

ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΕΛΛΑΔΟΣ

Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Ανάπτυξη εφαρμογής ιστού συλλογής και αναζήτησης
δρομολογίων ΚΤΕΛ με βελτιστοποίηση διαδρομών βάσει
κριτηρίων χρήστη**

Φοιτητής:

Μαυρίδης Παναγιώτης
Αριθμός Μητρώου: 2021027

Επιβλέπων:

Ουγιάρογλου Στέφανος

Θεσσαλονίκη, Αύγουστος 2026

Ανάπτυξη εφαρμογής ιστού συλλογής και αναζήτησης δρομολογίων ΚΤΕΛ με βελτιστοποίηση διαδρομών βάσει κριτηρίων χρήστη

Κωδικός Δ.Ε.: 26177

Ονοματεπώνυμο φοιτητή: Μαυρίδης Παναγιώτης

Ονοματεπώνυμο εισηγητή: Ουγιάρογλου Στέφανος

Ημερομηνία ανάληψης: 12-03-2026

Ημερομηνία περάτωσης: _____

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή Παναγιώτη Μαυρίδη που την εκπόνησε. Ο συγγραφέας/δημιουργός διατηρεί στο ακέραιο και κατ' αποκλειστικότητα το σύνολο των δικαιωμάτων επί του έργου. Δεν εκχωρείται σε τρίτους, συμπεριλαμβανομένων του Διεθνούς Πανεπιστημίου της Ελλάδος και του επιβλέποντος καθηγητή, άδεια αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό ή ψηφιακής διάχυσης της εργασίας, σε ηλεκτρονική ή οποιαδήποτε άλλη μορφή. Απαγορεύεται η αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση και τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά, της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

Περίληψη

Οι υπεραστικές οδικές συγκοινωνίες της Ελλάδας οργανώνονται σε περισσότερα από εξήντα αυτοτελή ΚΤΕΛ νομαρχιακής εμβέλειας, καθένα από τα οποία δημοσιεύει το χρονοδιάγραμμά του ανεξάρτητα και σε μορφή που απευθύνεται αποκλειστικά σε ανθρώπινο αναγνώστη. Δεν υπάρχει ενιαία, μηχανικά αναγνώσιμη πηγή, με αποτέλεσμα καμία μετακίνηση που διασχίζει περισσότερες από μία περιοχές ευθύνης να μην μπορεί να αναζητηθεί αυτόματα. Η παρούσα εργασία αντιμετωπίζει το κενό αυτό σε ολόκληρη την αλυσίδα, από τη συλλογή των δεδομένων έως την παρουσίασή τους στον τελικό χρήστη.

Εντοπίστηκαν τέσσερις τεχνολογικές υποδομές πίσω από τους ιστότοπους των φορέων και οι μη τεκμηριωμένες υπηρεσίες τους ανακτήθηκαν με αντίστροφη μηχανίκευση, η οποία περιέλαβε και την αποκρυπτογράφηση των αποκρίσεων μίας εξ αυτών. Η συλλογή υλοποιήθηκε με ασύγχρονους και πολυνηματικούς μηχανισμούς, με έλεγχο ταυτοχρονισμού και πολιτική επαναλήψεων. Η ταύτιση των στάσεων επιλύθηκε με γεωκωδικοποίηση, χωρική ομαδοποίηση και ημιαυτόματη επικύρωση μέσω διαδραστικού χάρτη, αντιστοιχίζοντας 1.157 επιμέρους καταχωρίσεις σε 353 κανονικές οντότητες. Το τελικό σύνολο περιλαμβάνει 106.380 δρομολόγια από 35 ΚΤΕΛ και 2.310 κόμβους, φορτωμένα σε σχεσιακή βάση με εγγυημένη ταυτοδυναμία.

Επί του συνόλου αυτού αναπτύχθηκε ο ΜΙΤΟΣ, σύστημα δρομολόγησης που υιοθετεί το χρονοεξαρτώμενο μοντέλο, εκφράζει τον χρόνο σε απόλυτα λεπτά ώστε να υποστηρίζει ορίζοντα πολλαπλών ημερών, και διαχειρίζεται ρητά τις ανταποκρίσεις και τις διανυκτερεύσεις. Κεντρικό στοιχείο της υλοποίησης αποτελεί ένα στάδιο οπισθοδρομικής βελτιστοποίησης που ελαχιστοποιεί τους χρόνους αναμονής χωρίς να μεταβάλλει τον χρόνο τελικής άφιξης. Το σύστημα διατίθεται ως διαδικτυακή εφαρμογή με διαδραστικό χάρτη και ως διεπαφή γραμμής εντολών. Η αποτίμηση έδειξε μέσο χρόνο απόκρισης 274 χιλιοστών του δευτερολέπτου και τήρηση των αναλλοίωτων ορθότητας σε 3.533 ελεγχθέντα σκέλη, με μόνη εξαίρεση την υπέρβαση του ορίου αναμονής σε 0,3% αυτών — σπάνιο και εντοπισμένο ελάττωμα της μετα-βελτιστοποίησης, η οποία βελτίωσε το 15,6% των διαδρομών. Το 92,2% των λύσεων συνθέτει δρομολόγια περισσότερων του ενός φορέα — απαντήσεις, δηλαδή, που καμία υφιστάμενη λύση δεν μπορεί εξ ορισμού να δώσει.

