο y= r+γmax$_{\alpha'}$Q(s', α'; θ_i^-), όπου θ_i^- κάποια προηγούμενα βάρη, δηλαδή:

$$L_i(\theta_i) = E_{s,a,r,s'}[(r + \gamma max_{a'} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i))^2] \quad (3.13)$$

Οι αστάθειες, οι οποίες είναι συχνό φαινόμενο σε νευρωνικά δίκτυα που εκπαιδεύονται με ενισχυτική μάθηση, επιλύονται με μία νέα παραλλαγή του Q-learning αλγορίθμου χρησιμοποιώντας τις εξής τρείς βασικές τεχνικές:

- Επανάληψη εμπειρίας (Experience replay)
- Δίκτυο στόχων (Target network) και
- Πολιτική ε-greedy

Η επανάληψη εμπειρίας είναι μία τεχνική για την σταθεροποίηση της εκπαίδευσης της συνάρτησης Q(s,α), επιτρέποντας στον πράκτορα να επανεξετάσει και να μάθει από προηγούμενες ενέργειες. Η εμπειρία του πράκτορα e_t= (s_t, α_τ, r_t, s_{t+1}), δηλαδή η κατάσταση, δράση, ανταμοιβή κα επόμενη κατάσταση σε κάθε χρονικό βήμα αποθηκεύονται σε μια μνήμη επανάληψης (replay memory),η οποία είναι μία δομή της μορφής D_t=e_1,…,e_t. Κατά τη διάρκεια της εκπαίδευσης, ο πράκτορας λαμβάνει τυχαία και με ομοιόμορφη κατανομή δείγματα μιας παρτίδας εμπειριών από τη μνήμη επανάληψης για να ενημερώσει τη συνάρτηση Q(s,α) με βάση τον κανόνα ενημέρωσης του Q-learning. Με αυτό τον τρόπο αφαιρείται η συσχέτιση στην ακολουθία παρατήρησης, εξομαλύνονται οι αλλαγές στη διανομή δεδομένων και αποφεύγονται τοπικά ελάχιστα στη συνάρτηση $L_i(\theta_i)$. Πρακτικά, αποθηκεύονται στη μνήμη οι N τελευταίες εμπειρίες Επιτρέπεται, επίσης, στον πράκτορα να μάθει από διαφορετικές εμπειρίες για να σταθεροποιήσει τη διαδικασία μάθησης.

Το δίκτυο στόχων είναι ένα αντίγραφο του νευρωνικού δικτύου Q ή δικτύου ενεργειών (action network), το οποίο χρησιμοποιείται για την προσέγγιση της συνάρτησης Q(s,α). Δηλαδή, για να μείνουν σταθεροί οι στόχοι κατά τη διάρκεια της εκπαίδευσης χρησιμοποιούνται δύο νευρωνικά δίκτυα. Το δίκτυο Q ενημερώνεται σε κάθε βήμα και παράγει τις τιμές q. Το δίκτυο στόχων παράγει

τα ξεχωριστά διανύσματα βάρους θ_i^- για να δημιουργήσει ένα χρονικό κενό μεταξύ της συνάρτησης ενέργειας- αξίας q(s, α) και της συνάρτησης τιμής ενέργειας του δικτύου Q. Στην αρχή και τα δύο δίκτυα ξεκινάνε με τις ίδιες τυχαίες τιμές. Με το πέρας C βημάτων, όπου C είναι η χρονική περίοδος που επιλέγεται ως υπερπαράμετρος, οι παράμετροι του δικτύου στόχου, θ_i^-, ενημερώνονται μόνο με τις παραμέτρους του δικτύου Q, θ_i και διατηρούνται σταθερές μεταξύ των μεμονωμένων ενημερώσεων. Μαθηματικά αυτό εκφράζεται ως εξής:

$$\theta_i \leftarrow \theta_i + \alpha(r + \gamma max_{a'} Q^T(s', a'; \theta^-) - Q(s, a; \theta)\nabla_\theta Q(s, a; \theta) \quad (3.14)$$

Η επιλογή ενός ξεχωριστού δικτύου στόχων καθιστά απίθανη την απόκλιση, καθώς προσθέτει μια χρονική καθυστέρηση μεταξύ της ενημέρωσης της τιμής Q και της ενημέρωσης των τιμών στόχου Q^T. Επιπλέον το δίκτυο στόχων σταθεροποιεί τη διαδικασία εκπαίδευσης του DQN επιτρέποντας στο δίκτυο στόχων να είναι σχετικά σταθερό ενώ ενσωματώνει τις πιο πρόσφατες αλλαγές στο δίκτυο Q. Τέλος, η πολιτική ε-greedy, με ε το οποίο μειώνεται σταθερά και τελικά διατηρείται μικρό επιτρέπει την ισορροπία ανάμεσα στην εξερεύνηση και την εκμετάλλευση.Ο ψευδοκώδικας του αλγορίθμου DQN φαίνεται παρακάτω:

Algorithm 2: Deep Q-learning

Initialize replay memory D to capacity N
Initialize action-value function Q with random weights θ
Initialize target action-value function Q^T with weights $\theta^- = \theta$
for episode=1 to M **do**
  With probability ε select a random action α_t otherwise select $\alpha_t = \text{argmax}_\alpha Q(s_t, \alpha; \theta)$
  Execute action α_t in emulator and observe reward r_t and next observation o_{t+1}
  Set $s_{t+1} = s_t, \alpha_t, o_{t+1}$
  Store transition $(s_t, \alpha_t, r_t, s_{t+1})$ in D
  Sample random minibatch of transitions $(s_j, \alpha_j, r_j, s_{j+1})$ from D
  if episode terminates at step J +1 **then**
    $y_j \leftarrow j$
  else:
    $y_j \leftarrow r_j + \gamma \text{max}_{\alpha'} Q(s_{t+1}, \alpha'; \theta^-)$
  end if
  Perform a gradient descent step on $(y_j - Q(s_j, \alpha_j; \theta))^2$ with respect to the network parameters θ
  Every C steps reset $Q^T = Q$
  end for
end for

3.2.7. Μέθοδος κλίσης πολιτικής (Policy gradient method)

Οι αλγόριθμοι, που έχουν παρουσιασθεί ως τώρα, μεγιστοποιούν την ανταμοιβή προσεγγίζοντας τη βέλτιστη πολιτική μέσω κάποιας συνάρτησης αξίας (value-based algorithms). Υπάρχουν όμως και αλγόριθμοι ενισχυτικής μάθησης, οι οποίοι προσεγγίζουν απευθείας τη βέλτιστη πολιτική δημιουργώντας και αποθηκεύοντας στη μνήμη τους κατά τη διάρκεια της μάθησης μία αναπαράσταση πολιτικής, δηλαδή μία αντιστοίχιση $\pi: s \rightarrow a$. Στόχος είναι η εύρεση των παραμέτρων θ της βέλτιστης πολιτικής π*(θ).

Μια τυπική προσέγγιση για την επίλυση τέτοιων προβλημάτων μεγιστοποίησης είναι η χρήση κλίσης ανόδου ή καθόδου (gradient ascent/descent). Η πολιτική αξιολογείται άμεσα από τον υπολογισμό του αναμενόμενου αθροίσματος των ανταμοιβών υπό αυτή την πολιτική, που συμβολίζεται ως J. Μαθηματικά το J μπορεί να εκφραστεί ως:

$$J = E_{\tau \sim P_{\theta}(\tau)}[\sum_{t} r(s_t, a_t)] \quad (3.15)$$

Όπου το τ αντιπροσωπεύει μία τροχιά καταστάσεων- ενεργειών και το $P_{\theta}(\tau)$ αντιστοιχεί στην πιθανότητα να υπάρχει η τροχιά τ ακολουθώντας την πολιτική π(θ).

Η κλίση της απόδοσης μπορεί να υπολογιστεί ως εξής (*Sutton & Barto, 2018*):

$$\nabla_{\theta} J(\theta) = \frac{1}{N} \sum_{i=1}^{N} (\sum_{t=1}^{T} \nabla_{\theta} log \pi_{\theta}(\alpha_{i,t}|s_{i,t}) (\sum_{t=1}^{T} r(s_{i,t}|a_{i,t})) \quad (3.16)$$

Όπου το N αντιστοιχεί στον αριθμό των τροχιών. Έτσι, υπολογίζοντας τον μέσο όρο σε διαφορετικές τροχιές, μπορεί να ληφθεί χονδρικά μια προσέγγιση της προσδοκώμενης ανταμοιβής. Ως εκ τούτου, είναι προφανές ότι μπορεί να εφαρμοστεί η κλίση ανόδου για την ενημέρωση της πολιτικής για την αύξηση του J.

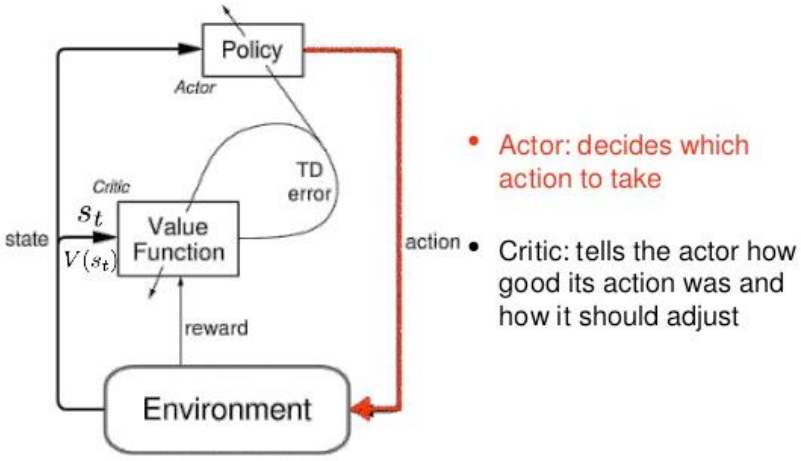
Τα πλεονεκτήματα των μεθόδων που βασίζονται στην πολιτική είναι κυρίως δύο. Πρώτον, μπορεί ο αλγόριθμος να εκπαιδευτεί σε στοχαστικές πολιτικές και δεύτερον, αυτές οι μέθοδοι είναι κατάλληλες για ενέργειες, οι οποίες λαμβάνουν συνεχείς τιμές, λόγω του ότι η ενέργεια πραγματοποιείται δειγματοληπτικά στο π(α|s), με αποτέλεσμα συνεχείς τιμές.

Ωστόσο, ένα σημαντικό μειονέκτημα των αλγορίθμων που βασίζονται στην πολιτική είναι ότι η αξιολόγηση της αναμενόμενης απόδοσης πάσχει από μεγάλη διασπορά και, επομένως, καθιστά τις ενημερώσεις της πολιτικής ασταθείς και αργές, μερικές φορές αποτυγχάνοντας να βρεθεί η βέλτιστη λύση. Για να ξεπεραστεί αυτό το πρόβλημα προτείνεται η εξέταση αλγορίθμων που βασίζονται τόσο στην πολιτική, όσο και στις συναρτήσεις αξίας.

Αυτός ο συνδυασμός οδηγεί στην τρίτη κατηγορία ενισχυτικής μάθησης, στις μεθόδους Δράστη-Κριτή (Actor-Critic).

3.2.8. Αλγόριθμοι Δράστη-Κριτή (Actor-Critic)

Οι αλγόριθμοι Δράστη-Κριτή (Actor-Critic) είναι μία παραλλαγή των αλγορίθμων κλίσης πολιτικής. Διαθέτουν δύο κεντρικά νευρωνικά δίκτυα, έναν δράστη (actor), ο οποίος αποφασίζει την πολιτική π(α|s,θ), δηλαδή τις επόμενες ενέργειες, και έναν κριτή (critic), ο οποίος αξιολογεί την πολιτική με βάση την αξία της κατάστασης. Η διαδικασία αυτή περιγράφεται και στην Εικόνα 10.



Εικόνα 10: Διάγραμμα ροής αλγορίθμου actor-critic (Πηγή: datascience.stackexchange.com)

Για να μειωθεί η διασπορά της αναμενόμενης απόδοσης χρησιμοποιείται μία συνάρτηση βάσης b(s) (baseline function), η οποία δεν πρέπει να εξαρτάται από τις ενέργειες. Στη συγκεκριμένη περίπτωση μία καλή επιλογή συνάρτησης βάσης είναι η συνάρτηση αξίας-κατάστασης. Άρα, η προηγούμενη εξίσωση γίνεται:

$$\nabla_\theta J(\theta) = \sum_S d^\pi(s) \sum_a \frac{\partial \pi(s,a)}{\partial \theta} \left(Q^\pi(s,a) - V^\pi(s)\right) \quad (3.17)$$

Η διαφορά $Q^\pi(s,\alpha)$- $V^\pi(s)$ ονομάζεται συνάρτηση πλεονεκτήματος ($A^\pi(s,\alpha)$) και είναι άγνωστη, άρα πρέπει να εκτιμηθεί από τον κριτή. Έτσι, ο κριτής μπορεί να διαθέτει δύο μοντέλα, ένα για την συνάρτηση αξίας κατάστασης-ενέργειας και ένα για τη συνάρτηση αξίας κατάστασης. Ο κριτής μαθαίνει τις παραμέτρους συνήθως μέσω του σφάλματος χρονικών διαφορών, καθώς η συνάρτηση σφαλμάτων χρονικών διαφορών (TD error) είναι ένας αμερόληπτος εκτιμητής της συνάρτησης πλεονεκτήματος:

$$E_{\pi_\theta}[\delta^{\pi_\theta}|s,a] = E_{\pi_\theta}[r + \gamma V^{\pi_\theta}(s')|s,a] - V^{\pi_\theta}(s)$$

$$\rightarrow E_{\pi_\theta}[\delta^{\pi_\theta}|s,a] = Q^{\pi_\theta}(s,a) - V^{\pi_\theta}(s) \qquad (3.18)$$

Και η κλίση της πολιτικής υπολογίζεται ως εξής:

$$\nabla_\theta E_{\pi_\theta}[R] = E_{\pi_\theta}[\nabla_\theta \log \pi_\theta(s,a)\, \delta^{\pi_\theta}] \qquad (3.19)$$

Πρακτικά όμως επειδή στην πραγματικότητα υπολογίζεται ένα προσεγγιστικό σφάλμα χρονικών διαφορών, ο κριτής δεν χρειάζεται να έχει δύο μοντέλα αλλά μόνο ένα, αυτό για την εκτίμηση της συνάρτησης αξίας-καταστάσεων.

3.2.9. Αλγόριθμος Proximal Policy Optimization (PPO)

Ο αλγόριθμος Proximal Policy Optimization ή PPO προτάθηκε για πρώτη φορά το 2017 από τους *Schulman et al.* και είναι βασισμένος στο μοντέλο δράστη-κριτή. Η κύρια ιδέα του αλγορίθμου PPO είναι ότι μετά από μία ενημέρωση η νέα πολιτική δεν πρέπει να απέχει πολύ από την προηγούμενη. Για να το πετύχει αυτό ο αλγόριθμος προσπαθεί να εκπαιδευτεί με το κατάλληλο βήμα, το οποίο να μην είναι πολύ μικρό, με αποτέλεσμα να έχουμε αργή σύγκλιση, ούτε πολύ μεγάλο, ώστε να μην μπορεί να συγκλίνει. Αυτό ονομάζεται περιοχή εμπιστοσύνης.

Ο αλγόριθμος PPO διατηρεί δύο διαφορετικά νευρωνικά δίκτυα πολιτικής. Πρώτον, το δίκτυο της νέας πολιτικής $\pi_\theta(\alpha_t|s_t)$ που πρέπει να ανανεωθεί και δεύτερον της προηγούμενης πολιτικής $\pi_{\theta k}(\alpha_t|s_t)$, η οποία είχε δημιουργηθεί από παλαιότερη εμπειρία. Για να μπορέσει η νέα πολιτική να μην απέχει πολύ από την προηγούμενη, καθορίζει τον λόγο τους $r_t(\theta)$:

$$r_t(\theta) = \frac{\pi_\theta(\alpha_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} \quad (3.20)$$

και στη συνέχεια υπολογίζει την συνάρτηση απώλειας σύμφωνα με την παρακάτω εξίσωση:

$$L_{\theta_k}^{CLIP}(\theta) = E_{\tau\sim\pi_k}[\sum_{t=0}^{T}\left[\min\left(r_t(\theta)A_t^{\pi_k}, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)A_t^{\pi_k}\right]\right] \quad (3.21)$$

όπου Ε το τετραγωνικό σφάλμα, A_t το πλεονέκτημα και ε μία σταθερή παράμετρος, συνήθως 0,1 ή 0,2. Άρα οι όροι 1-ε και 1+ε δεν εξαρτώνται από το θ, αποδίδοντας έτσι μηδενική διαβάθμιση. Κατά συνέπεια, δείγματα εκτός της αξιόπιστης περιοχής απορρίπτονται αποτελεσματικά, αποθαρρύνοντας τις υπερβολικά μεγάλες ενημερώσεις. Ως αποτέλεσμα, δεν περιορίζεται ρητά η ίδια η ενημέρωση πολιτικής, αλλά απλώς αγνοούνται τα πλεονεκτήματα που προκύπτουν από υπερβολικά αποκλίνουσες πολιτικές.

Παρακάτω φαίνεται ο ψευδοκώδικας του αλγορίθμου PPO:

Algorithm 3: Proximal Policy Optimization

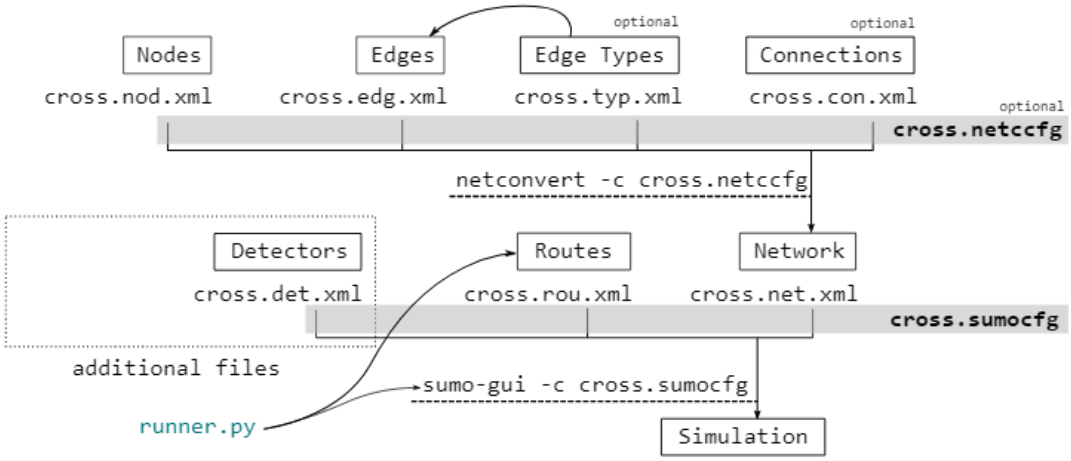
for iteration=1,2,… **do**
  for actor=1,2,…,N **do**
    Run policy $\pi_{\theta_{old}}$ in environment for T timesteps
    Compute advantage estimates $\widehat{A_1}$,…,$\widehat{A_T}$
  end for
  Optimize surrogate L wrt θ, with K epochs and minibatch size M≤NT
  $\theta_{old} \leftarrow \theta$
end for

3.3. Προσομοίωση Αστικής Κινητικότητας

Η Προσομοίωση Αστικής Κινητικότητας (Simulation of Urban MObility-SUMO) είναι ένα πακέτο προσομοίωσης ανοιχτού κώδικα. Επιτρέπει την προσομοίωση αστικών οδικών δικτύων πολυτροπικών (multimodal) μέσων μεταφοράς συμπεριλαμβανομένων και των πεζών και συνοδεύεται από ένα μεγάλο σύνολο εργαλείων για τη δημιουργία σεναρίων (*Alvarez Lopez et al., 2018)*. Παρουσιάστηκε για πρώτη φορά το 2001 από το Γερμανικό Κέντρο Αεροδιαστημικής (Deutsche Zentrum für Luft- und Raumfahr- DLR).

Στο SUMO περιλαμβάνεται πληθώρα εργαλείων υποστήριξης, τα οποία μπορούν να προστεθούν στις προσομοιώσεις, που αυτοματοποιούν βασικές εργασίες για τη δημιουργία, την εκτέλεση και την αξιολόγηση προσομοιώσεων κυκλοφορίας, όπως εισαγωγή δικτύου, υπολογισμοί διαδρομής, οπτικοποίηση και υπολογισμός εκπομπών. Μία προσομοίωση στο SUMO αποτελείται από τουλάχιστον τρία εισαγώγιμα αρχεία (inputs) μορφής xml, το αρχείο δικτύου (network file), το αρχείο ζήτησης (route file) και το αρχείο διαμόρφωσης (configuration file). Οποιαδήποτε άλλη πληροφορία για την προσομοίωση μπορεί να συμπληρωθεί με την μορφή πρόσθετου αρχείου (additional file).

Το αρχείο δικτύου τις οδούς του δικτύου σε μορφή συνδέσεων (edges), τα σημεία αρχής, τέλους και τομής των συνδέσεων (nodes) και μπορεί να περιλαμβάνει στοιχεία σηματοδότησης και προτεραιοτήτων. Το αρχείο ζήτησης (route file) περιλαμβάνει όλα τα οχήματα ή ροές οχημάτων που θα εισαχθούν στην προσομοίωση και τις διαδρομές τους. Τέλος το αρχείο διαμόρφωσης είναι το αρχείο το οποίο συνδέει το αρχείο δικτύου και ζήτησης και φορτώνεται στο λογισμικό ώστε να «τρέξει» μία προσομοίωση. Στο αρχείο διαμόρφωσης ορίζονται και τα εξαγώγιμα αρχεία (outputs), στα οποία καταγράφονται τα δεδομένα που εξάγονται από την προσομοίωση. Ο τρόπος με τον οποίο συνδέονται όλα τα αρχεία φαίνεται στην Εικόνα 11.



Εικόνα 11:Συνοπτική παρουσίαση αρχείων προσομοίωσης
(Πηγή: intelaligent.github.io)

3.3.1.Μοντέλα προσομοίωσης

Τα μικροσκοπικά χαρακτηριστικά της οδήγησης των οχημάτων καθορίζονται από την αλληλεπίδραση τριών μοντέλων, τα οποία παρατίθενται παρακάτω:

- Μοντέλο ακολουθούντος οχήματος (Car-following model)
- Μοντέλο διασταύρωσης (Intersection model) και
- Μοντέλο αλλαγής λωρίδας (Lane-change model)

Το μοντέλο ακολουθούντος οχήματος καθορίζει την ταχύτητα ενός οχήματος σε σχέση με το όχημα μπροστά του. Δηλαδή, κάθε όχημα προσπαθεί πάντα να κρατήσει μια απόσταση ασφαλείας με το προπορευόμενο όχημα και προσαρμόζεται πάντα στη συμπεριφορά επιβράδυνσης του προπορευόμενου οχήματος. Ως ασφαλής ταχύτητα ορίζεται ως η ταχύτητα που επιλέγουν οι οδηγοί των οχημάτων, μέσα στο όριο ταχύτητας που ορίζει ο νόμος και σκοπό να εξασφαλίσουν ασφαλή οδήγηση ανάλογα με τις περιστάσεις, τις συνθήκες του δρόμου και τις συνθήκες κυκλοφορίας.

Το μοντέλο διασταύρωσης καθορίζει τη συμπεριφορά των οχημάτων σε διαφορετικούς τύπους διασταυρώσεων όσον αφορά τους κανόνες διέλευσης, την αποδοχή της διασταύρωσης ή συγχώνευσης της κυκλοφορίας δύο ρευμάτων και την αποφυγή κορεσμού των κόμβων.

Τέλος, το μοντέλο αλλαγής λωρίδας καθορίζει την επιλογή λωρίδας σε δρόμους με πολλές λωρίδες. Τα οχήματα αλλάζουν λωρίδα κυκλοφορίας για πολλούς λόγους, υποχρεωτικούς αλλά και προαιρετικών ελιγμών. Το μοντέλο αλλαγής λωρίδας στο SUMO αναγνωρίζει επί του παρόντος τέσσερις λόγους αλλαγής λωρίδας:

- Στρατηγική αλλαγή λωρίδας, δηλαδή το όχημα πρέπει να χρησιμοποιήσει άλλη λωρίδα για να συνεχίσει την πορεία του
- Συνεργατική αλλαγή λωρίδας, δηλαδή το όχημα θέλει να απελευθερώσει μία λωρίδα για ένα άλλο όχημα
- Ενίσχυση ταχύτητας, δηλαδή το όχημα θέλει να επιταχύνει αλλάζοντας σε μία ταχύτερη λωρίδα και
- Διατήρηση δεξιάς λωρίδας καθώς οι αριστερές λωρίδες παραχωρούνται στα ταχύτερα οχήματα.

Το μοντέλο αλλαγής λωρίδας είναι επίσης υπεύθυνο για την προσαρμογή της ταχύτητας των οχημάτων ώστε να επιτρέπεται η πραγματοποίηση ελιγμών αλλαγής λωρίδας. Αυτό έχει τεράστια σημασία ιδίως σε περιπτώσεις πυκνής κυκλοφορίας, επειδή τα οχήματα οφείλουν να διατηρούν ασφαλείς αποστάσεις από όλα τα οχήματα της λωρίδας που προσπαθούν να προσεγγίσουν για να αποφύγουν τις συγκρούσεις. Η επίτευξη ασφαλών αποστάσεων απαιτεί συχνά αλλαγές ταχύτητας από το όχημα που εκτελεί ελιγμό καθώς και από τα οχήματα στη λωρίδα στόχου του. Μεταξύ των αποφάσεων που συνήθως λαμβάνονται από το μοντέλο αλλαγής λωρίδας είναι εάν ένα όχημα που μπλοκάρει τη λωρίδα στόχο πρέπει να προσπεραστεί ή εάν είναι προτιμότερο να επιβραδυνθεί η κυκλοφορία και να ληφθεί αυτό το όχημα ως προπορευόμενο.

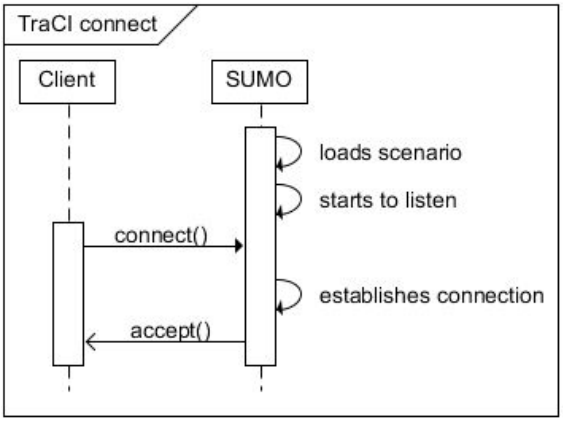
3.3.2 Διεπαφές Προγραμματισμού Επιφανειών- TraCI

Το SUMO μπορεί να προσφέρει περισσότερες ικανότητες στο χρήστη, καθώς παρέχει διάφορες Διεπαφές Προγραμματισμού Εφαρμογών (Application Programming Interface-API) για τον

απομακρυσμένο έλεγχο της προσομοίωσης. Το πιο γνωστό τέτοιο εργαλείο είναι η βιβλιοθήκη TraCI (Traffic Control Interface), το οποίο χρησιμοποίει μία αρχιτεκτονική τύπου TCP πελάτη-διακομιστή. Ως διακομιστής λειτουργεί το πρόγραμμα προσομοίωσης SUMO, ενώ ως πελάτης λειτουργεί κάποιο πρόγραμμα γραμμένο σε μία από τις παρακάτω γλώσσες προγραμματισμού που υποστηρίζονται από το TraCI:

- Python
- C++
- MATLAB και
- Java

Τα προγράμματα αυτά ονομάζονται ελεγκτές (controllers) και λαμβάνουν πληροφορίες από τον διακομιστή σχετικά με την πορεία της προσομοίωσης και στη συνέχεια στέλνουν εντολές πίσω στον διακομιστή, όπως φαίνεται και στην Εικόνα 12. Σκοπός της βιβλιοθήκης αυτής είναι η πρόσβαση στο πρόγραμμα προσομοίωσης, η διαχείρισή του και η εύκολη και άμεση ρύθμιση των παραμέτρων της προσομοίωσης κατά τη διάρκειά της, δηλαδή ο πλήρης απομακρυσμένος έλεγχός της.



Εικόνα 12: Αρχιτεκτονική TraCI τύπου πελάτη-διακομιστή (Πηγή: sumo.dlr.de)

Κεφάλαιο 4: Εφαρμογή και αποτελέσματα

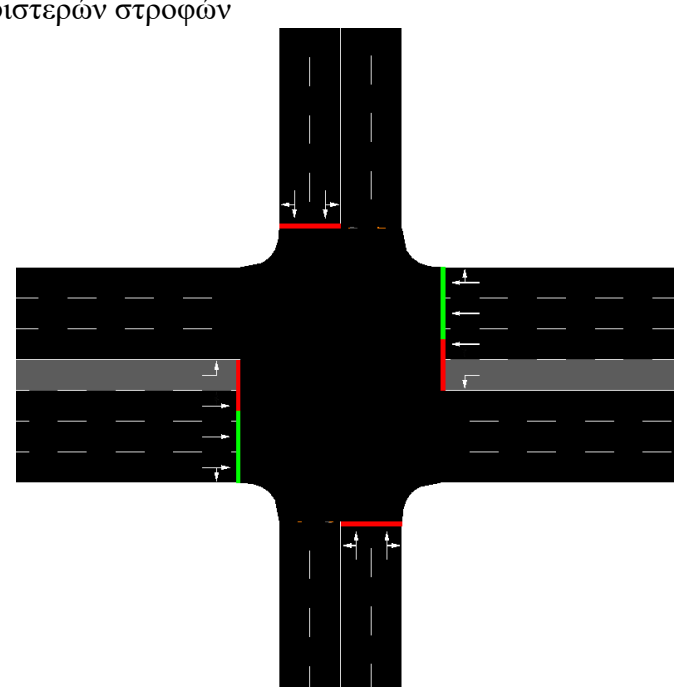
4.1. Περιβάλλον επίλυσης

4.1.1. Το δίκτυο

Πρώτο βήμα για να μπορέσει να δομηθεί το πρόβλημα είναι η δημιουργία ενός δικτύου στο οποίο θα δοκιμαστούν οι δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας. Για το σκοπό αυτό δημιουργήθηκε ένας οδικό τμήμα τριών (3) διαδοχικών κόμβων με απόσταση 250 μέτρα. Η κύρια αρτηρία αποτελείται από επτά (7) λωρίδες κυκλοφορίας, εκ των οποίων η μία (1) είναι αποκλειστικών αριστερών στροφών. Οι τρεις δευτερεύουσες οδοί είναι τεσσάρων (4) λωρίδων κυκλοφορίας, δύο (2) για κάθε κατεύθυνση. Ολόκληρη η δομή του δικτύου φαίνεται στην Εικόνα 16.

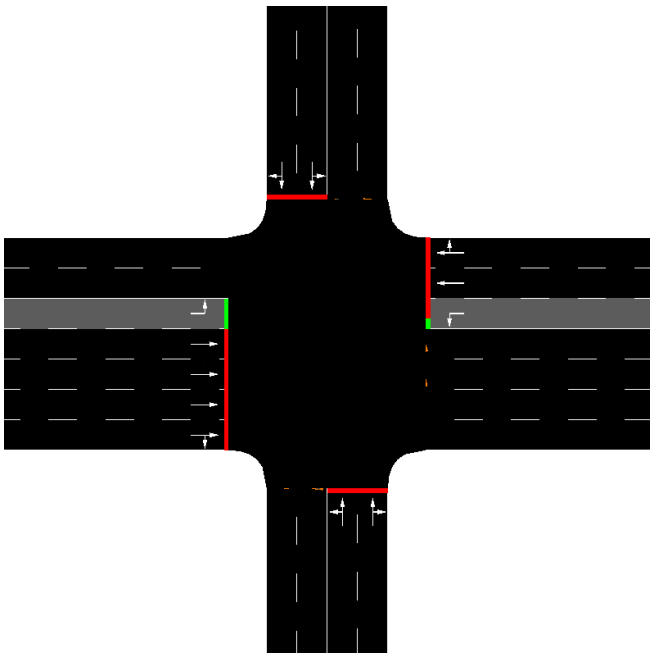
Στην κύρια αρτηρία δημιουργήθηκε μία εναλλασσόμενη λωρίδα κυκλοφορίας, έτσι οι δυνατές διαμορφώσεις του οδικού χώρου παρουσιάζονται στις Εικόνες 13,14 και 15 και είναι οι εξής τρεις:

- Διαμόρφωση 1-Τρεις λωρίδες κυκλοφορίας και στις δύο κατευθύνσεις, συν μία λωρίδα αποκλειστικών αριστερών στροφών



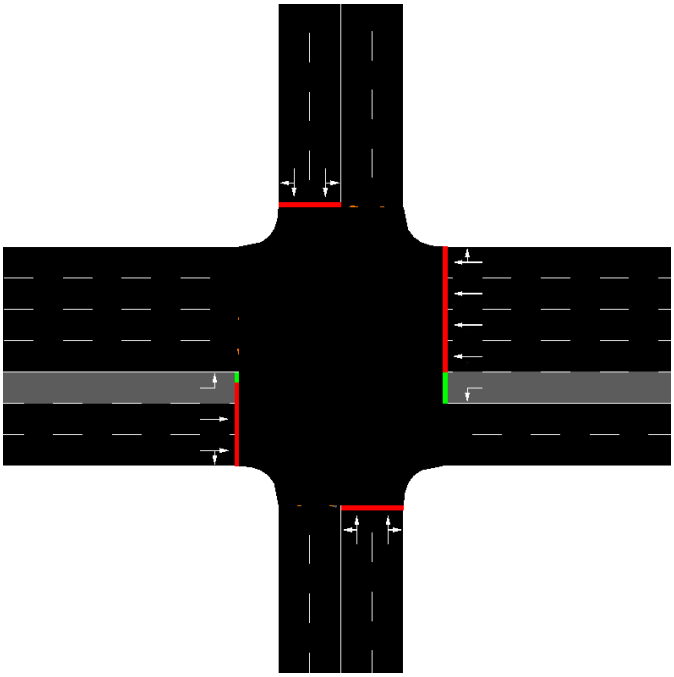
Εικόνα 13:Διαμόρφωση 1

- Διαμόρφωση 2-Τέσσερις λωρίδες κυκλοφορίας στην κατεύθυνση προς τα ανατολικά και δύο προς τα δυτικά, συν μία λωρίδα αποκλειστικών αριστερών στροφών και



Εικόνα 14: Διαμόρφωση 2

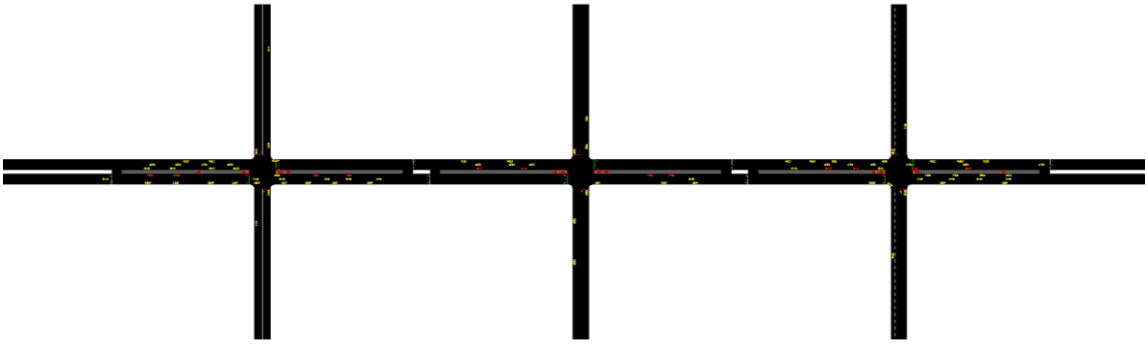
- Διαμόρφωση 3-Τέσσερις λωρίδες κυκλοφορίας στην κατεύθυνση προς τα δυτικά και δύο προς τα ανατολικά, συν μία λωρίδα αποκλειστικών αριστερών στροφών



Εικόνα 15: Διαμόρφωση 3

Το δίκτυο δημιουργήθηκε μέσω του προγράμματος netedit, το οποίο είναι ένα οπτικό πρόγραμμα επεξεργασίας οδικών δικτύων που μπορεί να χρησιμοποιηθεί για τη δημιουργία δικτύων από την

αρχή και για την τροποποίηση όλων των πτυχών υπαρχόντων δικτύων. Το πρόβλημα στην δημιουργία δικτύων με εναλλασσόμενες λωρίδες κυκλοφορίας είναι ότι προς το παρόν το netedit δεν αναγνωρίζει την ιδιότητας της εναλλαγής κατεύθυνσης στις λωρίδες που δημιουργεί. Για την επίλυση του ζητήματος αυτού δημιουργήθηκαν αλληλεπικαλυπτόμενες λωρίδες και στις δύο κατευθύνσεις, στις οποίες επιτρέπεται ή απαγορεύεται η διέλευση οχημάτων μέσω TraCI ανάλογα με τη διαμόρφωση των λωρίδων.