Abstract

Greece's intercity road transport is organised around more than sixty autonomous regional operators (KTEL), each publishing its timetable independently and in a form intended solely for human readers. No unified, machine-readable source exists, which means that no journey crossing more than one operator's area of responsibility can be queried automatically. This thesis addresses that gap across the entire chain, from data acquisition to end-user presentation.

Four technological platforms were identified behind the operators' websites, and their undocumented services were accessed through reverse engineering, which included decrypting the responses of one of them. Collection was implemented with asynchronous and multithreaded mechanisms featuring concurrency control and a retry policy. Stop entity resolution was solved through geocoding, spatial clustering and semi-automatic validation via an interactive map, mapping 1,157 individual entries onto 353 canonical entities. The resulting dataset comprises 106,380 services from 35 operators across 2,310 nodes, loaded into a relational database with guaranteed idempotency.

On this foundation MITOS was developed: a routing system adopting the time-dependent model, expressing time in absolute minutes so as to support a multi-day horizon, and handling transfers and overnight stays explicitly. A central element of the implementation is a backward optimisation stage that minimises waiting times without altering the final arrival time. The system is delivered both as a web application with an interactive map and as a command-line interface. Evaluation showed a mean response time of 274 milliseconds and compliance with the correctness invariants across 3,533 examined legs, the sole exception being the waiting-time limit, exceeded in 0.3% of them — a rare and localised defect of the post-optimisation stage, which improved 15.6% of journeys. Some 92.2% of the solutions combine services of more than one operator — answers, that is, which no existing solution can by definition provide.

Περιεχόμενα

Εισαγωγή	1
1.1 Σχεδιασμός ταξιδιού με μέσα συγκοινωνιών	1
1.2 Τρόπος λειτουργίας υπεραστικών συγκοινωνιών στην Ελλάδα	3
1.3 Το πρόβλημα: κατακερματισμός δεδομένων και απουσία ενιαίου προτύπου	5
1.4 Κίνητρο	6
1.5 Στόχοι της εργασίας	7
1.6 Συνεισφορά.....	8
1.7 Οργάνωση της εργασίας.....	10
Θεωρητικό υπόβαθρο	11
2.1 Θεωρία γράφων: βασικοί ορισμοί	11
2.1.1 Γράφοι, κατευθυνόμενοι γράφοι, βάρη ακμών.....	11
2.1.2 Μονοπάτια, κύκλοι, συνεκτικότητα.....	12
2.1.3 Αναπαράσταση γράφων και υπολογιστικές επιπτώσεις.....	12
2.2 Το πρόβλημα συντομότερης διαδρομής	13
2.2.1 Ορισμός και παραλλαγές.....	13
2.2.2 Η αρχή της βελτιστότητας και η υποδομή βέλτιστης λύσης	13
2.2.3 Ο αλγόριθμος Dijkstra: περιγραφή και ψευδοκώδικας	14
2.2.4 Απόδειξη ορθότητας	14
2.2.5 Ανάλυση πολυπλοκότητας	15
2.2.6 Προϋποθέσεις εφαρμογής και όρια του κλασικού Dijkstra	15
2.3 Χρονοεξαρτώμενα δίκτυα	16
2.3.1 Η έννοια της χρονοεξαρτώμενης συνάρτησης κόστους.....	16
2.3.2 Η ιδιότητα FIFO (non-overtaking) και η σημασία της για την ορθότητα.....	16
2.3.3 Επέκταση του Dijkstra σε χρονοεξαρτώμενα δίκτυα.....	17
2.3.4 Απόδειξη ορθότητας υπό την υπόθεση FIFO	17
2.4 Μοντελοποίηση δικτύων συγκοινωνιών βάσει χρονοδιαγράμματος	18
2.4.1 Ιδιαιτερότητες: διακριτές αναχωρήσεις, περιοδικότητα, ανταποκρίσεις	18
2.4.2 Το time-expanded μοντέλο.....	18
2.4.3 Το time-dependent μοντέλο.....	19
2.4.4 Συγκριτική αποτίμηση των δύο μοντέλων	19
2.5 Το πρόβλημα πρωιμότερης άφιξης (Earliest Arrival Problem).....	20

2.5.1 Τυπικός ορισμός	20
2.5.2 Περιορισμοί ελάχιστου χρόνου ανταπόκρισης.....	20
2.5.3 Επέκταση σε ορίζοντα πολλαπλών ημερών	21
2.5.4 Η έννοια του απόλυτου χρόνου ως ενοποιητική αναπαράσταση	21
2.6 Πολυκριτηριακή δρομολόγηση	21
2.6.1 Κυριαρχία και βελτιστότητα κατά Pareto	21
2.6.2 Τυπικά αντικρουόμενα κριτήρια.....	22
2.6.3 Λεξικογραφική διάταξη ως πρακτική προσέγγιση	22
2.6.4 Ελαχιστοποίηση αναμονής και αναζήτηση τελευταίας αναχώρησης	22
2.7 Αλγόριθμοι δρομολόγησης σε δίκτυα συγκοινωνιών	23
2.7.1 Connection Scan Algorithm (CSA)	23
2.7.2 RAPTOR	23
2.7.3 Transfer Patterns	24
2.7.4 Συγκριτική αξιολόγηση και κριτήρια επιλογής	24
2.8 Ενοποίηση ετερογενών πηγών δεδομένων.....	25
2.8.1 Το πρόβλημα ταύτισης οντοτήτων (entity resolution)	25
2.8.2 Μέτρα ομοιότητας και χωρική ομαδοποίηση.....	25
2.8.3 Αμφισημία τοπωνυμίων και ο ρόλος του γεωγραφικού συμφραζομένου.....	26
2.9 Ο προτεινόμενος αλγόριθμος	26
2.9.1 Τυπική διατύπωση του προβλήματος της εργασίας.....	26
2.9.2 Το επιλεγμένο μοντέλο και η αιτιολόγησή του	26
2.9.3 Περιγραφή και ψευδοκώδικας.....	27
2.9.4 Το στάδιο οπισθοδρομικής βελτιστοποίησης.....	27
2.9.5 Ορθότητα, τερματισμός και πολυπλοκότητα	28
2.9.6 Παραδοχές και θεωρητικοί περιορισμοί	29
2.10 Σχετικές εργασίες.....	29
Επίλογος κεφαλαίου	30
Γλώσσες και Τεχνολογίες	32
3.1 Python και το οικοσύστημα βιβλιοθηκών	32
3.2 Τεχνολογίες ανάκτησης δεδομένων: requests, BeautifulSoup, aiohttp / asyncio.....	33
3.3 SQLite και σχεσιακή αποθήκευση.....	34
3.4 Σειριοποίηση δεδομένων: JSON και pickle	35
3.5 Flask και αρχιτεκτονική REST.....	35
3.6 Τεχνολογίες frontend: HTML, CSS, JavaScript.....	36

3.7 Γεωχωρικές υπηρεσίες: OpenStreetMap, Nominatim, OSRM, Leaflet.....	37
3.8 Το πρότυπο GTFS.....	38
3.9 Κρυπτογραφία στην επικοινωνία με APIs (AES-CBC, pycryptodome)	38
3.10 Εργαλεία ανάπτυξης και ελέγχου εκδόσεων	39
Επίλογος κεφαλαίου	39
Ανάκτηση δεδομένων και δημιουργία σχεσιακής Βάσης Δεδομένων	41
4.1 Καταγραφή και ταξινόμηση των πηγών δεδομένων.....	41
4.1.1 Πλατφόρμα IME / ktelbus.gr	42
4.1.2 Πλατφόρμα e-analysis / e-ticketing.gr	42
4.1.3 Πλατφόρμα 3pi-telematics	42
4.1.4 Αυτόνομα συστήματα (ΚΤΕΛ Ηλείας).....	43
4.2 Μεθοδολογία αντίστροφης μηχανίκευσης των διαδικτυακών υπηρεσιών	43
4.2.1 Ανάλυση κίνησης δικτύου και εντοπισμός endpoints.....	43
4.2.2 Αποκρυπτογράφηση των αποκρίσεων του IME API.....	44
4.3 Σχεδίαση και υλοποίηση των scrapers.....	45
4.3.1 Ασύγχρονη συλλογή με asyncio και έλεγχος ταυτοχρονισμού	45
4.3.2 Πολυνηματική συλλογή με ThreadPoolExecutor	45
4.3.3 Πολιτική επαναλήψεων, rate limiting και καταγραφή σφαλμάτων	46
4.4 Στρατηγική δειγματοληψίας χρόνου και εξαγωγή εβδομαδιαίου προτύπου	46
4.5 Ενδιάμεση μορφή αποθήκευσης και σχήμα των JSON αρχείων	47
4.6 Ενοποίηση οντοτήτων στάσεων	48
4.6.1 Το πρόβλημα των συνωνύμων και των ομωνύμων	48
4.6.2 Γεωκωδικοποίηση στάσεων μέσω Nominatim.....	49
4.6.3 Χωρική ομαδοποίηση γειτονικών στάσεων	49
4.6.4 Διαδραστικό εργαλείο χάρτη για ημιαυτόματη επικύρωση	50
4.6.5 Το αρχείο canonical stops και η context-aware αντιστοίχιση	50
4.7 Συμπλήρωση ελλিপών χρόνων διαδρομής με OSRM	51
4.8 Σχεδίαση της σχεσιακής βάσης δεδομένων	51
4.8.1 Σχήμα, ευρετήρια και περιορισμοί.....	51
4.8.2 Διαδικασία φόρτωσης και ιδιότητα idempotency	52
4.9 Στατιστικά και ποιοτική αποτίμηση του τελικού συνόλου δεδομένων	53
Επίλογος κεφαλαίου	54
Ανάπτυξη του MITOY	55
5.1 Λειτουργικές απαιτήσεις	55