Εικόνα 16: Συνολική εικόνα του δικτύου σε διαμόρφωση 1

4.1.2. Η εναλλαγή των λωρίδων

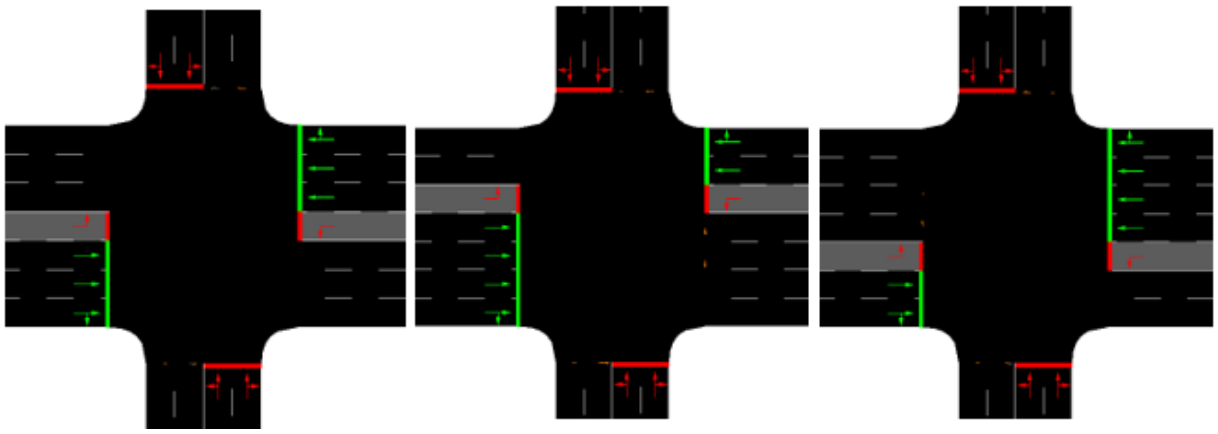
Η εναλλαγή της διαμόρφωσης των λωρίδων γίνεται μέσω ενός αλγορίθμου. Σαν παραδοχή λαμβάνεται ένα περιβάλλον συνδεσιμότητας ανάμεσα στα οχήματα και σε όλο το κοντινό τους περιβάλλον (V2X). Η κύρια οδική αρτηρία αναγνωρίζει τον αριθμό οχημάτων που την καταλαμβάνουν κάθε χρονική στιγμή. Όταν οι συνθήκες κυκλοφορίας στις δύο κατευθύνσεις κυκλοφορίας μεταβληθούν σημαντικά, λαμβάνεται απόφαση για διακοπή της κυκλοφορίας σε μία ή δύο λωρίδες κυκλοφορίας της κατεύθυνσης με το λιγότερο φόρτο και αμέσως μετά τη διάθεσή τους στην κατεύθυνση με το μεγαλύτερο φόρτο. Για να γίνει αυτό όταν ληφθεί η απόφαση εναλλαγής δίνεται ένα σήμα εκκένωσης από την οδική υποδομή προς τα οχήματα της λωρίδας, η οποία πρόκειται να αλλάξει κατεύθυνση. Μόλις όλα τα οχήματα έχουν απομακρυνθεί από τη λωρίδα δίνεται σήμα στα οχήματα της κατεύθυνσης στην οποία διατέθηκε η λωρίδα ότι μπορούν να τη χρησιμοποιήσουν.

Για αλλαγή από διαμόρφωση 2 σε διαμόρφωση 3 και το αντίθετο, ακολουθείται η ίδια διαδικασία σταδιακά δύο φορές, περνώντας απαραίτητα από τη διαμόρφωση 1.

4.1.3. Σηματοδότηση κόμβων

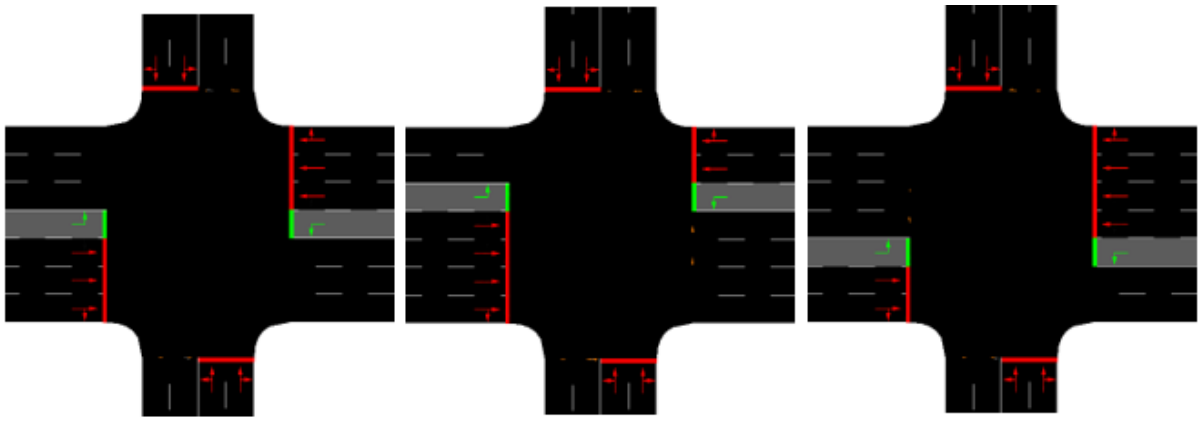
Οι κόμβοι του δικτύου είναι σηματοδοτούμενοι με αυτόματους σηματοδότες (actuated traffic lights) και για όλες τις διαμορφώσεις των λωρίδων υπάρχουν τέσσερεις φάσεις σηματοδότησης, οι οποίες φαίνονται παρακάτω και στις Εικόνες 17,18,19 και 20:

- Φάση 1- Διέλευση ευθείων κινήσεων και δεξιών στροφών από την κύρια αρτηρία



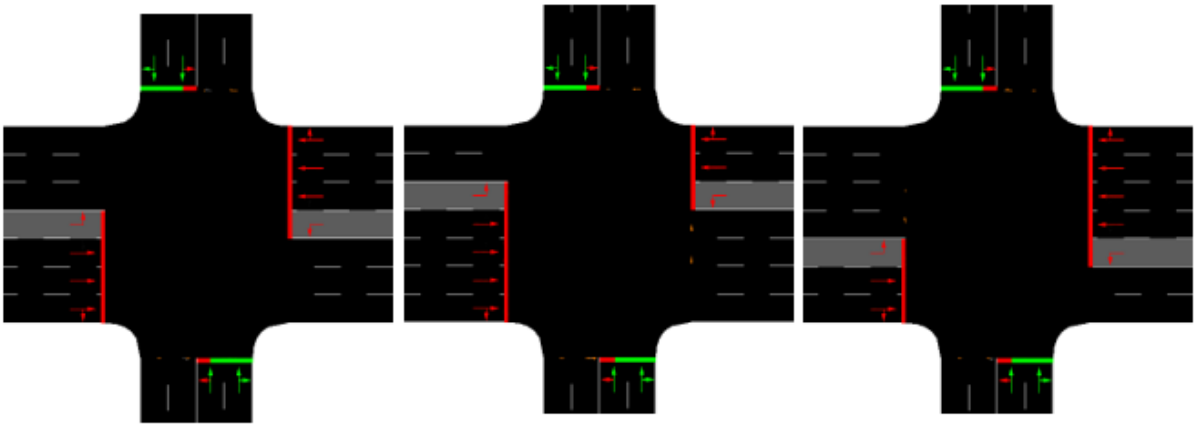
Εικόνα 17:1η Φάση σηματοδότησης για όλες τις διαμορφώσεις

- Φάση 2- Διέλευση αποκλειστικών αριστερών στροφών από την κύρια οδό



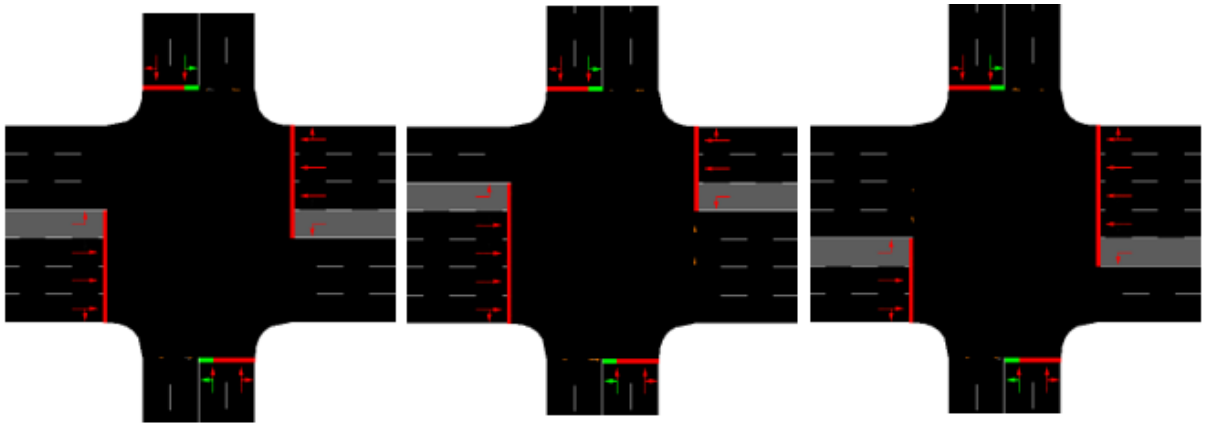
Εικόνα 18:2η Φάση σηματοδότησης για όλες τις διαμορφώσεις

- Φάση 3- Διέλευση ευθείων κινήσεων και δεξιών στροφών από τις δευτερεύουσες οδούς του δικτύου



Εικόνα 19:3η Φάση σηματοδότησης για όλες τις διαμορφώσεις

- Φάση 4- Διέλευση αποκλειστικών αριστερών στροφών από τις δευτερεύουσες οδούς του δικτύου



Εικόνα 20:4η Φάση σηματοδότησης για όλες τις διαμορφώσεις

Ο όρος αυτοματοποιημένη σηματοδότηση αναφέρεται σε φωτεινούς σηματοδότες που αλλάζουν φάση ανάλογα με την κυκλοφορία. Ο τρόπος με τον οποίο βοηθούν την κυκλοφορία είναι να παρατείνουν τις φάσεις σηματοδότησης κάθε φορά που εντοπίζεται συνεχές ρεύμα κυκλοφορίας. Ως συνεχές ρεύμα κυκλοφορίας έχει οριστεί στο πρόβλημα ένα ρεύμα του οποίου τα οχήματα έχουν χρονική απόσταση μικρότερη από 3s. Ο σηματοδότης μεταβαίνει στην επόμενη φάση αφού ανιχνεύσει χρονική απόσταση μεταξύ διαδοχικών οχημάτων μεγαλύτερη των 3s. Αυτό επιτρέπει την καλύτερη κατανομή του χρόνου πρασίνου μεταξύ των φάσεων και επίσης επηρεάζει τη διάρκεια του κύκλου ως απόκριση στις δυναμικές συνθήκες κυκλοφορίας. Επιπλέον, μία φάση σηματοδότησης δεν μπορεί να υπερβεί τα 45s, αλλά ούτε και να είναι μικρότερη από 5s.

Τα χρονικά κενά που καθορίζουν τις επεκτάσεις φάσης συλλέγονται από φωρατές κυκλοφορίας (detectors) , οι οποίοι τοποθετούνται αυτόματα σε ρυθμιζόμενη απόσταση από τη γραμμή στάσης στις προσβάσεις των σηματοδοτούμενων κόμβων. Η απόσταση αυτή εξαρτάται από την ελάχιστη χρονική διάρκεια μιας φάσης (minDur) και τη χρονική απόσταση μεταξύ των οχημάτων κατά τη διέλευσή τους από τη γραμμή στάσης (passingTime) σύμφωνα με τον τύπο:

$$\left({}^{minDur}/_{passingTime} + 0.5\right) * 7.5 \quad (4.1)$$

Ο σκοπός αυτού του ορίου είναι να επιτρέψει σε όλα τα οχήματα μεταξύ του ανιχνευτή και της γραμμής στάσης να περάσουν τη διασταύρωση εντός του χρόνου minDur. Στη συγκεκριμένη περίπτωση, ο ελάχιστος χρόνος φάσης είναι 5s και η χρονική απόσταση μεταξύ των οχημάτων κατά τη διέλευσή τους από τη γραμμή στάσης είναι 2s, άρα οι φωρατές έχουν τοποθετηθεί σε απόσταση 22,5m από τη γραμμή στάσης.

4.1.4. Σενάρια κυκλοφορικής ζήτησης εκπαίδευσης

Για την εκπαίδευση του αλγορίθμου χρησιμοποιήθηκαν εκατό διαφορετικά σενάρια ζήτησης. Τα σενάρια αυτά έχουν έναν βασικό κυκλοφοριακό φόρτο της κύριας αρτηρίας, ο οποίος κυμαίνεται από 1000 έως 2000 οχήματα την ώρα. Η διαφορά στο φόρτο των δύο κατευθύνσεων δημιουργείται

από τη χρήση ενός ποσοστού, το οποίο κυμαίνεται από 50%-70%. Αυτό το ποσοστό του βασικού κυκλοφοριακού φόρτου αποδίδεται στην μία κατεύθυνση αρχικά και στα μέσα της προσομοίωσης αποδίδεται στην αντίθετη κατεύθυνση. Κάθε σενάριο ζήτησης διαρκεί 3600s και άρα η εναλλαγή των φόρτων γίνεται στα 1800s. Στόχος των διαφορετικών σεναρίων είναι η εξοικείωση του αλγορίθμου σε διαφορετικές συνθήκες ζήτησης και τελικώς η ορθότερη εκμάθηση του αλγορίθμου.

4.2. Επίλυση του προβλήματος

4.2.1. Προκαταρτική θεώρηση αλγορίθμου ενισχυτικής μάθησης

Στο κεφάλαιο 3 παρουσιάστηκαν δύο διαφορετικοί τύποι αλγορίθμων αυτοί που προσεγγίζουν την βέλτιστη πολιτική μέσα από συναρτήσεις αξίας και κυρίως ο αλγόριθμος DQN και αυτοί που προσεγγίζουν απευθείας τη βέλτιστη πολιτική και κυρίως ο αλγόριθμος PPO.

Το σύνολο των ενεργειών του προβλήματος είναι διακριτό (3 πιθανές ενέργειες για κάθε κατάσταση). Αυτό καθιστά και τους δύο αλγόριθμους κατάλληλους για την εκπαίδευση ενός μοντέλου. Αρχικά, επιλέχθηκε η εκπαίδευση του μοντέλου με τη χρήση του αλγορίθμου DQN. Ωστόσο, κατά την εκπαίδευση παρατηρήθηκε μεγάλη αστάθεια στην αμοιβή που λάμβανε ο πράκτορας και δεν παρουσιάστηκε σύγκλιση προς τη βέλτιστη ανταμοιβή με το πέρας των επεισοδίων.

Αυτό μπορεί να οφείλεται στην πιθανή στοχαστικότητα της βέλτιστης πολιτικής, όπου μια συγκεκριμένη κατάσταση δεν αντιστοιχεί σε μία μοναδική ενέργεια. Ο αλγόριθμος PPO αντιμετωπίζει αποτελεσματικότερα τέτοια προβλήματα, γι' αυτό επιλέγεται για την τελική εκπαίδευση του μοντέλου.

Παρόλα αυτά, ακόμα και ο αλγόριθμος PPO αντιμετωπίζει προβλήματα σύγκλισης κατά την εκπαίδευση. Επομένως, η εκπαίδευση του μοντέλου για την επιλογή της διαμόρφωσης των λωρίδων κυκλοφορίας της κύριας αρτηρίας διαρθρώνεται σε δύο στάδια: ένα στάδιο προ-εκπαίδευσης και ένα μετέπειτα στάδιο εκπαίδευσης.

4.2.2. Ανάπτυξη και Εκπαίδευση μοντέλου

Η προ-εκπαίδευση στοχεύει στην απόκτηση μεταβιβάσιμης γνώσης από τα δεδομένα εκπαίδευσης για την ευκολότερη και ορθότερη εκπαίδευση του τελικού μοντέλου. Τόσο στην εκπαίδευση όσο και στην προ-εκπαίδευση χρησιμοποιήθηκε ο αλγόριθμος PPO.

Ο αλγόριθμος της προ-εκπαίδευσης και της εκπαίδευσης διαφέρουν στην αμοιβή r_t που επιστρέφει το περιβάλλον. Θεωρήθηκαν ως κλασσικά προβλήματα ενισχυτικής μάθησης, στα οποία ένας πράκτορας αλληλεπιδρά με το περιβάλλον του. Στις συγκεκριμένες περιπτώσεις το περιβάλλον είναι το οδικό δίκτυο της παραγράφου 4.1.1 και κάθε χρονικό διάστημα t=600s ο πράκτορας καλείται να επιλέξει μία ενέργεια a_t, δηλαδή μία από τις τρεις διαμορφώσεις, οι οποίες έχουν οριστεί. Ορίστηκε το συγκεκριμένο χρονικό διάστημα ανάμεσα στις αποφάσεις του πράκτορα ως ιδανικό για να αποτυπώσει την μεταβολή στην κατάσταση του δικτύου. Μεγαλύτερα χρονικά διαστήματα από αυτό θα οδηγούσαν στο να αγνοηθούν κάποιες καταστάσεις δικτύου και θα και η εκπαίδευση θα ολοκληρωνόταν έχοντας περάσει από λιγότερες καταστάσεις. Αντίθετα, παρόλο που σε ένα μικρότερο χρονικό διάστημα ο πράκτορας θα περνούσε από περισσότερες καταστάσεις, οι ενέργειες που θα λάμβανε θα ήταν περισσότερο τυχαίες.

Μετά από κάθε ενέργεια που δίνεται στο περιβάλλον, αυτό αλλάζει την εσωτερική του κατάσταση s_t και επιστρέφει ένα σύνολο παρατηρήσεων και την ανταμοιβή. Ως παρατηρήσεις o_t έχουν οριστεί τα ακόλουθα μεγέθη:

- Την προηγούμενη ενέργεια α_{t-1}
- Την ενέργεια α_t
- Τον χρόνο εκκένωσης της λωρίδας που αλλάζει κατεύθυνση και
- Πόσα βήματα απομένουν μέχρι τη λήξη του επεισοδίου

Η ανταμοιβή που έχει οριστεί για τη φάση της προ-εκπαίδευσης βασίζεται σε έναν πίνακα ιδανικών ενεργειών, ο οποίος έχει δημιουργηθεί για κάθε σενάριο ζήτησης εκπαίδευσης. Η λογική αυτού του πίνακα είναι να θεωρεί ως ιδανική τη διαμόρφωση που δίνει μία επιπλέον λωρίδα στην κατεύθυνση με το μεγαλύτερο φόρτο. Ως εκ τούτου, η ανταμοιβή r_t ορίστηκε ως εξής:

$$r_t = \begin{cases} 0, \alpha\nu\ \alpha_t = a_{t,\iota\delta\alpha\nu\iota\kappa\acute{o}} \\ -1, \alpha\nu\ \alpha_t = a_{t,\iota\delta\alpha\nu\iota\kappa\acute{o}} \pm 1 \\ -2,\ \ \sigma\varepsilon\ \kappa\acute{a}\theta\varepsilon\ \acute{a}\lambda\lambda\eta\ \pi\varepsilon\rho\acute{\iota}\pi\tau\omega\sigma\eta \end{cases} \quad (4.2)$$

Στόχος, δηλαδή, του μοντέλου για αρχή είναι να μάθει να αναγνωρίζει την κατεύθυνση της μεγαλύτερης κυκλοφοριακής ροής, ώστε να αυξήσει την ανταμοιβή του.

Έπειτα, στην περίπτωση της εκπαίδευσης, η ανταμοιβή r_t που επιστρέφει το περιβάλλον μετά από κάθε ενέργεια είναι ο μέγιστος μέσος χρόνος διαδρομής που παρατηρείται στις δύο κατευθύνσεις, δηλαδή:

$$r_t = -\max(meanTravelTime_{east}, meanTravelTime_{west}) \quad (4.3)$$

Έχει επιλεχθεί αυτό το μέγεθος γιατί είναι καλός δείχτης της εξυπηρέτησης των οχημάτων και στις δύο κατευθύνσεις. Στόχος είναι μέσα από πολλά τα επεισόδια παρατηρήσεων, ενεργειών, ανταμοιβών (s_t, a_t, r_t, s_{t+1}) ο αλγόριθμος να εκπαιδευτεί στο να αναγνωρίζει καταστάσεις και να λαμβάνει ενέργειες, έτσι ώστε να μεγιστοποιήσει τη συνολική ανταμοιβή που λαμβάνει στο τέλος, στην προκειμένη περίπτωση να την οδηγήσει πιο κοντά στο 0.