5.2 Μη λειτουργικές απαιτήσεις	56
5.3 Case studies	57
5.4 Αρχιτεκτονική συστήματος.....	57
5.4.1 Επισκόπηση και διάγραμμα ροής δεδομένων.....	57
5.4.2 Διαχωρισμός επιπέδων	58
5.4.3 Προϋπολογισμός γράφου και μηχανισμός cache.....	59
5.5 Ανάπτυξη του backend	60
5.5.1 Κατασκευή του γράφου δρομολογίων από τη βάση	60
5.5.2 Αναπαράσταση του χρόνου σε απόλυτα λεπτά και υποστήριξη πολλαπλών ημερών..	60
5.5.3 Υλοποίηση του αλγορίθμου αναζήτησης διαδρομών	61
5.5.4 Διαχείριση ανταποκρίσεων και διανυκτερεύσεων.....	61
5.5.5 Μετα-βελτιστοποίηση διαδρομής: ελαχιστοποίηση χρόνων αναμονής.....	62
5.5.6 Κριτήρια κατάταξης και απαλοιφή διπλοτύπων	63
5.5.7 Ασαφής αναζήτηση και κανονικοποίηση ελληνικού κειμένου	63
5.5.8 Σχεδίαση του REST API	64
5.5.9 Διεπαφή γραμμής εντολών (CLI)	64
5.6 Ανάπτυξη του frontend	65
5.6.1 Σχεδίαση της διεπαφής χρήστη.....	65
5.6.2 Αναζήτηση στάσεων με αυτόματη συμπλήρωση	66
5.6.3 Οπτικοποίηση διαδρομών και διαδραστικός χάρτης	68
5.6.4 Παραμετροποίηση αναζήτησης από τον χρήστη	68
5.6.5 Αποκρίσιμος σχεδιασμός.....	69
Επίλογος κεφαλαίου	69
Παρουσίαση του MITOY	70
6.1 Εγκατάσταση και εκτέλεση.....	70
6.2 Σενάριο 1: άμεση σύνδεση.....	70
6.3 Σενάριο 2: διαδρομή με μία ή περισσότερες ανταποκρίσεις.....	71
6.4 Σενάριο 3: διαδρομή με διανυκτέρευση.....	72
6.5 Σενάριο 4: εβδομαδιαία σύνοψη διαθεσιμότητας.....	72
6.6 Χρήση μέσω γραμμής εντολών	73
6.7 Αξιολόγηση απόδοσης.....	73
6.8 Αξιολόγηση ορθότητας έναντι πραγματικών δρομολογίων.....	74
6.9 Σύγκριση με υφιστάμενες λύσεις	75
Επίλογος κεφαλαίου	76

Συμπεράσματα και Μελλοντικές επεκτάσεις.....	77
7.1 Σύνοψη της εργασίας	77
7.2 Συμπεράσματα.....	78
7.2.1 Εκπλήρωση των στόχων.....	78
7.2.2 Τεχνικά και ερευνητικά συμπεράσματα.....	78
7.2.3 Η κατάσταση των ανοικτών συγκοινωνιακών δεδομένων στην Ελλάδα.....	79
7.3 Περιορισμοί.....	80
7.3.1 Περιορισμοί που πηγάζουν από τις πηγές δεδομένων	80
7.3.2 Παραδοχές του μοντέλου	80
7.3.3 Εγκυρότητα και χρονική ισχύς των δεδομένων.....	80
7.4 Μελλοντικές επεκτάσεις	81
7.4.1 Πολυτροπική δρομολόγηση.....	81
7.4.2 Διασυννοριακή επέκταση και ενοποίηση με ευρωπαϊκά δίκτυα.....	81
7.4.3 Δρομολόγηση υπό αβεβαιότητα	81
7.4.4 Ενσωμάτωση δεδομένων πραγματικού χρόνου και δυναμική επαναδρομολόγηση.....	82
7.4.5 Εξατομικευμένη και επεξηγήσιμη δρομολόγηση	82
7.4.6 Από τον σχεδιασμό στον ανασχεδιασμό του δικτύου: ανάλυση προσβασιμότητας ως εργαλείο πολιτικής.....	82
7.4.7 Προς ένα εθνικό πρότυπο ανοικτών συγκοινωνιακών δεδομένων	82
Επίλογος κεφαλαίου	83
Βιβλιογραφία.....	84

Κατάλογος Σχημάτων

Σχήμα 4.1: Το σχήμα της βάσης δεδομένων: ενιαίος, μη κανονικοποιημένος πίνακας δρομολογίων με τα τρία ευρετήρια που υποστηρίζουν τα ερωτήματα του συστήματος.	52
Σχήμα 5.1: Διάγραμμα ροής δεδομένων του ΜΙΤΟΥ. Η ροή είναι αυστηρά μονόδρομη και κάθε ενδιάμεσο προϊόν αποτελεί αυτοτελές αρχείο.	58
Σχήμα 5.2: Η αρχική οθόνη της διαδικτυακής εφαρμογής. Διακρίνονται η φόρμα αναζήτησης με τα τέσσερα πάντοτε ορατά πεδία, το πλήκτρο εναλλαγής αφετηρίας και προορισμού, η πτυσσόμενη ενότητα των προχωρημένων επιλογών και, κάτω από αυτήν, ο διαδραστικός χάρτης με τα συσσωματώματα των στάσεων. Η περιοχή αποτελεσμάτων παραμένει κρυφή έως την πρώτη αναζήτηση.	66
Σχήμα 5.3: Αυτόματη συμπλήρωση στο πεδίο της αφετηρίας. Οι υποδείξεις εμφανίζονται μετά τον δεύτερο χαρακτήρα και καλύπτουν τόσο την κανονική οντότητα όσο και τις επιμέρους καταχωρίσεις που περιέχουν τον ίδιο όρο.	67
Σχήμα 5.4: Αυτόματη συμπλήρωση στο πεδίο του προορισμού, με την αφετηρία ήδη επιλεγμένη. Ο νομός που συνοδεύει κάθε υπόδειξη προέρχεται από τη γεωκωδικοποίηση της ενότητας 4.6 και επιτρέπει στον χρήστη να διακρίνει ομώνυμες ή παραπλήσιες στάσεις πριν από την επιλογή.....	67
Σχήμα 6.1: Η διαδρομή Θεσσαλονίκη προς ΚΤΕΛ Σπάρτης όπως αποδίδεται στη διεπαφή, με τη γραμμή της Τρίτης αναπτυγμένη σε πλήρη κάρτα· η λύση είναι πανομοιότυπη με εκείνη της Δευτέρας που περιγράφεται στο κείμενο. Διακρίνονται τα δύο σκέλη, ο φορέας και η πλατφόρμα προέλευσης καθενός, η ενδιάμεση αναμονή των τριάντα εννέα λεπτών στην Αθήνα και οι συγκεντρωτικοί χρόνοι της διαδρομής.	71
Σχήμα 6.2: Η εβδομαδιαία σύνοψη διαθεσιμότητας σε συνεπτυγμένη μορφή, εδώ για τη διαδρομή της ενότητας 6.3: μία γραμμή ανά ημέρα με την ένδειξη διαθεσιμότητας και τη βέλτιστη διαδρομή της ημέρας, η οποία αναπτύσσεται κατ' απαίτηση σε πλήρη κάρτα.	73

Κατάλογος Πινάκων

Πίνακας 4.1: Σύνθεση του τελικού συνόλου δεδομένων ανά πηγή (στιγμιότυπο Ιουνίου 2026)..	53
Πίνακας 4.2: Κατανομή των δρομολογίων ως προς την εβδομαδιαία εκτέλεση.	54
Πίνακας 6.1: Χρόνοι απόκρισης της αναζήτησης σε δείγμα 600 τυχαίων ζευγών αφετηρίας και προορισμού. Μέγιστη παρατηρηθείσα τιμή σε ολόκληρο το δείγμα: 795 ms.....	74
Πίνακας 6.2: Αποτελέσματα της αυτοματοποιημένης επαλήθευσης αναλλοίωτων επί 400 τυχαίων ερωτημάτων, τα οποία παρήγαγαν 1.144 διαδρομές και 3.533 σκέλη.....	75
Πίνακας 6.3: Αποτίμηση της μετα-βελτιστοποίησης σε δείγμα 1.123 διαδρομών δύο ή περισσότερων σκελών.	75
Πίνακας 6.4: Διαθεσιμότητα λύσεων ανά ημέρα εκκίνησης — τα δύο άκρα του παρατηρηθέντος εύρους.	76

Κεφάλαιο 1

Εισαγωγή

Η δυνατότητα ενός πολίτη να μετακινηθεί από ένα σημείο της χώρας σε ένα άλλο με τα μέσα μαζικής μεταφοράς προϋποθέτει, πριν από κάθε ζήτημα κόστους ή υποδομής, κάτι πολύ απλούστερο: να γνωρίζει ότι η μετακίνηση αυτή είναι εφικτή, με ποια διαδοχή δρομολογίων πραγματοποιείται και σε ποιον χρόνο ολοκληρώνεται. Η πληροφορία αυτή, που στα περισσότερα ευρωπαϊκά δίκτυα θεωρείται δεδομένη και προσφέρεται από αυτοματοποιημένους σχεδιαστές ταξιδιού, παραμένει στην Ελλάδα διάσπαρτη, ετερογενής και σε μεγάλο βαθμό μη μηχανικά επεξεργάσιμη για το υπεραστικό οδικό δίκτυο. Η παρούσα εργασία εξετάζει το κενό αυτό και προτείνει μια ολοκληρωμένη απάντηση, από τη συλλογή των δεδομένων έως την αλγοριθμική επίλυση του προβλήματος και την παρουσίασή του σε τελικό χρήστη. Το παρόν κεφάλαιο οριοθετεί το αντικείμενο: περιγράφει τι σημαίνει σχεδιασμός ταξιδιού με μέσα συγκοινωνιών, πώς λειτουργεί το ελληνικό υπεραστικό σύστημα, ποιο ακριβώς είναι το πρόβλημα που αντιμετωπίζεται, τι κινητοποίησε την ενασχόληση με αυτό, ποιοι στόχοι τέθηκαν, τι προσφέρει η εργασία και πώς οργανώνεται η συνέχεια της.

1.1 Σχεδιασμός ταξιδιού με μέσα συγκοινωνιών

Ο σχεδιασμός ταξιδιού με μέσα μαζικής μεταφοράς αφορά την αυτόματη εύρεση μιας ακολουθίας δρομολογίων που μεταφέρει έναν επιβάτη από μια αφετηρία σε έναν προορισμό, ικανοποιώντας ταυτόχρονα τους περιορισμούς που επιβάλλει το χρονοδιάγραμμα των μέσων. Ένα σύστημα που επιτελεί αυτή τη λειτουργία ονομάζεται σχεδιαστής ταξιδιού (journey planner) και αποτελεί σήμερα τυπικό συστατικό της καθημερινής μετακίνησης: υπηρεσίες όπως το Google Maps, το Rome2Rio, το Transit ή οι εθνικοί σχεδιαστές που λειτουργούν οι διαχειριστές δικτύων σε πολλές ευρωπαϊκές χώρες απαντούν σε ερωτήματα της μορφής «πώς πηγαίνω από το Α στο Β αύριο το πρωί» σε κλάσματα του δευτερολέπτου, ενσωματώνοντας δεκάδες φορείς εκμετάλλευσης και εκατοντάδες χιλιάδες αναχωρήσεις.