Ο αλγόριθμος PPO που κατασκευάστηκε αποτελείται από 2 νευρωνικά δίκτυα, το δίκτυο του κριτή και το δίκτυο του δράστη. Το καθένα από τα οποία έχει 2 κρυμμένα επίπεδα με 64 νευρώνες το καθένα (Multilayer percepton policy- Mlppolicy), τα οποία έχουν είσοδο το διάνυσμα παρατηρήσεων και 3 εξόδους, μία για κάθε ενέργεια. Οι τιμές των υπερπαραμέτρων που χρησιμοποιήθηκαν φαίνονται στον Πίνακας 1.

Πίνακας 1: Υπερπαράμετροι αλγορίθμου PPO

Παράμετρος	**Περιγραφή**	**Τιμή**
policy	Το μοντέλο πολιτικής που χησιμοποιείται	Mlppolicy
learning_rate	Ρυθμός μάθησης	0.003
batch_size	Μέγεθος minibatch	50
gamma	Συντελεστής έκπτωσης	0.85

Επιπλέον, για να επισπευστεί η διαδικασία της εκμάθησης του αλγορίθμου χρησιμοποιήθηκε η τεχνική των παράλληλων περιβαλλόντων. Συγκεκριμένα, έτρεξαν ταυτόχρονα 20 περιβάλλοντα. Γενικά, με τη συγκεκριμένη τεχνική παρατηρούνται οι παρακάτω βελτιώσεις:

- Μείωση της συσχέτισης στο σύνολο των δεδομένων λόγω της συλλογής δεδομένων από πολλαπλές τροχιές (trajectories).
- Ταχύτερη συνολική συλλογή δεδομένων, γεγονός που βελτιώνει το χρονικό διάστημα που απαιτείται για να ληφθεί το ίδιο αποτέλεσμα.

Η πρώτη βελτίωση δεν παρατηρείται σε όλες τις περιπτώσεις αλλά πρέπει τα περιβάλλοντα που εκτελούνται παράλληλα να διαθέτουν τις παρακάτω ιδιότητες:

- Να χρησιμοποιούν μεθόδους εντός πολιτικής (on-policy) ή μεθόδους στις οποίες δεν υπάρχει μνήμη επανάληψης και
- Να χρησιμοποιούν συναρτήσεις προσέγγισης για τη συνάρτηση πολιτικής ή και αξίας.

Οι παραπάνω ιδιότητες ισχύουν για τον αλγόριθμο PPO, ο οποίος χρησιμοποιήθηκε για την ανάπτυξη του μοντέλου εναλλαγής των λωρίδων.

Η υλοποίηση του αλγορίθμου έγινε στο προγραμματιστικό περιβάλλον της python. Με τη βοήθεια των παρακάτω βιβλιοθηκών:

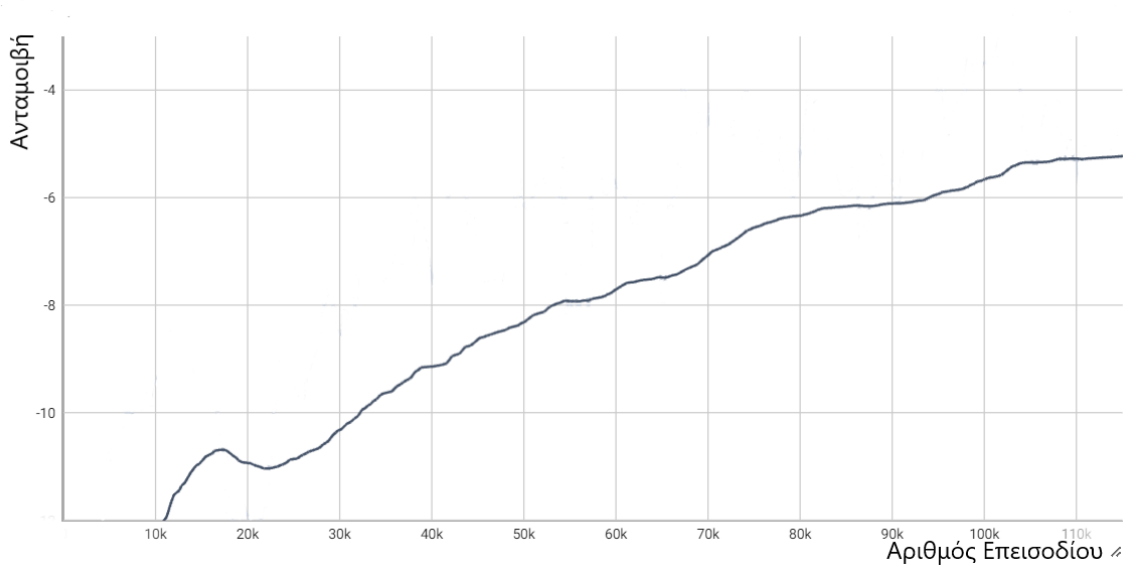
- Gymnasium, μία βιβλιοθήκη ανοιχτού κώδικα από την OpenAI που περιέχει ένα σύνολο τυποποιημένων περιβαλλόντων, ειδικά για αλγορίθμους ενισχυτικής μάθησης.
- Baselines, μια ακόμα βιβλιοθήκη ανοιχτού κώδικα από την OpenAI που περιέχει βελτιστοποιημένους αλγόριθμους ενισχυτικής μάθησης και
- TensorFlow, μία βιβλιοθήκη ανοιχτού κώδικα για μηχανική μάθηση, η οποία παρέχει χρήσιμα εργαλεία για την παρακολούθηση της εκπαίδευσης και της απόδοσης των μοντέλων.

Κεφάλαιο 5: Αξιολόγηση αποτελεσμάτων

5.1. Απόδοση μοντέλων

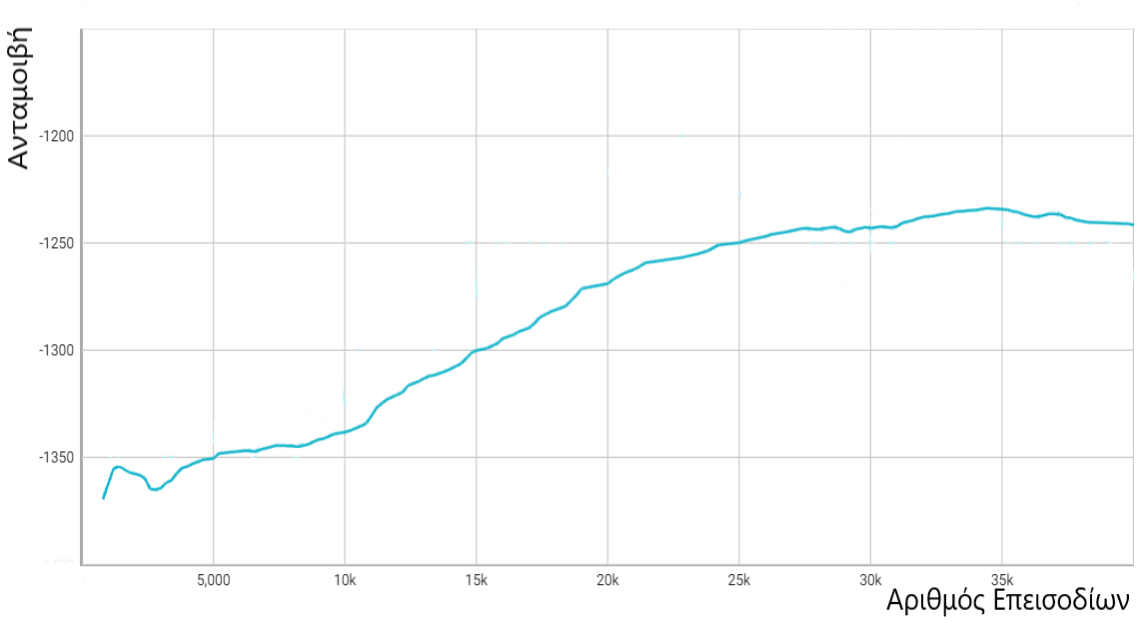
Στο προηγούμενο κεφάλαιο αναλύθηκε η μεθοδολογία που χρησιμοποιήθηκε για την εξαγωγή του βέλτιστου μοντέλου εναλλαγής λωρίδων. Στο παρόν κεφάλαιο παρουσιάζεται και αξιολογείται το μοντέλο αυτό σε διάφορα κυκλοφοριακά και γεωμετρικά σενάρια. Στόχος των διαφορετικών σεναρίων αξιολόγισης είναι η εξαγωγή συμπερασμάτων για την επιρροή των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας στα κυκλοφοριακά μεγέθη κυκλοφοριακά μεγέθη, όπως για παράδειγμα ο χρόνος διαδρομής ή η μέση ταχύτητα των οχημάτων του δικτύου και σε περιβαλλοντικά μεγέθη, όπως οι εκπομπές αερίων αλλά και η εξαγωγή συμπερασμάτων για την επιρροή της γεωμετρίας του δικτύου στην απόδοση του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας.

Αρχικά, όπως αναφέρθηκε και στην παράγραφο 4.2. το μοντέλο εκπαιδεύτηκε σε δύο φάσεις, σε προεκπαίδευση και το μοντέλο που προέκυψε από την προεκπαίδευση εκπαιδεύτηκε με τη νέα ανταμοιβή. Παρακάτω, στο Διάγραμμα 3 φαίνεται η πορεία της ανταμοιβής με το πέρας των επεισοδίων κατά τη φάση της προεκπαίδευσης.



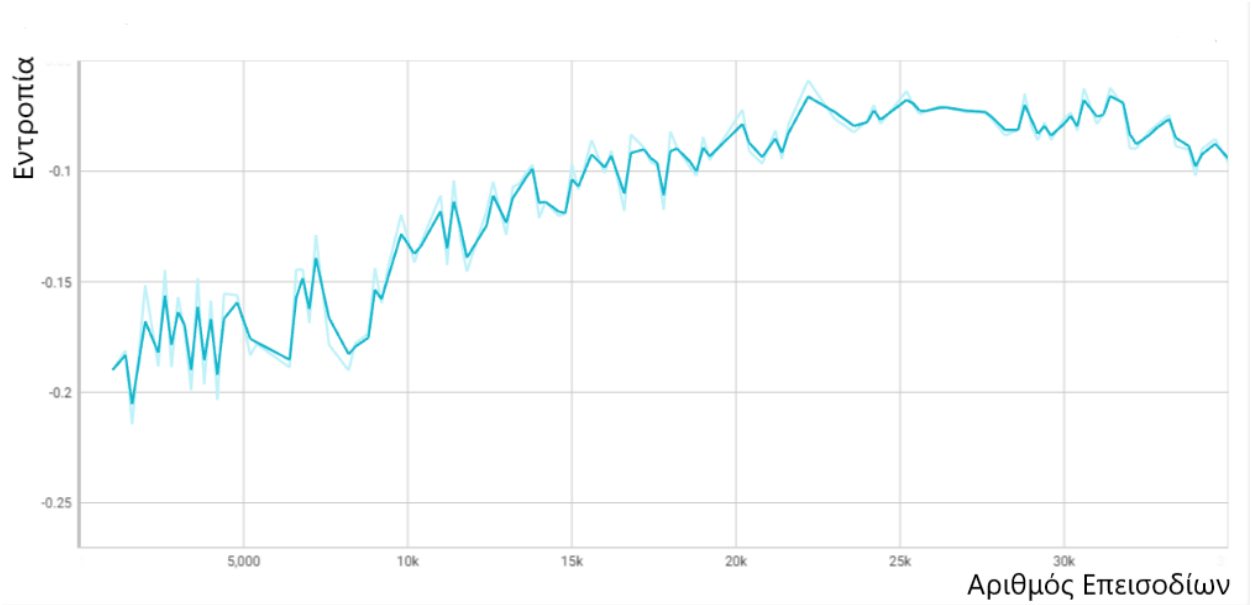
Διάγραμμα 3:Ανταμοιβή με το πέρας των επεισοδίων κατά τη φάση της προ-εκπαίδευσης.

Αυτό που μπορεί να παρατηρηθεί είναι ότι με το πέρας των επεισοδείων η ανταμοιβή που λαμβάνει ο πράκτορας αυξάνεται και μετά το εκατοστό χιλιοστό επεισόδιο αρχίζει και σταθεροποιείται κοντά στην τιμή -4. Το αντίστοιχο διάγραμμα αλλά για τη φάση της εκπαίδευσης φαίνεται παρακάτω στο Διάγραμμα 4.



Διάγραμμα 4:Ανταμοιβή με το πέρας των επεισοδίων κατά τη φάση της εκπαίδευσης.

Και σε αυτή την περίπτωση παρατηρούμε αύξηση της ανταμοιβής και σταθεροποίηση της μετά το τεσσαρακοστό χιλιοστό επεισόδιο. Μία ακόμα μετρική, η οποία φανερώνει ότι ο αλγόριθμος μαθαίνει με το πέρας των επεισοδίων είναι η εντροπία. Η εντροπία δείχνει πόσο τυχαίες είναι οι αποφάσεις που παίρνει ο πράκτορας και θα πρέπει να μειώνεται αργά κατά τη διάρκεια μιας επιτυχημένης διαδικασίας μάθησης. Παρακάτω στο Διάγραμμα 5 φαίνεται η εντροπίας για το τελικό μοντέλο.



Διάγραμμα 5: Διάγραμμα εντροπίας κατά τη φάση εκπαίδευσης

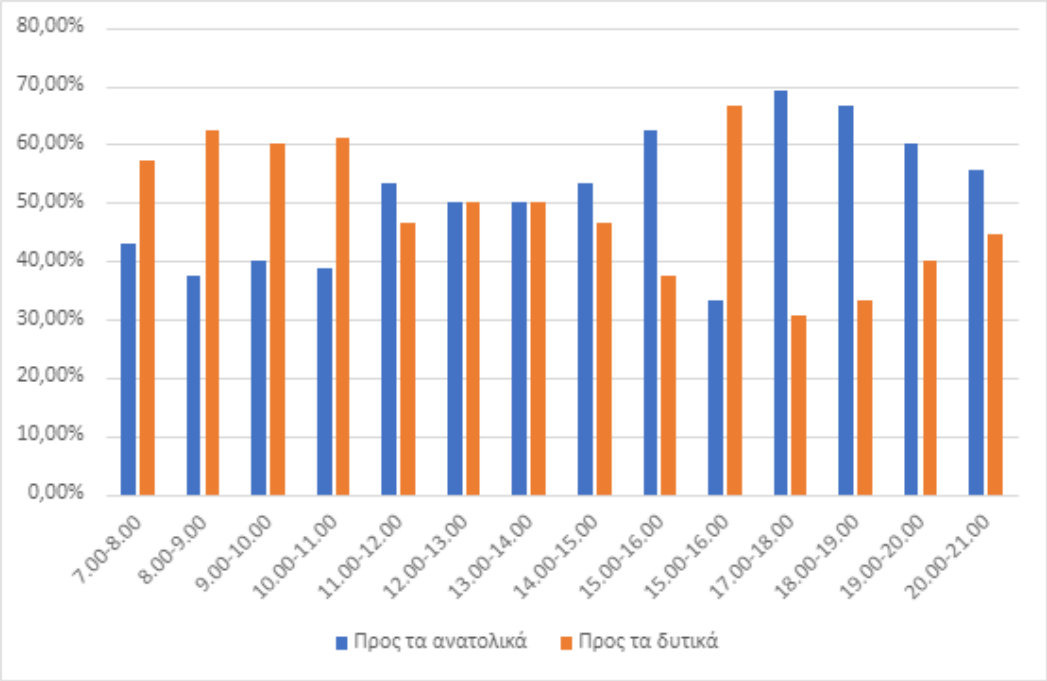
Η σημαντικότερη βελτίωση που παρατηρήθηκε από την προ-εκπαίδευση στην εκπαίδευση αφορά τις περιπτώσεις χαμηλών κυκλοφοριακών φόρτων και στην πρώτη διαμόρφωση, δηλαδή τη διαμόρφωση με τρεις λωρίδες κυκλοφορίας ανά κατεύθυνση. Καθώς πλέον ο αλγόριθμος μαθαίνει να διατηρεί τη πρώτη διαμόρφωση όταν η διαφορά ανάμεσα στη ζήτηση των δύο κατευθύνσεων είναι μικρή ή όταν ο συνολικός φόρτος του δικτύου είναι χαμηλός.

5.2. Σενάρια Αξιολόγησης

Το καλύτερο μοντέλο εναλλασσόμενων λωρίδων που αναπτύχθηκε μέσω του αλγορίθμου PPO αξιολογείται σε διαφορετικά σενάρια ζήτησης, που προσομοιώνουν καθημερινές κυκλοφοριακές καταστάσεις. Επιπλέον, αξιολογείται και σε διαφορετικές γεωμετρικές συνθήκες του δικτύου, για να εξεταστεί η επιρροή της απόστασης των κόμβων του δικτύου στην απόδοση του μοντέλου.

5.2.1. Σενάρια κυκλοφοριακής ζήτησης

Αρχικά, το πρώτο σενάριο, στο οποίο αξιολογείται το μοντέλο, είναι ένα σενάριο πρωινής και απογευματινής αιχμής. Δηλαδή, οι χαρακτηριστικές περίοδοι που ο φόρτος λαμβάνει υψηλές τιμές το πρωί και το απόγευμα και η κυκλοφορία μετατοπίζεται από τη μία κατεύθυνση στην άλλη καθώς οι οδηγοί μετακινούνται από και προς τους χώρους εργασίας τους. Συγκεκριμένα δημιουργήθηκαν τρία σενάρια δεκατεσσάρων ωρών, που διαφέρουν στην κυκλοφοριακή ζήτηση, αλλά βασίζονται στην ίδια ποσοστιαία κατανομή του φόρτου στις δύο κατευθύνσεις. Αυτή η κατανομή βασίζεται στην έρευνα των *Zheng et al., 2019* και φαίνεται στο Διάγραμμα 6:



Διάγραμμα 6:Ποσοστιαία κατανομή της κυκλοφοριακής ζήτησης στις δύο κατευθύνσεις κύριας οδού για περίοδο δεκατεσσάρων ωρών.

Το δεύτερο σενάριο που δημιουργήθηκε είναι ένα σενάριο εξυπηρέτησης μεγάλης εκδήλωσης. Συγκεκριμένα, δημιουργήθηκαν τρία σενάρια τα οποία διαφέρουν στον αριθμό των ατόμων που συμμετέχουν στην εκδήλωση. Η ποσοστιαία χρονική κατανομή των οχημάτων από και προς την εκδήλωση βασίστηκε στην έρευνα των *Amini et al., 2016* και έχει τα χαρακτηριστικά που φαίνονται στον Πίνακας 2.

Πίνακας 2:Ποσοστιαία κατανομή αφίξεων και αναχωρήσεων σε μεγάλη εκδήλωση

Άφιξη	**Ποσοστό**
1-2 ώρες πριν	32%
<1 ώρα πριν	56%
Μετά την έναρξη	12%
Αναχώρηση	**Ποσοστό**
Πριν τη λήξη	10%
<1 ώρα μετά	72%
>1 ώρα μετά	18%

Επιπλέον, για τη δημιουργία των σεναρίων έγιναν και οι εξής παραδοχές:

- Ποσοστό συμμετεχόντων που χρησιμοποιούν τα μέσα μαζικής μεταφοράς: 40%
- Μέση κατάληψη οχημάτων: 2,1 άτομα ανά όχημα.