Παρότι η λειτουργία μοιάζει εκ πρώτης όψεως συγγενής με την πλοήγηση σε οδικό δίκτυο, τα δύο προβλήματα διαφέρουν ουσιωδώς. Στην οδική πλοήγηση το δίκτυο είναι συνεχώς διαθέσιμο: κάθε ακμή μπορεί να διασχιστεί οποιαδήποτε στιγμή και το κόστος της είναι, σε πρώτη προσέγγιση, σταθερό. Αντίθετα, σε ένα δίκτυο συγκοινωνιών η μετάβαση από μια στάση σε μια άλλη είναι

δυνατή μόνο σε συγκεκριμένες, διακριτές χρονικές στιγμές — τις προγραμματισμένες αναχωρήσεις. Το κόστος μιας μετάβασης δεν εξαρτάται μόνο από το ζεύγος των στάσεων αλλά και από τη χρονική στιγμή άφιξης του επιβάτη στην πρώτη από αυτές, διότι καθορίζει πόσο θα αναμείνει μέχρι την επόμενη διαθέσιμη αναχώρηση. Το πρόβλημα, δηλαδή, είναι εγγενώς χρονοεξαρτώμενο.

Η χρονική αυτή διάσταση εισάγει επιπλέον περιορισμούς που δεν έχουν αντίστοιχο στην οδική δρομολόγηση. Η μετεπιβίβαση απαιτεί έναν ελάχιστο χρόνο ανταπόκρισης, ώστε ο επιβάτης να προλάβει να μεταβεί από την αποβίβαση στην επόμενη επιβίβαση· η καθυστέρηση ή η ματαίωση ενός σκέλους ακυρώνει ολόκληρη τη συνέχεια της διαδρομής· ορισμένα δρομολόγια εκτελούνται μόνο συγκεκριμένες ημέρες της εβδομάδας ή περιόδους του έτους· και όταν καμία σύνδεση δεν είναι διαθέσιμη εντός της ίδιας ημέρας, η βέλτιστη λύση μπορεί να περιλαμβάνει διανυκτέρευση σε ενδιάμεσο σταθμό. Η αναζήτηση επομένως δεν διεξάγεται σε έναν στατικό γράφο αλλά σε μια δομή όπου ο χρόνος αποτελεί ισότιμη συνιστώσα της κατάστασης.

Ένα δεύτερο χαρακτηριστικό του προβλήματος είναι ότι η έννοια της «καλύτερης» διαδρομής δεν είναι μονοσήμαντη. Ο επιβάτης δεν ενδιαφέρεται μόνο για την ώρα άφιξης· τον αφορούν εξίσου το πλήθος των μετεπιβίβσεων, ο συνολικός χρόνος αναμονής, η ώρα αναχώρησης, το κόστος και η αξιοπιστία των ανταποκρίσεων. Τα κριτήρια αυτά είναι συχνά αντικρουόμενα: μια διαδρομή με μία επιπλέον μετεπιβίβαση μπορεί να φτάνει νωρίτερα, ενώ μια απευθείας σύνδεση μπορεί να απαιτεί πρωινή αναχώρηση. Η θεωρητική απάντηση δίνεται μέσω της βελτιστότητας κατά Pareto, όμως στην πράξη τα συστήματα καταφεύγουν σε λεξικογραφικές διατάξεις ή σε σταθμισμένες συναρτήσεις κόστους, ώστε να παρουσιάσουν στον χρήστη ένα μικρό, ερμηνεύσιμο σύνολο εναλλακτικών.

Πέρα από την αλγοριθμική του διάσταση, ο σχεδιασμός ταξιδιού είναι πρωτίστως πρόβλημα δεδομένων. Ένας σχεδιαστής είναι τόσο καλός όσο το χρονοδιάγραμμα που τροφοδοτείται: απαιτεί πλήρη κατάλογο στάσεων με γεωγραφικές συντεταγμένες και σταθερά αναγνωριστικά, ακριβείς ώρες αναχώρησης και άφιξης, κανόνες εκτέλεσης ανά ημέρα, καθώς και συνεπή αντιστοίχιση των στάσεων μεταξύ διαφορετικών φορέων. Η ανάγκη αυτή οδήγησε στη διαμόρφωση προτύπων ανταλλαγής, με επικρατέστερο το GTFS (General Transit Feed Specification), το οποίο περιγράφει σε μια τυποποιημένη συλλογή αρχείων τις γραμμές, τις στάσεις, τα δρομολόγια και το ημερολόγιο εκτέλεσής τους. Η υιοθέτηση ενός κοινού προτύπου είναι ακριβώς εκείνο που επιτρέπει σε τρίτους να αναπτύξουν εφαρμογές χωρίς να διαπραγματευτούν χωριστά με κάθε φορέα.