Τέλος, δημιουργήθηκε και ένα σενάριο, το οποίο προσομοιάζει την περίπτωση οδικού ατυχήματος ή το κλείσιμο ενός μέρους μιας λωρίδας λόγω έκτακτης ανάγκης. Συγκεκριμένα, δημιουργήθηκε σενάριο δύο ωρών, στο οποία ένα όχημα ακινητοποιείται στα πρώτα δεκαπέντε λεπτά της προσομοίωσης στην άνω λωρίδα με κατεύθυνση προς τα δυτικά και επανέρχεται πάλι στην κυκλοφορία μετά το πέρας μίας ώρας. Ο φόρτος του σεναρίου αυτού και στις δύο κατευθύνσεις είναι ο ίδιος.

5.2.2. Σενάρια γεωμετρίας του δικτύου

Τα γεωμετρικά σενάρια στα οποία αξιολογείται το μοντέλο είναι τέσσερα και διαφοροποιούνται στις αποστάσεις που έχουν οι τρεις κόμβοι του οδικού δικτύου, στο οποίο αναπτύχθηκε το μοντέλο. Αναλυτικά, τα σενάρια παρουσιάζονται παρακάτω:

- Το αρχικό σενάριο, δηλαδή με απόσταση 250 μέτρων ανάμεσα στους κόμβους
- Το οδικό τμήμα των τριών κόμβων με την μεταξύ τους απόσταση να είναι 350 μέτρα
- Το οδικό τμήμα των τριών κόμβων με την μεταξύ τους απόσταση να είναι 150 μέτρα

Τελικώς, για λόγους σύγκρισης όλοι οι συνδυασμοί των παραπάνω σεναρίων αξιολογούνται σε τρείς περιπτώσεις:

- Στο μοντέλο δυναμικά εναλλασσόμενων λωρίδων που εκπαιδεύτηκε
- Στο μοντέλο της προεκπαίδευσης, το οποίο δίνει μία επιπλέον λωρίδα κυκλοφορίας στην κατεύθυνση με τον μεγαλύτερο φόρτο και
- Σε ένα στατικό περιβάλλον λωρίδων κυκλοφορίας

5. 3. Συγκριτική αξιολόγηση αποτελεσμάτων

Τα αποτελέσματα όλων των σεναρίων αξιολογούνται σε μετρικές που αφορούν τόσο κυκλοφοριακά χαρακτηριστικά, όσο και σε εκπομπές καυσαερίων και κατανάλωση καυσίμων. Συγκεκριμένα, στα κυκλοφοριακά χαρακτηριστικά λαμβάνονται υπόψιν τα εξής:

- Η ταχύτητα (Speed), ως ο μέσος όρος των μέσων ταχυτήτων των οχημάτων που εξυπηρετήθηκαν κατά τη διάρκεια της προσομοίωσης
- Η διάρκεια ταξιδιού (Duration), ως ο μέσος όρος της μέσης διάρκειας των ταξιδιών των οχημάτων που εξυπηρετήθηκαν κατά τη διάρκεια της προσομοίωσης και
- Η συνολική διάρκεια ταξιδιού και καθυστέρηση (Total travel time & Delay), ως ο μέσος όρος της μέσης συνολικής διάρκειας και καθυστέρησης των οχημάτων που εξυπηρετήθηκαν κατά τη διάρκεια της προσομοίωσης.

Κατά τη σύγκριση προσομοιώσεων με ίδιο χρόνο λήξης, τα αποτελέσματα μπορεί να αλλοιωθούν γιατί αυτές οι προσομοιώσεις μπορεί να διαφέρουν ως προς τον αριθμό των οχημάτων που αναχώρησαν και έφθασαν. Για αυτό το λόγο από τα παραπάνω χαρακτηριστικά μεγέθη το πιο δίκαιο για την σύγκριση διαφορετικών προσομοιώσεων είναι το τελευταίο, δηλαδή η συνολική διάρκεια ταξιδιού και καθυστέρηση, καθώς συνυπολογίζει και τα οχήματα που δεν πρόλαβαν να εισέλθουν στην προσομοίωση. Ο τύπος με τον οποίο υπολογίζεται το μέγεθος αυτό φαίνεται παρακάτω:

$$Total\ Travel\ Time\ \&\ Delay = inserted \times (duration + departDelay) + waiting \times departDelayWaiting \quad (5.1)$$

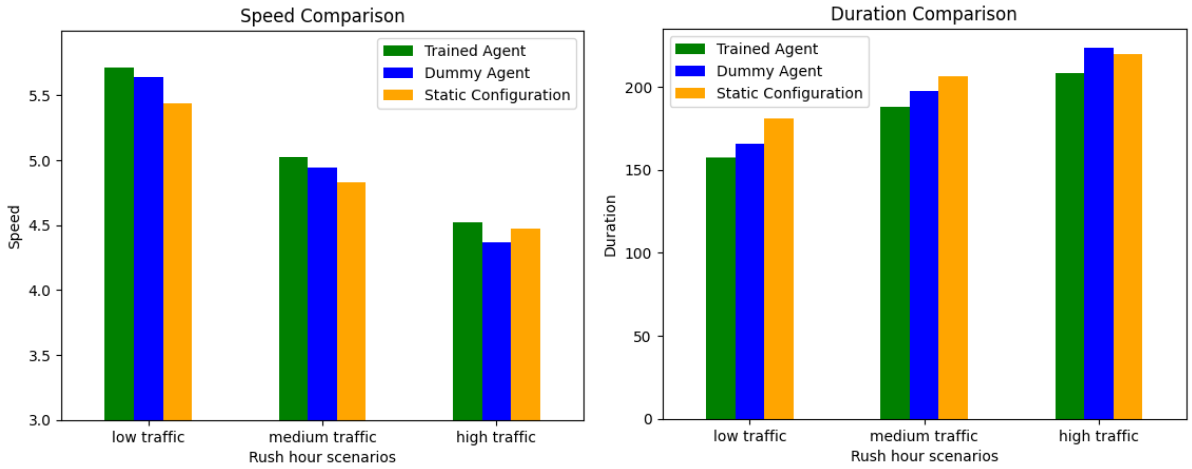
Όπου:

- inserted: ο αριθμός των οχημάτων που έχουν εισαχθεί στην προσομοίωση
- duration: η μέση διάρκεια ταξιδιού
- departDelay: ο μέσος χρόνος που τα οχήματα απαιτήθηκε να περιμένουν πριν ξεκινήσουν τη διαδρομή τους
- waiting: ο αριθμός οχημάτων με καθυστερημένη εισαγωγή που περιμένουν ακόμη για εισαγωγή στο τέλος της προσομοίωσης
- departDelayWaiting: ο μέσος χρόνος αναμονής των οχημάτων που δεν μπόρεσαν να εισαχθούν λόγω έλλειψης χώρου στην προσομοίωση

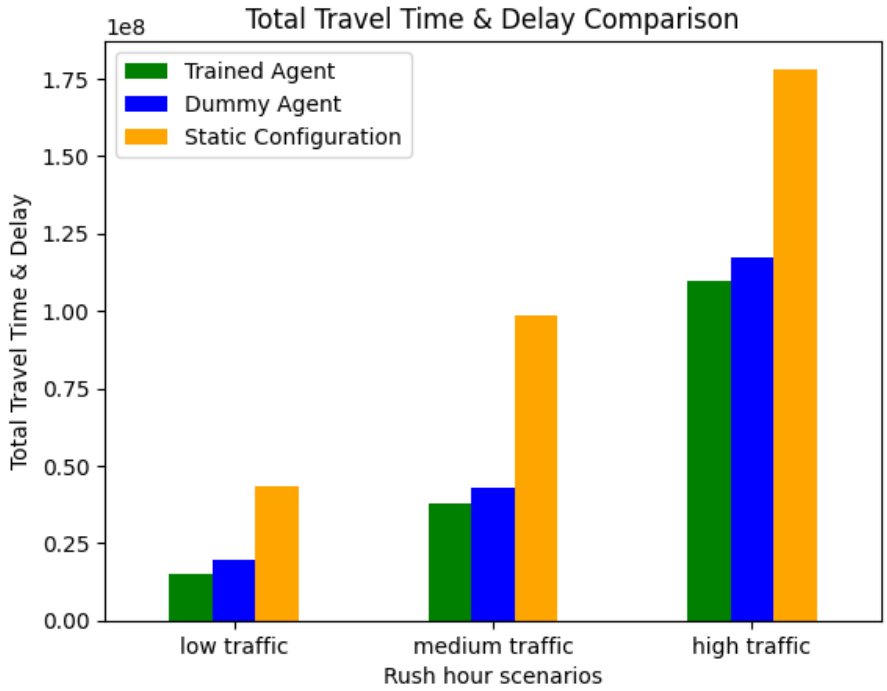
Στις εκπομπές καυσαερίων συγκρίνεται η μέση εκπομπή όλων των οχημάτων για το διοξείδιο του άνθρακα (CO_2), το μονοξείδιο του άνθρακα (CO) και για τα οξείδια του αζώτου (NO_x). Επίσης συγκρίνεται και η μέση κατανάλωση καυσίμων για όλα τα οχήματα της προσομοίωσης.

5.3.1. Συγκριτικά αποτελέσματα σεναρίων πρωινής-απογευματινής αιχμής

Στα Διαγράμματα 7 και 8 φαίνονται τα συγκριτικά αποτελέσματα στα σενάρια πρωινής και απογευματινής αιχμής, για το εκπαιδευμένο μοντέλο (trained agent), το μοντέλο το οποίο δίνει μία επιπλέον λωρίδα στην κατεύθυνση με το μεγαλύτερο φόρτο (dummy agent) και τη στατική διαμόρφωση του δικτύου (static configuration). Τα τρία διαφορετικά σενάρια αφορούν διαφορετικούς κυκλοφοριακούς φόρτους.

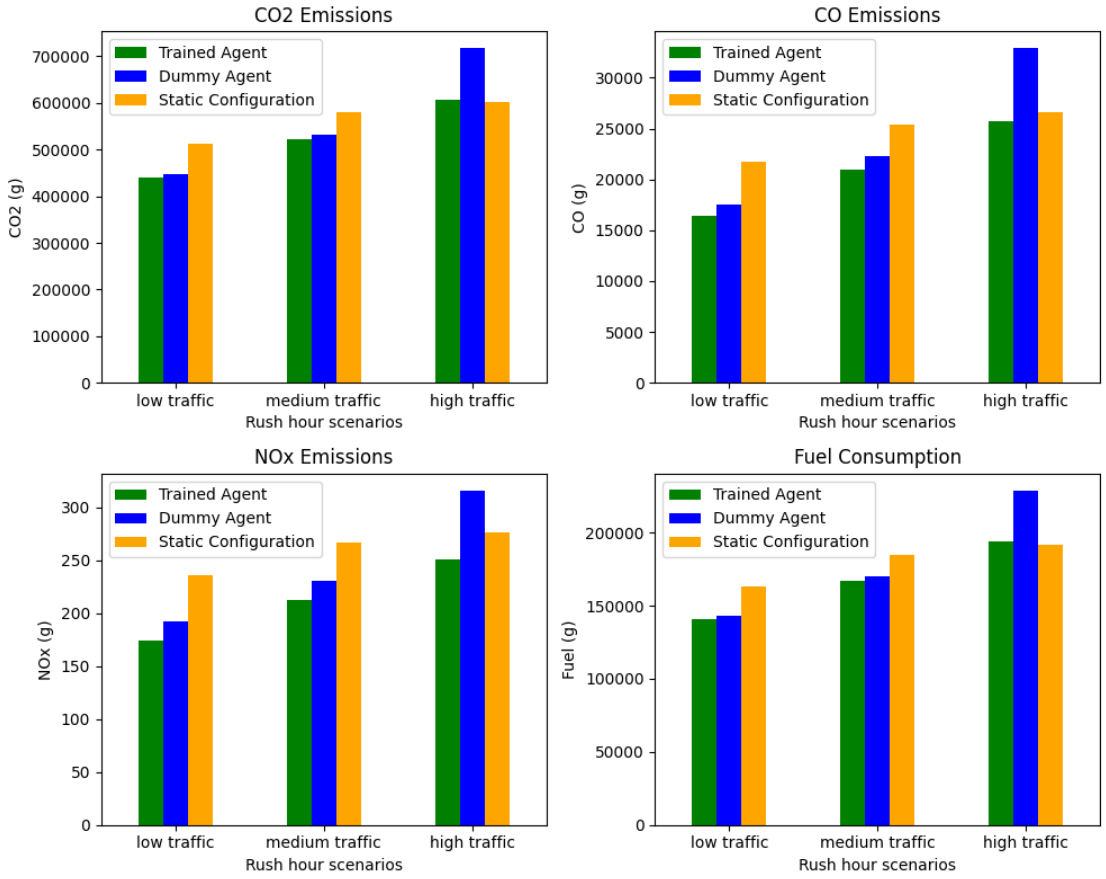


Διάγραμμα 7: Συγκριτικό διάγραμμα ταχυτήτων και διάρκειας διαδρομών- Πρωινή και απογευματινή αιχμή



Διάγραμμα 8: Συγκριτικό διάγραμμα συνολικής διάρκειας ταξιδιού και καθυστέρησης- Πρωινή και απογευματινή αιχμή

Αυτό που γίνεται αντιληπτό από τα διαγράμματα είναι βελτίωση των συνθηκών του δικτύου με τη χρήση των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας. Συγκεκριμένα, παρατηρήθηκε αύξηση στη μέση ταχύτητας των οχημάτων έως και 5% και μείωση της συνολικής διάρκειας ταξιδιού και καθυστέρησης έως και 25%. Επιπλέον, στο σενάριο με την μεγαλύτερη κυκλοφοριακή ζήτηση φαίνεται ότι η στατική διαμόρφωση του δικτύου λειτουργεί πιο αποδοτικά από το μοντέλο της προ-εκπαίδευσης.



Διάγραμμα 9: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων-Πρωινή και απογευματινή αιχμή.

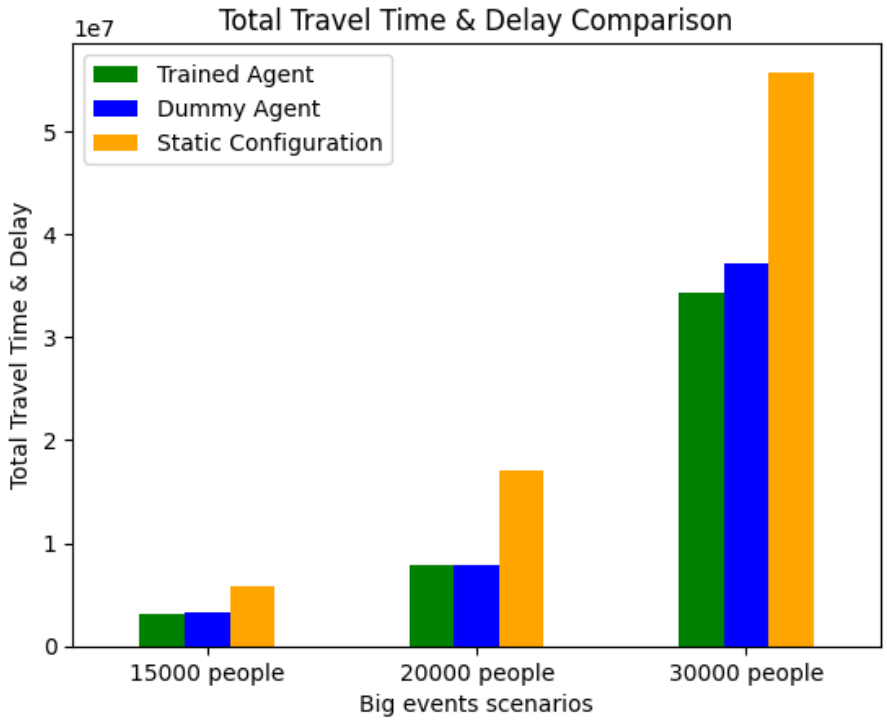
Όσον αφορά τις εκπομπές αερίων στο Διάγραμμα 9 γενικά παρατηρείται μείωση σχεδόν όλων των εκπομπών αερίων. Στο σενάριο του μεγάλου κυκλοφοριακού φόρτου, οι τιμές του CO_2, CO και της κατανάλωσης καυσίμων στο εκπαιδευμένο μοντέλο και τη στατική διαμόρφωση των λωρίδων δεν διαφέρουν σημαντικά. Παράλληλα σε αυτό το σενάριο, το μοντέλο της προ-εκπαίδευσης δίνει πολύ χειρότερα αποτελέσματα.

Στο ΠΑΡΑΡΤΗΜΑ Α της παρούσας εργασίας, φαίνονται τα διαγράμματα των αντίστοιχων μεγεθών των σεναρίων πρωινής και απογευματινής αιχμής για τις δύο επιπλέον διαφορετικές γεωμετρικές διαμορφώσεις του δικτύου που έχουν δημιουργηθεί. Η μορφή των αποτελεσμάτων αυτών δεν παρουσιάζει κάποια σημαντική διαφορά σε σχέση με αυτά του αρχικού δικτύου.

5.3.2. Συγκριτικά αποτελέσματα σεναρίων μεγάλων εκδηλώσεων

Στα Διαγράμματα 10 και 11 φαίνονται τα συγκριτικά αποτελέσματα στα σενάρια που αφορούν μεγάλες εκδηλώσεις, για το εκπαιδευμένο μοντέλο (trained agent), το μοντέλο το οποίο δίνει μία επιπλέον λωρίδα στην κατεύθυνση με το μεγαλύτερο φόρτο (dummy agent) και τη στατική διαμόρφωση του δικτύου (static configuration). Τα τρία διαφορετικά σενάρια αφορούν διαφορετική προσέλευση ατόμων στην εκδήλωση.

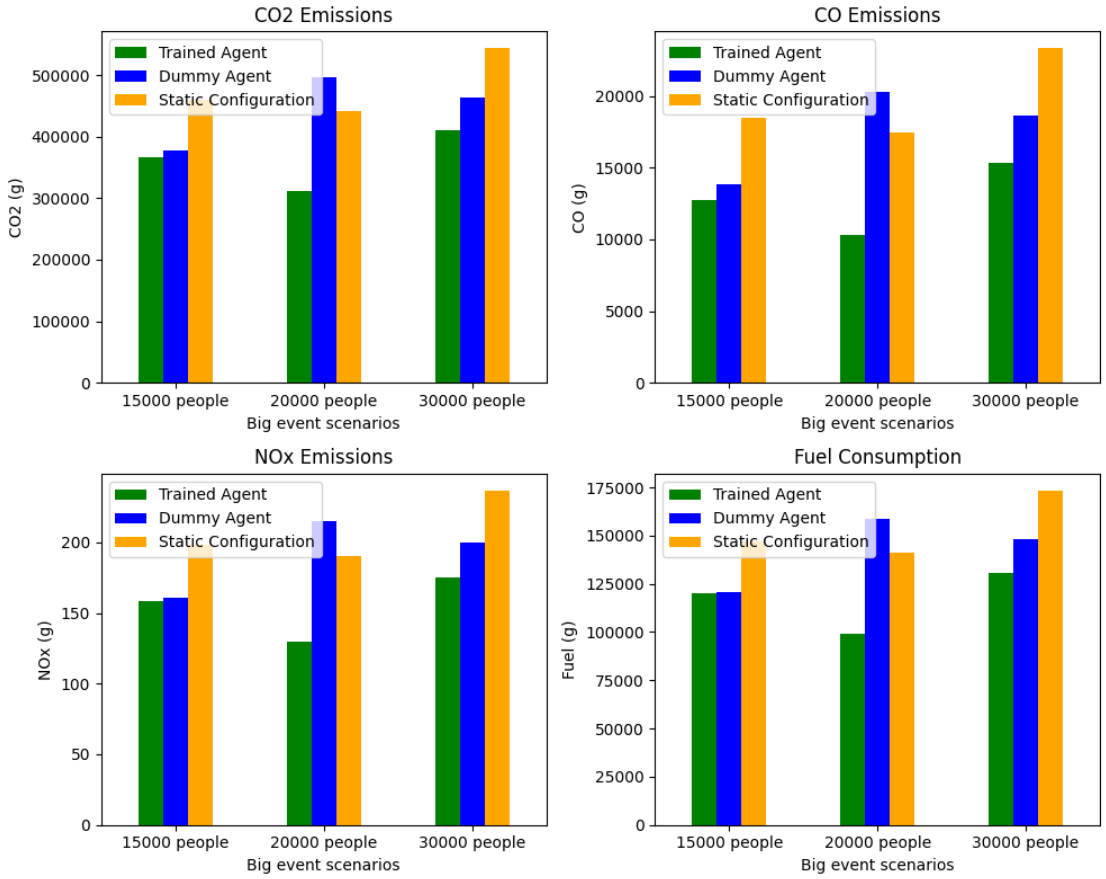
Διάγραμμα 10: Συγκριτικά διαγράμματα ταχυτήτων και διάρκειας διαδρομών - Μεγάλες εκδηλώσεις



Διάγραμμα 11: Συγκριτικό διάγραμμα συνολικής διάρκειας ταξιδιού και καθυστέρησης- Μεγάλες εκδηλώσεις

Και στην περίπτωση των μεγάλων εκδηλώσεων παρατηρούνται καλύτερες συνθήκες κυκλοφορίας με τη χρήση του μοντέλου που εκπαιδεύτηκε για τη δυναμική εναλλαγή των λωρίδων κυκλοφορίας. Συγκεκριμένα, παρατηρείται έως και 10% αύξηση των ταχυτήτων και έως και 40% μείωση του συνολικού χρόνου ταξιδιού και καθυστερήσεων, σε σχέση με τη στατική διαμόρφωση του οδικού χώρου. Πάλι παρατηρούνται καλύτερα αποτελέσματα και σε σχέση με το μοντέλο της προεκπαίδευσης.

Τις μεγαλύτερες βελτιώσεις σε αυτό το σενάριο τις παρατηρούμε στις συνθήκες του μεγαλύτερου κυκλοφοριακού φόρτου, σε αντίθεση με το σενάριο της πρωινής και απογευματινής αιχμής. Αυτό συμβαίνει, επειδή στην περίπτωση των μεγάλων εκδηλώσεων υπάρχει μεγαλύτερη διαφορά ανάμεσα στους κυκλοφοριακούς φόρτους των δύο κατευθύνσεων, σε σχέση με ένα σενάριο πρωινής και απογευματινής αιχμής. Άρα, η απόδοση μίας επιπλέον λωρίδας στην κατεύθυνση με τον μεγαλύτερο φόρτο δεν δημιουργεί σημαντικά χειρότερες συνθήκες στην αντίθετη κατεύθυνση.



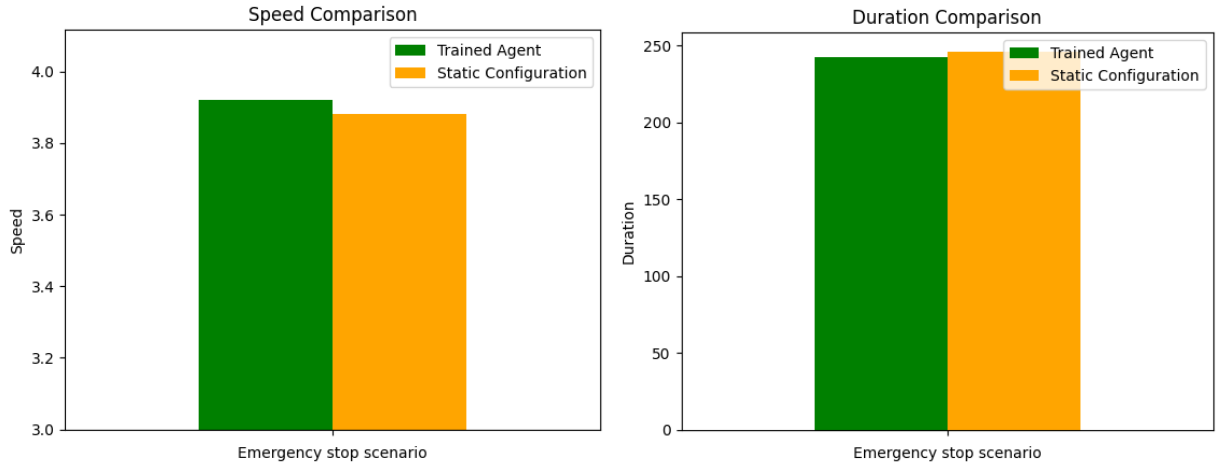
Διάγραμμα 12: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων-Μεγάλες εκδηλώσεις

Όσον αφορά τις εκπομπές αερίων γενικά παρατηρείται από το Διάγραμμα 12 μείωση όλων των εκπομπών αερίων. Στο σενάριο της μικρότερης προσέλευσης η μείωση των εκπομπών καυσαερίων είναι πολύ μικρότερη, σε σχέση σε το μοντέλο προ-εκπαίδευσης, το οποίο δίνει ικανοποιητικά αποτελέσματα.

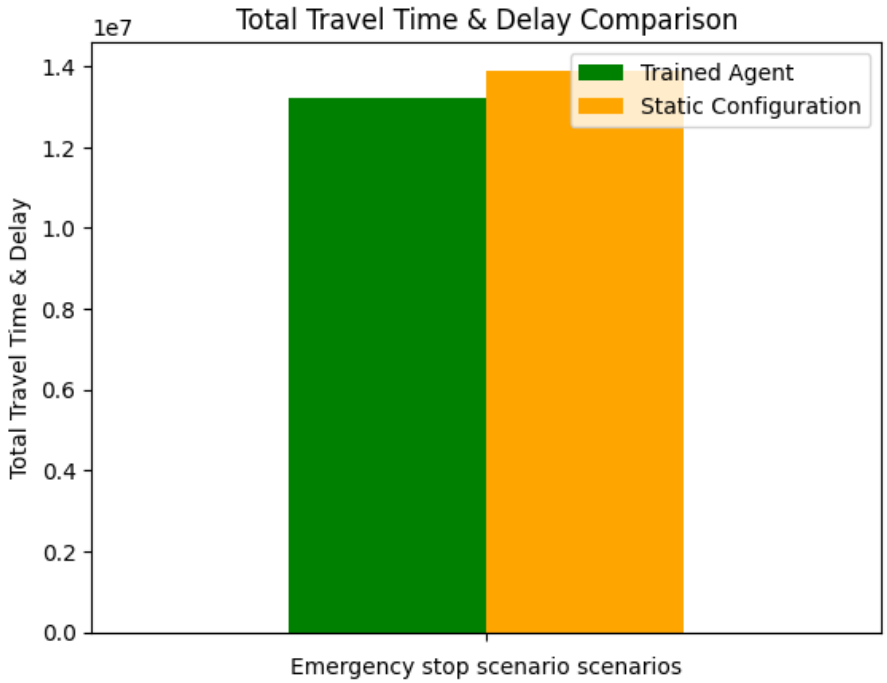
Στο ΠΑΡΑΡΤΗΜΑ Α της παρούσας εργασίας, φαίνονται τα διαγράμματα των αντίστοιχων μεγεθών των σεναρίων μεγάλων εκδηλώσεων για τις δύο επιπλέον διαφορετικές γεωμετρικές διαμορφώσεις του δικτύου που έχουν δημιουργηθεί. Η μορφή των αποτελεσμάτων αυτών δεν παρουσιάζει κάποια σημαντική διαφορά σε σχέση με αυτά του αρχικού δικτύου.

5.3.3. Συγκριτικά αποτελέσματα σεναρίων αναγκαστικής διακοπής κυκλοφορίας

Τέλος, στα Διαγράμματα 13 και 14 φαίνονται τα συγκριτικά αποτελέσματα στο σενάριο που αφορά την αναγκαστική στάση οχήματος και το κλείσιμο ενός μέρους μιας λωρίδας μήκους 100 μέτρων, για το εκπαιδευμένο μοντέλο (trained agent) και τη στατική διαμόρφωση του δικτύου (static configuration). Το σενάριο αυτό δεν αξιολογείται στο μοντέλο της προ-εκπαίδευσης, καθώς ο φόρτος που έχει χρησιμοποιηθεί είναι ίσος και στις δύο κατευθύνσεις. Άρα θα έδινε τα ίδια αποτελέσματα με τη στατική διαμόρφωση.

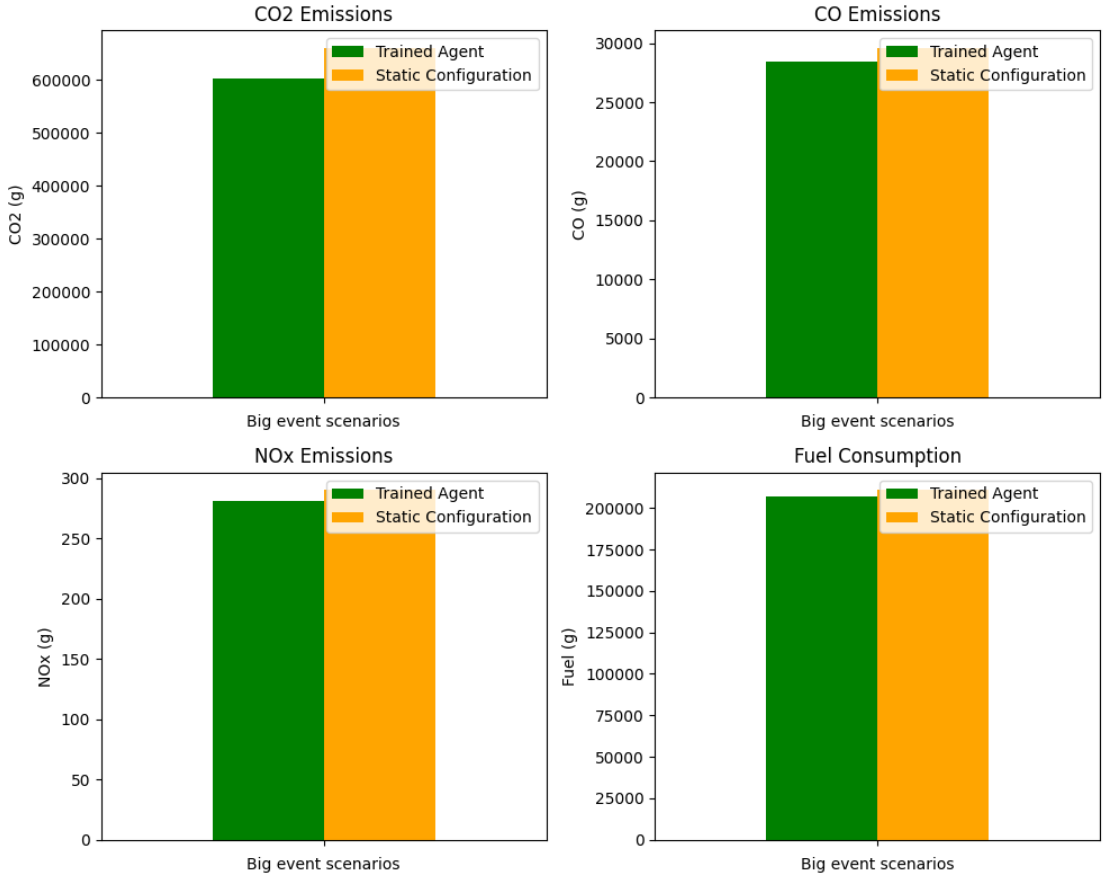


Διάγραμμα 13: Συγκριτικά διαγράμματα ταχυτήτων και διάρκειας διαδρομών-Αναγκαστική στάση



Διάγραμμα 14: Συγκριτικό διάγραμμα συνολικής διάρκειας ταξιδιού και καθυστέρησης-Αναγκαστική στάση

Στο σενάριο της αναγκαστικής στάσης ενός οχήματος στο οδικό δίκτυο, παρατηρούνται και πάλι καλύτερες συνθήκες με τη χρήση του μοντέλου που αναπτύχθηκε. Ωστόσο, η διαφορά σε σχέση με τη στατική διαμόρφωση δεν είναι πολύ μεγάλη. Σε αυτό παίζει ρόλο ο ίσος κυκλοφοριακός φόρτος και στις δύο κατευθύνσεις. Συγκεκριμένα, παρατηρείται αύξηση 1,5% στις ταχύτητες και μείωση 5% στο συνολικό χρόνο ταξιδιού και καθυστέρηση.



Διάγραμμα 15: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων-Αναγκαστική στάση

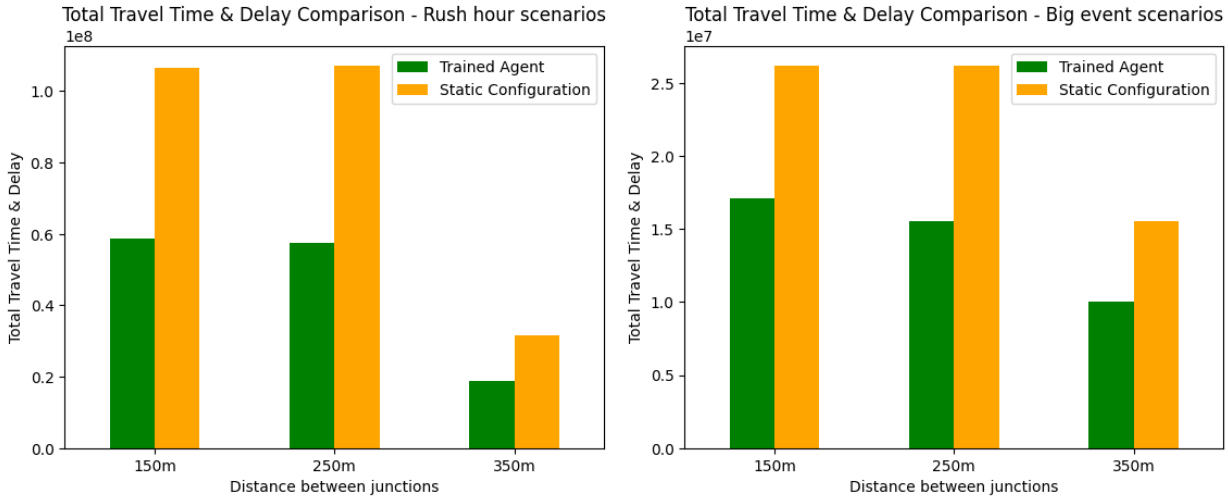
Και στο Διάγραμμα 15 φαίνεται μείωση των εκπομπών καυσαερίων και της κατανάλωσης καυσαερίων, όπως και στα προηγούμενα σενάρια που εξετάστηκαν.

Όπως και στις προηγούμενες περιπτώσεις τα υπόλοιπα διαγράμματα φαίνονται στο ΠΑΡΑΡΤΗΜΑ Α της παρούσας εργασίας.

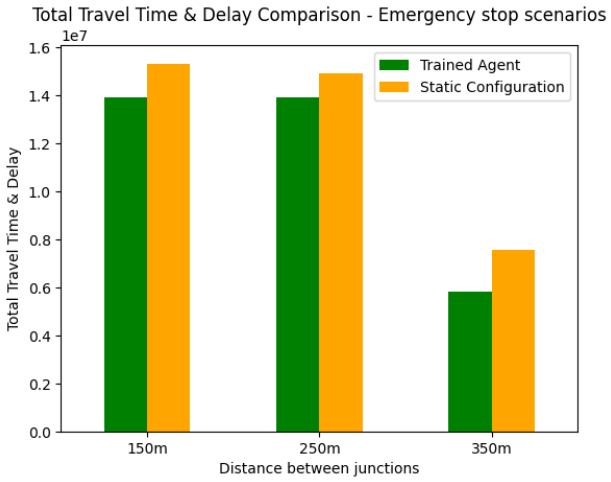
5.4. Σύγκριση διαφορετικών γεωμετρικών συνθηκών δικτύου

Εκτός της βελτίωσης των κυκλοφοριακών χαρακτηριστικών ενός οδικού δικτύου, στόχος της παρούσας εργασίας είναι η διερεύνηση της επιρροής της απόστασης μεταξύ των κόμβων στην αποδοτικότητα του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας. Για το λόγο αυτό συγκρίνεται ο μέσος συνολικός χρόνος ταξιδιού και καθυστερήσεων για κάθε σενάριο, ως η πιο δίκαιη σύγκριση, σε κάθε σενάριο που εξετάζεται και στις προηγούμενες παραγράφους.

Από τα Διαγράμματα Διάγραμμα 16Διάγραμμα 17 παρατηρείται μείωση του συνολικού χρόνου ταξιδιού και των καθυστερήσεων με την αύξηση της απόστασης των κόμβων σε ένα δίκτυο. Επιπλέον, τα οδικά τμήματα που είχαν απόσταση μεταξύ των κόμβων 150 μέτρα και 250 μέτρα δεν δίνουν πολύ διαφορετικά αποτελέσματα, δηλαδή οι κυκλοφοριακές συνθήκες του δικτύου καλυτερεύουν σημαντικά σε αποστάσεις μεταξύ των κόμβων άνω των 350 μέτρων.



Διάγραμμα 16: Σύγκριση συνολικού χρόνου ταξιδιού για διαφορετικές αποστάσεις μεταξύ των κόμβων – Πρωινή και απογευματινή αιχμή και Μεγάλες εκδηλώσεις



Διάγραμμα 17: Σύγκριση συνολικού χρόνου ταξιδιού για διαφορετικές αποστάσεις μεταξύ των κόμβων – Αναγκαστική στάση

Συγκεκριμένα, στο σενάριο της πρωινής και απογευματινής αιχμής παρατηρείται μείωση 55% του συνολικού χρόνου διαδρομής από τη γεωμετρική διαμόρφωση των 150 μέτρων σε αυτή των 350 μέτρων, στο σενάριο των μεγάλων εκδηλώσεων 37% και στο σενάριο της αναγκαστικής στάσης οχήματος 57%.