Στη σύγχρονη πρακτική διακρίνονται δύο επίπεδα πληροφορίας. Το στατικό επίπεδο περιγράφει το προγραμματισμένο χρονοδιάγραμμα, δηλαδή τι προβλέπεται να εκτελεστεί, και μεταβάλλεται σπάνια· το δυναμικό επίπεδο αποτυπώνει την πραγματική εκτέλεση σε πραγματικό χρόνο, με καθυστερήσεις, ματαιώσεις και έκτακτες τροποποιήσεις. Οι ώριμες υπηρεσίες συνδυάζουν και τα δύο, προσφέροντας δρομολόγηση που προσαρμόζεται στις τρέχουσες συνθήκες. Το στατικό επίπεδο, ωστόσο, παραμένει η αναγκαία βάση: χωρίς αξιόπιστο προγραμματισμένο χρονοδιάγραμμα δεν υπάρχει τίποτε επί του οποίου να εφαρμοστεί η πληροφορία πραγματικού χρόνου. Παράλληλα, η απόκριση οφείλει να είναι διαδραστική, γεγονός που εξηγεί την εντατική έρευνα των τελευταίων δύο δεκαετιών σε εξειδικευμένους αλγορίθμους και τεχνικές προεπεξεργασίας.

Η δημοσίευση τέτοιων δεδομένων δεν είναι απλώς τεχνική επιλογή αλλά και ζήτημα πολιτικής. Σε επίπεδο Ευρωπαϊκής Ένωσης, ο κατ' εξουσιοδότηση Κανονισμός (ΕΕ) 2017/1926 υποχρεώνει τα κράτη μέλη να καθιστούν διαθέσιμα τα στατικά δεδομένα μετακίνησης μέσω εθνικών σημείων πρόσβασης, ώστε να καταστεί δυνατή η ανάπτυξη πολυτροπικών υπηρεσιών πληροφόρησης [1]. Όπου η υποχρέωση αυτή έχει υλοποιηθεί ουσιαστικά, ο σχεδιασμός ταξιδιού έχει πάψει να είναι ερευνητικό ζητούμενο και έχει μετατραπεί σε υποδομή. Όπως θα φανεί στη συνέχεια, το ελληνικό υπεραστικό δίκτυο δεν βρίσκεται ακόμη σε αυτό το σημείο, και ακριβώς αυτή η απόσταση ορίζει το πεδίο της παρούσας εργασίας.

1.2 Τρόπος λειτουργίας υπεραστικών συγκοινωνιών στην Ελλάδα

Οι υπεραστικές οδικές συγκοινωνίες στην Ελλάδα εκτελούνται κατά κύριο λόγο από τα ΚΤΕΛ, τα Κοινά Ταμεία Εισπράξεων Λεωφορείων. Πρόκειται για έναν θεσμό με μακρά ιστορία, ο οποίος προέκυψε από τη συνένωση μεμονωμένων ιδιοκτητών λεωφορείων σε συλλογικά σχήματα διαχείρισης εσόδων και δρομολογίων. Με τον νόμο 2963/2001 τα σχήματα αυτά μετασχηματίστηκαν σε ανώνυμες εταιρείες, αποκτώντας σύγχρονη εταιρική μορφή χωρίς ωστόσο να μεταβληθεί η θεμελιώδης λογική οργάνωσής τους [2]. Η λογική αυτή είναι γεωγραφική: η χώρα διαιρείται σε περιοχές ευθύνης που ακολουθούν τα όρια των παλαιών νομών, και σε κάθε τέτοια περιοχή αντιστοιχεί, κατά κανόνα, ένα ΚΤΕΛ με αποκλειστικό δικαίωμα εκμετάλλευσης των υπεραστικών γραμμών που την αφορούν.

Η σημασία του θεσμού για τη χώρα είναι δυσανάλογα μεγάλη σε σύγκριση με άλλα ευρωπαϊκά κράτη, και τούτο για λόγους γεωγραφίας και ιστορικών επιλογών υποδομής. Το σιδηροδρομικό δίκτυο καλύπτει περιορισμένο τμήμα της επικράτειας, οργανωμένο κυρίως γύρω από τον άξονα Αθήνας–Θεσσαλονίκης και ελάχιστες παρακλαδώσεις, ενώ ολόκληρες περιφέρειες, όπως η Ήπειρος και η Πελοπόννησος στο μεγαλύτερο μέρος της, δεν εξυπηρετούνται σιδηροδρομικά. Οι αεροπορικές συνδέσεις είναι κατά κανόνα ακτινωτές προς τα δύο μεγάλα αστικά κέντρα και δεν εξυπηρετούν εγκάρσιες μετακινήσεις. Για τη συντριπτική πλειονότητα των διανομαρχιακών διαδρομών της ηπειρωτικής χώρας, επομένως, το υπεραστικό λεωφορείο δεν είναι απλώς μία επιλογή μεταξύ άλλων· είναι η μοναδική επιλογή δημόσιας μεταφοράς.

Το αποτέλεσμα είναι ένα δίκτυο που δεν σχεδιάστηκε ποτέ ως ενιαίο σύστημα, αλλά προέκυψε ως άθροισμα εξήντα και πλέον αυτοτελών, νομαρχιακής εμβέλειας δικτύων [2]. Κάθε εταιρεία καταρτίζει το δικό της χρονοδιάγραμμα, το οποίο εγκρίνεται από τις αρμόδιες περιφερειακές αρχές, καθορίζει τις στάσεις των γραμμών της και εκμεταλλεύεται τον δικό της κεντρικό σταθμό υπεραστικών λεωφορείων. Οι κομίστρα υπολογίζονται με βάση χιλιομετρικούς συντελεστές που ορίζονται κεντρικά από το αρμόδιο υπουργείο, γεγονός που εξασφαλίζει ομοιομορφία στην τιμολόγηση αλλά όχι στην πληροφόρηση [3]. Η εποπτεία, δηλαδή, είναι συγκεντρωτική ως προς το τιμολόγιο και αποκεντρωμένη ως προς το περιεχόμενο και τη δημοσιοποίηση των δρομολογίων.