Κεφάλαιο 6: Συμπεράσματα

6.1. Βασικά συμπεράσματα

Η ολοένα αυξανόμενη κυκλοφοριακή ζήτηση απαιτεί συνεχώς νέες τεχνικές για την ορθή και αποτελεσματική διαχείριση της κυκλοφορίας. Μία τέτοια τεχνική, η οποία έχει ξεκινήσει να ερευνάται είναι οι δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας. Η παρούσα διπλωματική εργασία είχε ως αντικείμενο την ανάπτυξη ενός μοντέλου δυναμικής εναλλαγής των λωρίδων κυκλοφορίας σε αστικό διάδρομο τριών κόμβων. Στόχος του μοντέλου αυτού ήταν η βελτίωση των χαρακτηριστικών της κυκλοφορίας, όπως είναι η ταχύτητα των οχημάτων του δικτύου, ο χρόνος διαδρομής και οι εκπομπές καυσαερίων. Για την επίτευξη αυτού του σκοπού το μοντέλο εκπαιδεύτηκε με τη χρήση του αλγορίθμου Proximal Policy Optimization (PPO), σε δύο στάδια, ένα στάδιο προ-εκπαίδευσης και ένα στάδιο εκπαίδευσης. Έπειτα το εκπαιδευμένο μοντέλο αξιολογήθηκε σε τρία σενάρια πραγματικών κυκλοφοριακών συνθήκων και διαφορετικών γεωμετριών του δικτύου. Η σύγκριση των αποτελεσμάτων έγινε για τρεις τρόπους διαχείρισης των λωρίδων κυκλοφορίας, δηλαδή για διαχείριση με το εκπαιδευμένο μοντέλο, έπειτα για ένα μοντέλο το οποίο αποδίδει μία επιπλέον λωρίδα κυκλοφορίας στην κατεύθυνση με τον μεγαλύτερο φόρτο και τέλος για στατική διαμόρφωση των λωρίδων κυκλοφορίας.

Με την εισαγωγή του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας με τη χρήση του μοντέλου που αναπτύχθηκε στην παρούσα εργασία προκύπτουν καλύτερες συνθήκες κυκλοφορίας στο δίκτυο. Συνολικά, από την ανάλυση των αποτελεσμάτων των σεναρίων διαπιστώνεται αύξηση της μέσης ταχύτητας των οχημάτων έως και 10% και μείωση του συνολικού χρόνου ταξιδιού και καθυστερήσεων των οχημάτων έως και 40% σε σχέση με τη στατική διαμόρφωση των λωρίδων κυκλοφορίας. Σημαντική είναι επίσης και η μείωση στις εκπομπές καυσαερίων και της κατανάλωσης καυσίμων.

Επιπλέον, πολύ σημαντικό επίτευγμα του μοντέλου που αναπτύχθηκε με τη χρήση του αλγορίθμου PPO είναι ότι δίνει καλύτερα αποτελέσματα από ένα μοντέλο, το οποίο παρέχει την επιπλέον λωρίδα κυκλοφορίας στην κατεύθυνση του μεγαλύτερου φόρτου. Με αυτό τον τρόπο αποδεικνύεται η αναγκαιότητα της ανάπτυξης μοντέλων ενισχυτικής μάθησης για την εφαρμογή μέτρων διαχείρισης της κυκλοφορίας, όπως είναι οι δυναμικά εναλλασσόμενες λωρίδες κυκλοφορίας.

Τέλος, δείχθηκε ότι η απόσταση των κόμβων του οδικού τμήματος παίζει σημαντικό ρόλο στην αποδοτικότητα του μέτρου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας. Όσο πιο απομακρυσμένοι είναι οι κόμβοι μεταξύ τους, τόσο περισσότερο μειώνεται ο μέσος χρόνος ταξιδιού και καθυστερήσεων, μείωση που μπορεί να φτάσει και το 57% στην περίπτωση της αναγκαστικής στάσης οχήματος.

6.2. Προτάσεις για περαιτέρω έρευνα

Για τη βαθύτερη διερεύνηση του αντικειμένου των δυναμικά εναλλασσόμενων λωρίδων κυκλοφορίας, τα παρακάτω αποτελούν σημαντικά σημεία.

Στην παρούσα εργασία διερευνήθηκε η εφαρμογή των δυναμικά εναλλασσόμενων λωρίδων σε μία οδική αρτηρία τριών κόμβων, ωστόσο η εφαρμογή του μέτρου αυτού σε επίπεδο δικτύου θα μπορούσε να δώσει πιο ρεαλιστικά αποτελέσματα, καθώς μπορεί να διερευνηθεί η επιρροή των

κυκλοφοριακών συνθηκών όλων των οδών του δικτύου στην αποδοτικότητα του μέτρου, αλλά και το αντίστροφο, δηλαδή η επιρροή των δυναμικά εναλλασσόμενων λωρίδων στις

Ενδιαφέρον θα είχε η διερεύνηση και άλλων διαμορφώσεων των λωρίδων κυκλοφορίας, όπως για παράδειγμα τέσσερεις λωρίδες στην κατεύθυνση προς τα ανατολικά και μία στην κατεύθυνση προς τα δυτικά και το αντίστροφο. Ακόμα, θα μπορούσε να εξεταστούν και οδικά τμήματα με λιγότερες και περισσότερες λωρίδες κυκλοφορίας, ώστε να βρεθεί η ορθότερη διαμόρφωση του δικτύου για την αποτελεσματικότερη εφαρμογή του μέτρου.

Τέλος, ενδιαφέρον θα είχε και η διερεύνηση δυναμικά εναλλασσόμενων λωρίδων αυτή τη φορά όμως ως προς τα μέσα που επιτρέπουν να κυκλοφορούν επάνω τους. Στόχος θα ήταν να βελτιστοποιηθεί όχι μόνο η κυκλοφορία των οχημάτων αλλά η συνολική χρήση του χώρου στα πλαίσια των χρήσεων του ελεύθερου χώρου στον αστικό ιστό.

Βιβλιογραφία

Abdelkader, G., Elgazzar, K., & Khamis, A. (2021). Connected Vehicles: Technology Review, State of the Art, Challenges and Opportunities. *Sensors*, *21*(22), Article 22. https://doi.org/10.3390/s21227712

Alhajyaseen, W. K. M., Najjar, M., Ratrout, N. T., & Assi, K. (2017). The effectiveness of applying dynamic lane assignment at all approaches of signalized intersection. *Case Studies on Transport Policy*, *5*(2), 224–232. https://doi.org/10.1016/j.cstp.2017.01.008

Alvarez Lopez, P., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wießner, E. (2018). Microscopic Traffic Simulation using SUMO. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2575–2582. https://www.itsc2019.org/

Amini, S., Papapanagiotou, E., & Busch, F. (2016). *Traffic Management for Major Events* (pp. 187–197). https://doi.org/10.14459/2016md1324021

Hausknecht, M., Au, T.-C., Stone, P., Fajardo, D., & Waller, T. (2011). Dynamic lane reversal in traffic management. *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 1929–1934. https://doi.org/10.1109/ITSC.2011.6082932

Lu, T., Yang, Z., Ma, D., & Jin, S. (2018). Bi-Level Programming Model for Dynamic Reversible Lane Assignment. *IEEE Access*, *6*, 71592–71601. https://doi.org/10.1109/ACCESS.2018.2881290

Mao, L., Li, W., Hu, P., Zhou, G., Zhang, H., & Dai, J. (2020). Design of Real-Time Dynamic Reversible Lane in Intelligent Cooperative Vehicle Infrastructure System. *Journal of Advanced Transportation*, *2020*, 1–8. https://doi.org/10.1155/2020/8838896

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I.,

King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), Article 7540. https://doi.org/10.1038/nature14236

Pérez-Méndez, D., Gershenson, C., Lárraga, M. E., & Mateos, J. L. (2021). Modeling adaptive reversible lanes: A cellular automata approach. *PLOS ONE*, *16*(1), e0244326. https://doi.org/10.1371/journal.pone.0244326

Reversible Lane Systems: Synthesis of Practice. (2006). Retrieved 29 March 2023, from https://ascelibrary.org/doi/epdf/10.1061/%28ASCE%290733-947X%282006%29132%3A12%28933%29

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal Policy Optimization Algorithms* (arXiv:1707.06347). arXiv. http://arxiv.org/abs/1707.06347

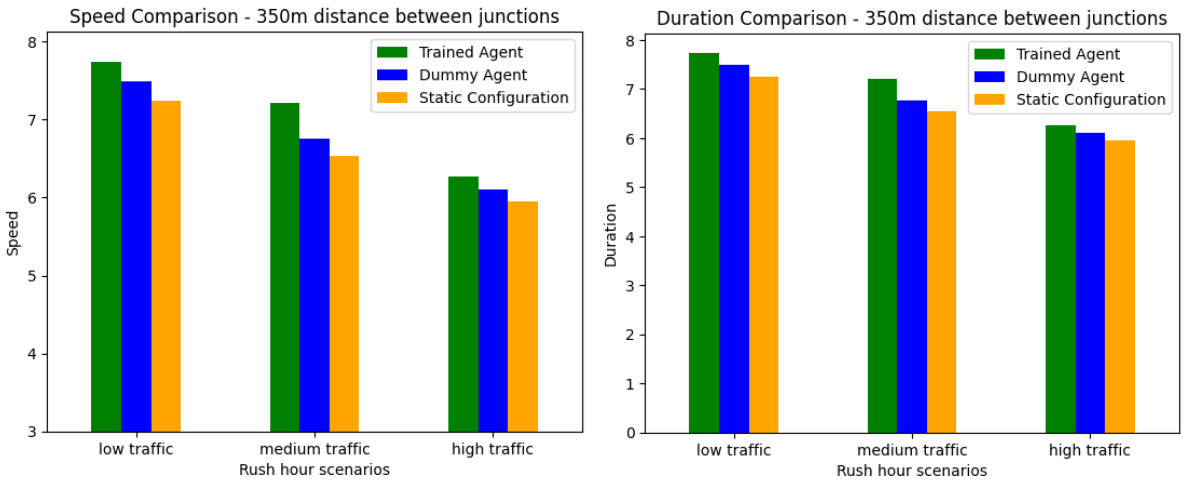
ITF*, Streets That Fit: Re-allocating Space for Better Cities*. (2022).

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning, second edition: An Introduction*. MIT Press.

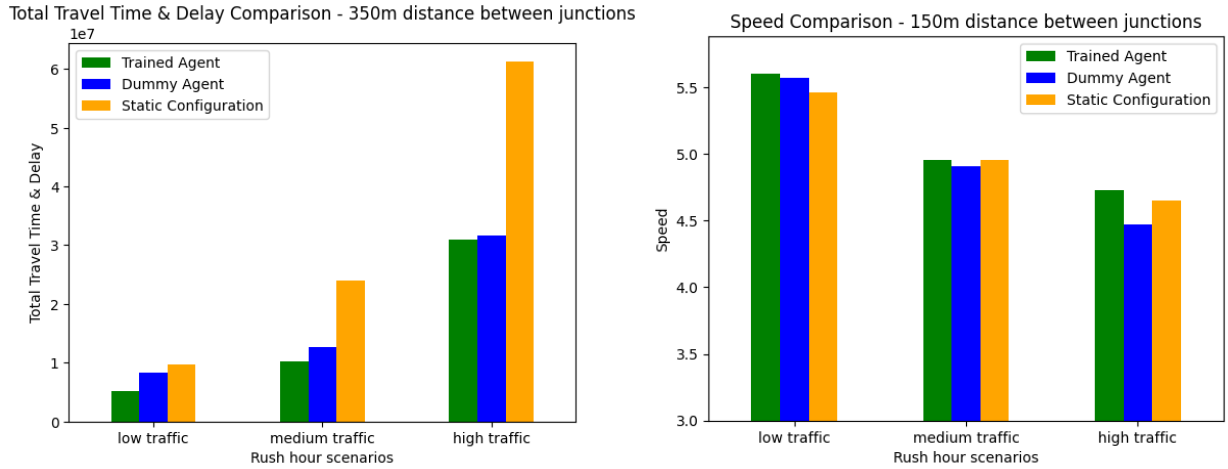
Uhlemann, E. (2015). Introducing Connected Vehicles [Connected Vehicles]. *IEEE Vehicular Technology Magazine*, *10*(1), 23–31. https://doi.org/10.1109/MVT.2015.2390920

Zheng, G., Xiong, Y., Zang, X., Feng, J., Wei, H., Zhang, H., Li, Y., Xu, K., & Li, Z. (2019). *Learning Phase Competition for Traffic Signal Control* (arXiv:1905.04722). arXiv. http://arxiv.org/abs/1905.04722

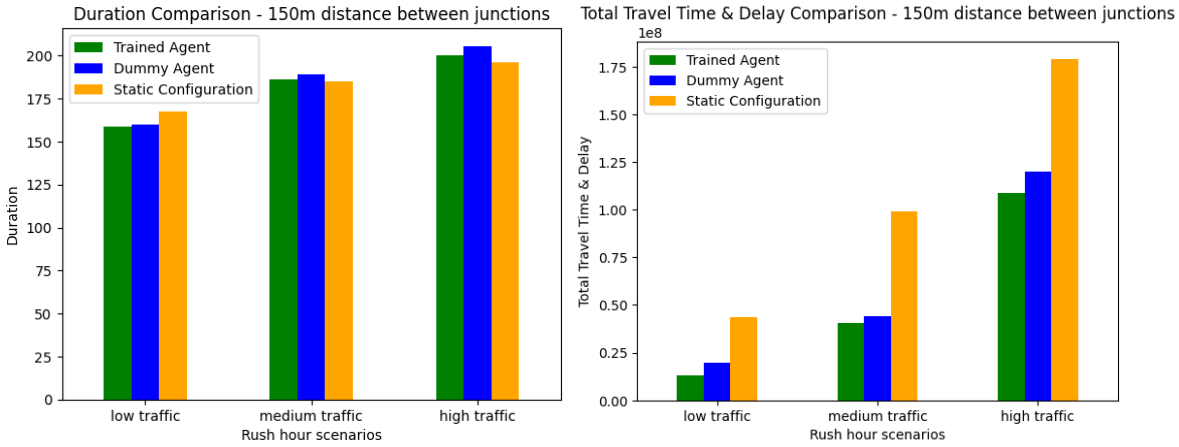
ΠΑΡΑΡΤΗΜΑ Α



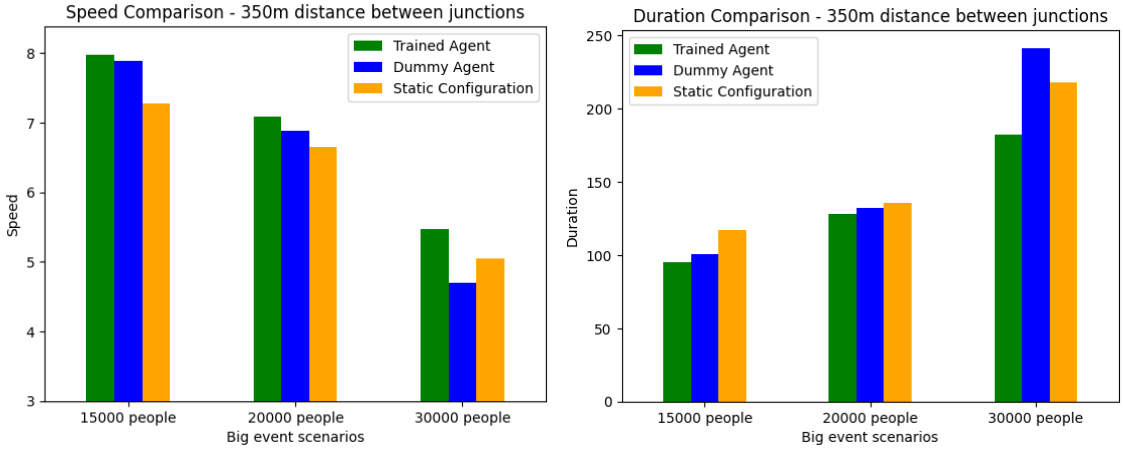
Διάγραμμα 18: Συγκριτικά διαγράμματα ταχυτήτων διάρκειας διαδρομών - Ώρες αιχμής- Μεγάλο δίκτυο



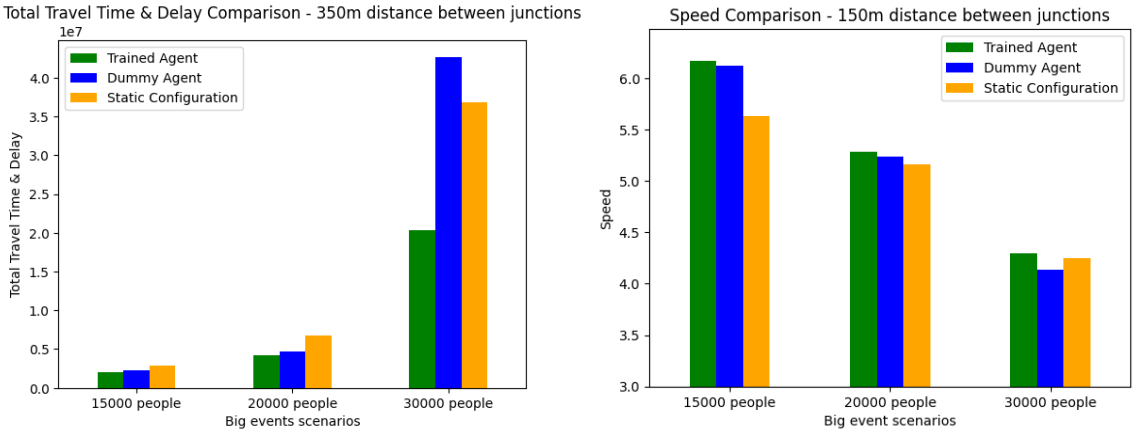
Διάγραμμα 19: Συγκριτικό διάγραμμα συνολικής διάρκειας ταξιδιού και καθυστέρησης(Μεγάλο δίκτυο) και ταχυτήτων (Μικρό δίκτυο) - Ώρες αιχμής



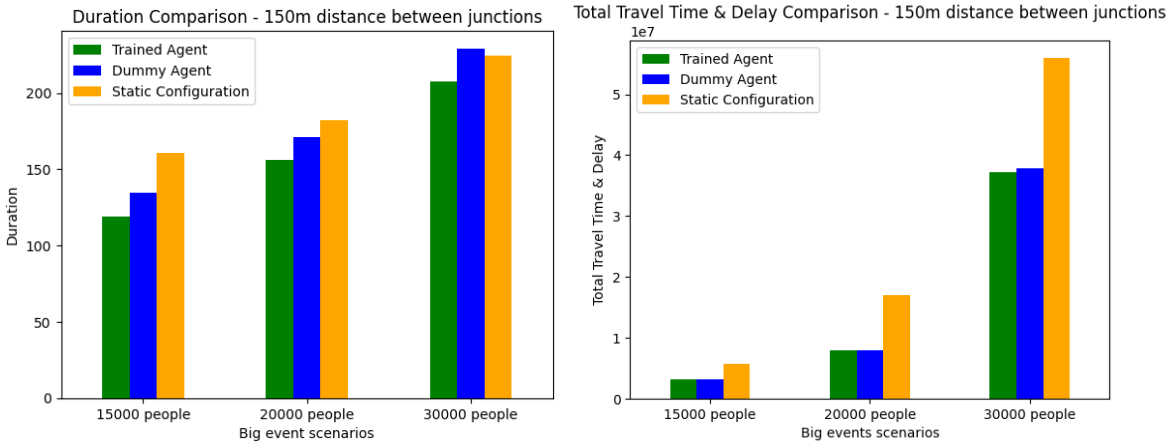
Διάγραμμα 20: Συγκριτικά διαγράμματα διάρκειας διαδρομών και συνολικής διάρκειας ταξιδιού και καθυστέρησης - Ώρες αιχμής- Μικρό δίκτυο



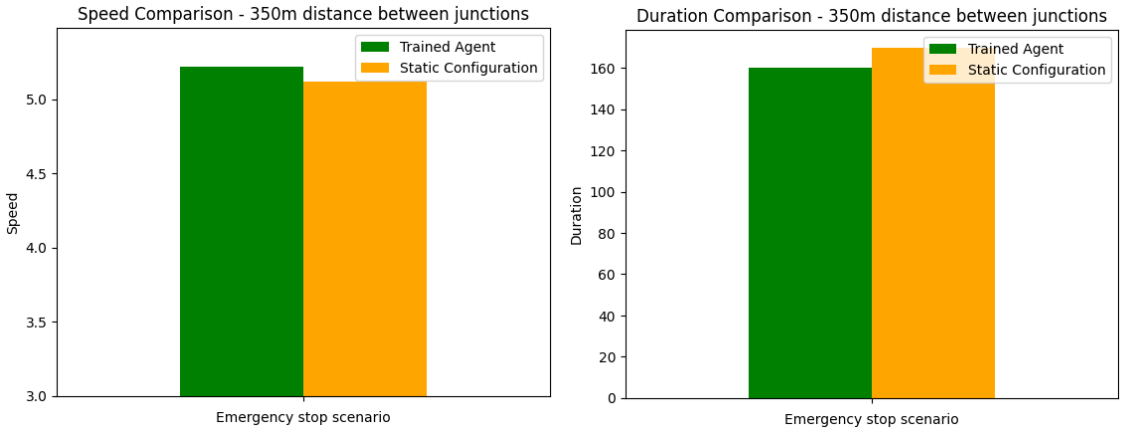
Διάγραμμα 21: Συγκριτικά διαγράμματα ταχυτήτων και διαγράμματα διάρκειας διαδρομών - Μεγάλες εκδηλώσεις-Μεγάλο δίκτυο



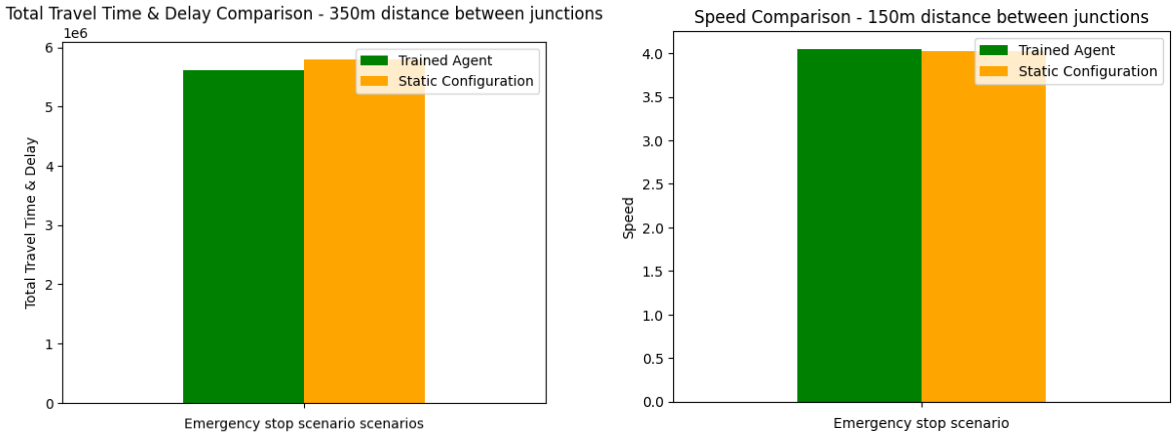
Διάγραμμα 22: Συγκριτικό διάγραμμα συνολικής διάρκειας ταξιδιού και καθυστέρησης (Μεγάλο δίκτυο) και ταχυτήτων (Μικρό δίκτυο) - Μεγάλες εκδηλώσεις



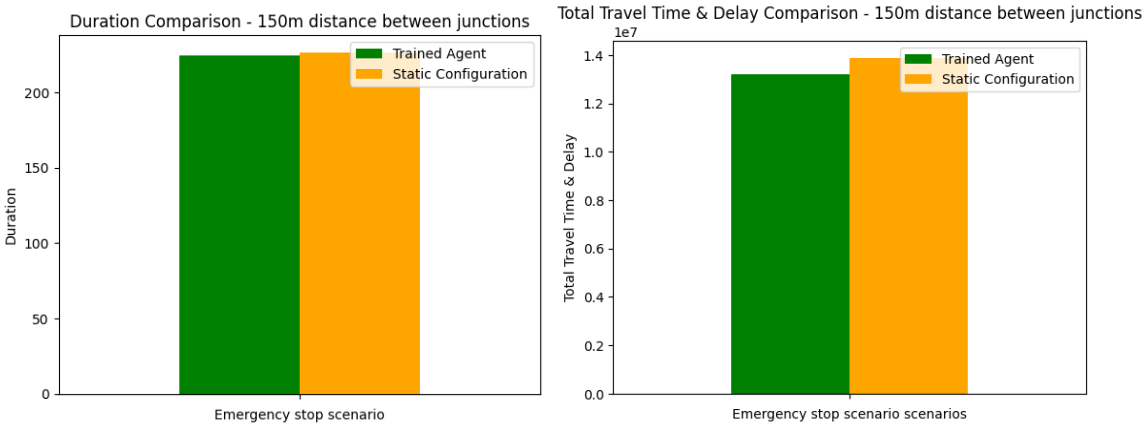
Διάγραμμα 23: Συγκριτικά διαγράμματα διάρκειας διαδρομών και συνολικής διάρκειας ταξιδιού και καθυστέρησης - Μεγάλες εκδηλώσεις- Μικρό δίκτυο



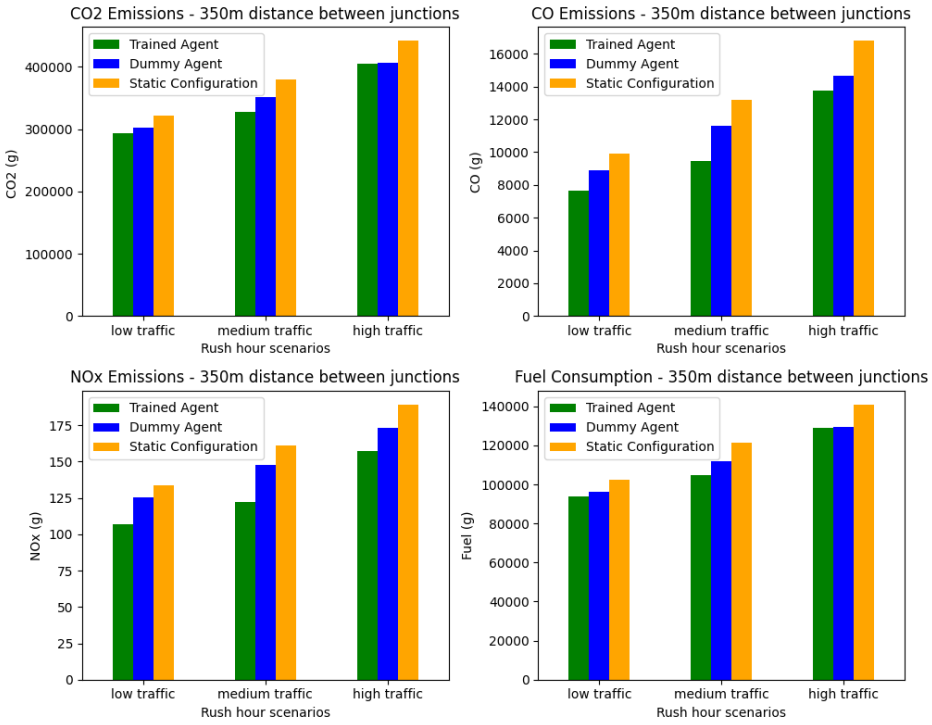
Διάγραμμα 24: Συγκριτικό διάγραμμα ταχυτήτων και διάρκειας διαδρομών -Αναγκαστική στάση- Μεγάλο δίκτυο



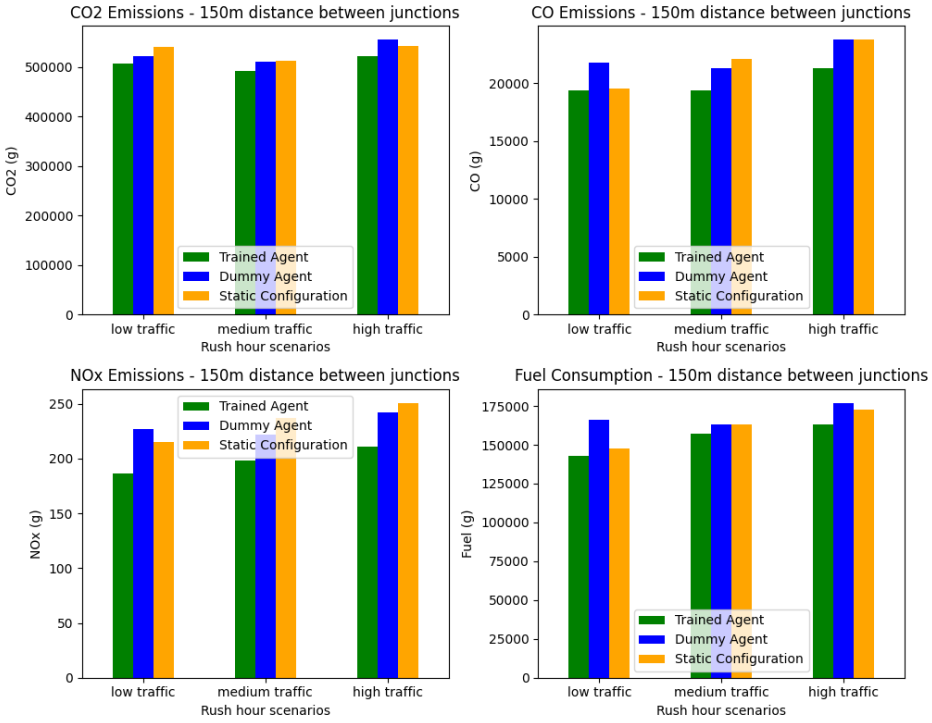
Διάγραμμα 25: Συγκριτικό διάγραμμα συνολικής διάρκειας ταξιδιού και καθυστέρησης (Μεγάλο δίκτυο) και ταχυτήτων (Μικρό δίκτυο) - Αναγκαστική στάση



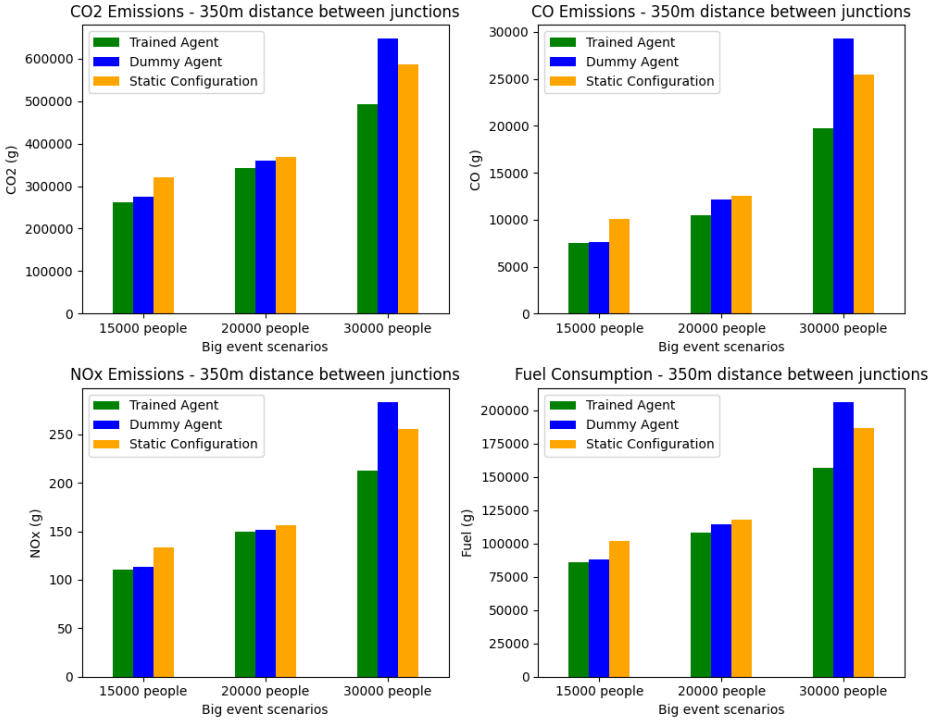
Διάγραμμα 26: Συγκριτικό διάγραμμα διάρκειας διαδρομών και συνολικής διάρκειας ταξιδιού και καθυστέρησης – Αναγκαστική στάση- Μικρό δίκτυο



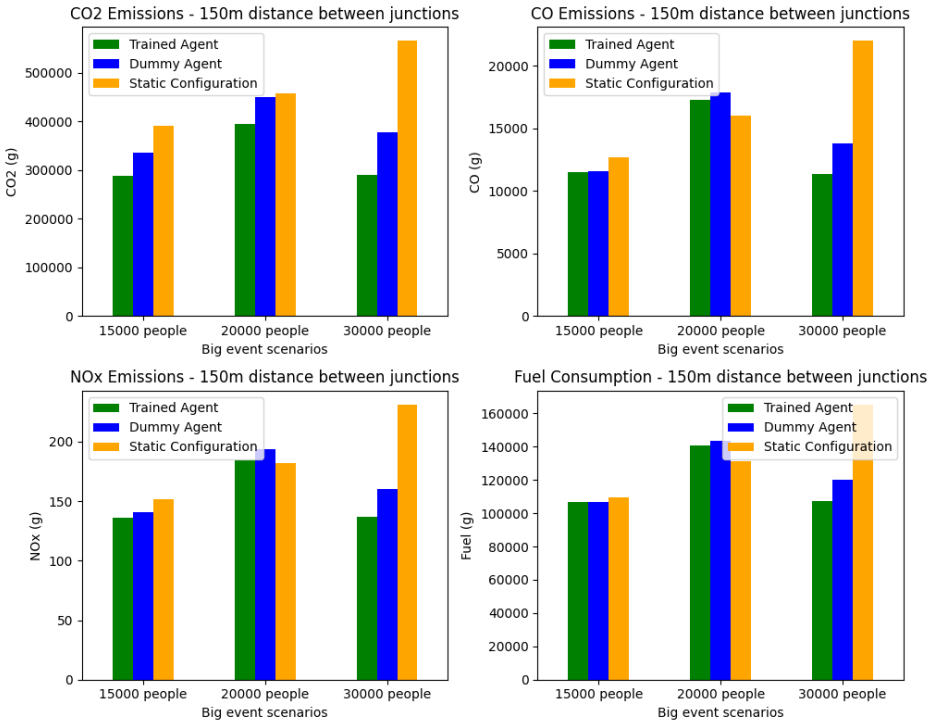
Διάγραμμα 27: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων-Ώρες αιχμής - Μεγάλο δίκτυο



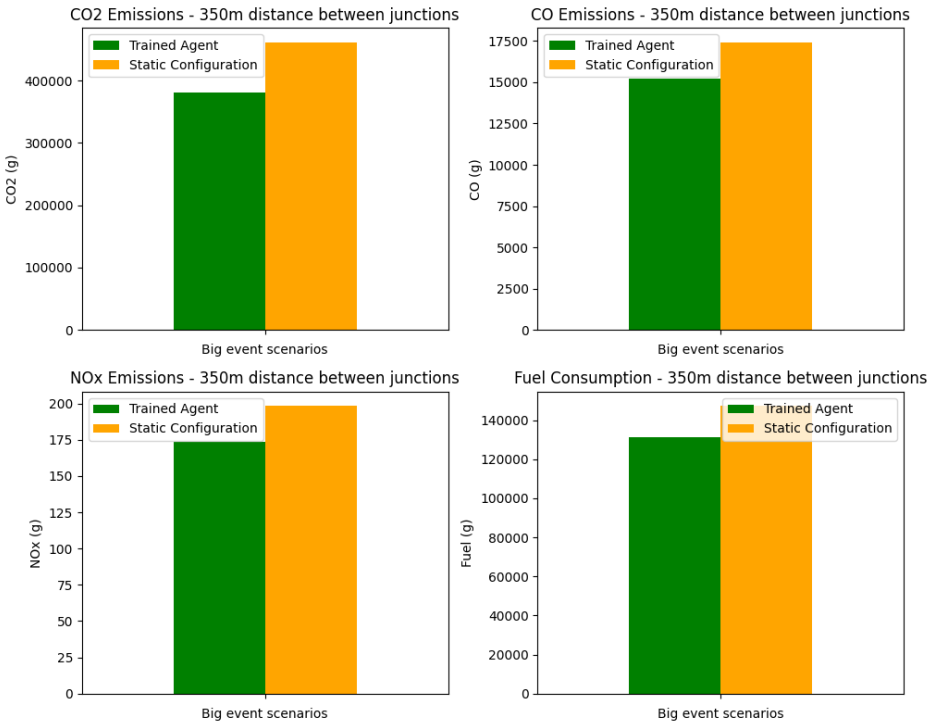
Διάγραμμα 28: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων-Ώρες αιχμής-Μικρό δίκτυο



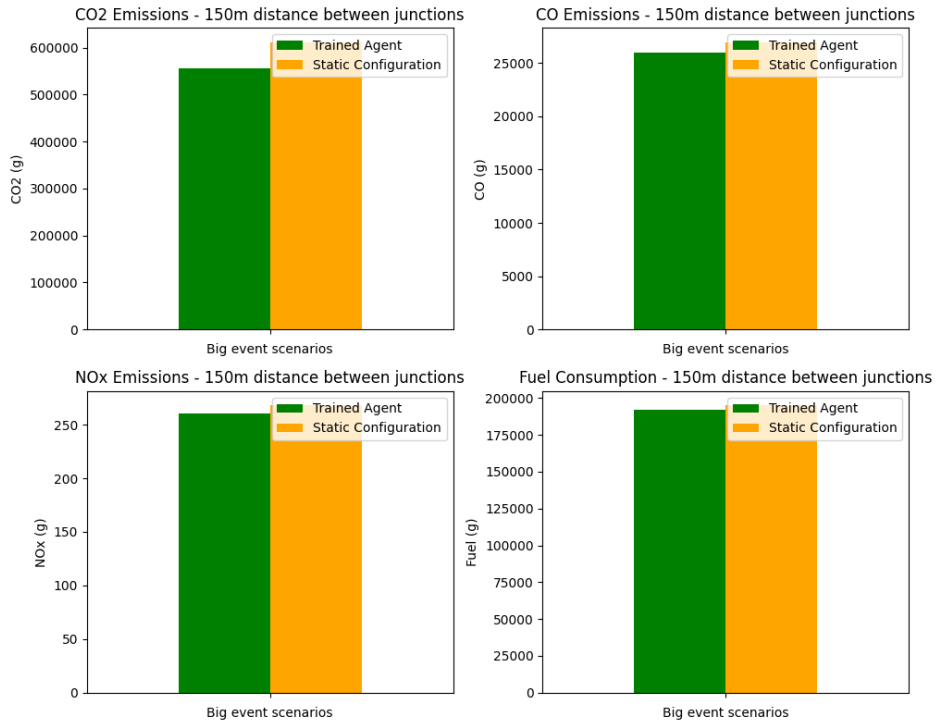
Διάγραμμα 29: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων-Μεγάλες εκδηλώσεις-Μεγάλο δίκτυο



Διάγραμμα 30: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων-Μεγάλες εκδηλώσεις-Μικρό δίκτυο



Διάγραμμα 31: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων- Αναγκαστική στάση-Μεγάλο δίκτυο



Διάγραμμα 32: Συγκριτικά διαγράμματα εκπομπής αερίων και καυσίμων- Αναγκαστική στάση- Μικρό δίκτυο