Η γεωγραφική κατάτμηση καθορίζει και τον τρόπο με τον οποίο εκτελούνται οι διανομαρχιακές συνδέσεις. Μια γραμμή που συνδέει δύο πρωτεύουσες νομών εκτελείται συνήθως από το ΚΤΕΛ του νομού αφετηρίας, ενώ η αντίστροφη κατεύθυνση μπορεί να εκτελείται από το ΚΤΕΛ του άλλου νομού ή από το ίδιο, ανάλογα με τις μεταξύ τους συμφωνίες. Πρακτικά αυτό σημαίνει ότι

το ίδιο φυσικό ταξίδι εμφανίζεται σε δύο διαφορετικά χρονοδιαγράμματα, ενδεχομένως με διαφορετική ονομασία των ίδιων στάσεων. Για τον επιβάτη η διάκριση αυτή είναι αδιάφορη· για οποιονδήποτε επιχειρεί να συγκροτήσει μια συνολική εικόνα του δικτύου, όμως, αποτελεί την πρώτη πηγή διπλοεγγραφών και ασυνεπειών.

Η δομή αυτή εμφανίζει τα όριά της κυρίως εκεί όπου η μετακίνηση διασχίζει περισσότερες από μία περιοχές ευθύνης. Δεν υφίσταται ενιαίο εισιτήριο για διαδρομές που απαιτούν αλλαγή λεωφορείου μεταξύ δύο ΚΤΕΛ, ούτε θεσμοθετημένος συντονισμός των ωραρίων στα σημεία όπου συναντώνται τα επιμέρους δίκτυα. Οι ανταποκρίσεις σε κόμβους όπως η Θεσσαλονίκη, η Λαμία, η Πάτρα ή τα Ιωάννινα δεν είναι σχεδιασμένες αλλά συμπτωματικές: προκύπτουν από την τυχαία σύμπτωση ενός δρομολογίου άφιξης με ένα δρομολόγιο αναχώρησης. Ο επιβάτης καλείται να τις εντοπίσει μόνος του, συνδυάζοντας πληροφορίες από δύο ή περισσότερους ανεξάρτητους φορείς, και αναλαμβάνει ο ίδιος τον κίνδυνο μιας χαμένης σύνδεσης.

Στην πολυπλοκότητα αυτή προστίθεται η χρονική μεταβλητότητα των χρονοδιαγραμμάτων. Τα δρομολόγια δεν είναι σταθερά καθ' όλη τη διάρκεια του έτους: διακρίνονται συνήθως σε θερινά και χειμερινά προγράμματα, ενώ σε πολλές γραμμές η συχνότητα μεταβάλλεται δραστικά κατά τη σχολική περίοδο, τις αργίες ή την τουριστική αιχμή. Ορισμένες συνδέσεις εκτελούνται μόνο δύο ή τρεις ημέρες την εβδομάδα, εξυπηρετώντας ημιορεινούς οικισμούς σε συνδυασμό με τις ημέρες λαϊκής αγοράς ή διοικητικών υπηρεσιών. Η πληροφορία για το ποιες ακριβώς ημέρες εκτελείται ένα δρομολόγιο είναι, συνεπώς, εξίσου κρίσιμη με την ίδια την ώρα αναχώρησης, και η απώλειά της καθιστά το χρονοδιάγραμμα πρακτικά άχρηστο για σχεδιασμό.

Σε επίπεδο πληροφόρησης, η εικόνα είναι εξίσου κατακερματισμένη. Ορισμένα ΚΤΕΛ διαθέτουν σύγχρονους ιστότοπους με δυνατότητα ηλεκτρονικής κράτησης, άλλα δημοσιεύουν τα δρομολογία τους ως στατικές σελίδες ή αρχεία PDF που ενημερώνονται σπάνια, και ορισμένα εξακολουθούν να παραπέμπουν σε τηλεφωνική επικοινωνία με τον σταθμό. Τα τελευταία χρόνια έχει σημειωθεί σημαντική πρόοδος μέσω της υιοθέτησης κοινών πλατφορμών ηλεκτρονικής έκδοσης εισιτηρίων και τηλεματικής, οι οποίες εξυπηρετούν πολλαπλά ΚΤΕΛ με κοινή τεχνική υποδομή. Η εργασία αυτή εντόπισε τέσσερις τέτοιες τεχνολογικές οικογένειες: την πλατφόρμα IME (ktelbus.gr), την πλατφόρμα e-analysis (e-ticketing.gr), την πλατφόρμα 3pi-telematics και, τέλος, αυτόνομα συστήματα που έχουν αναπτυχθεί κατά παραγγελία μεμονωμένων εταιρειών, όπως στην περίπτωση του ΚΤΕΛ Ηλείας.

Η σύγκλιση αυτή είναι πραγματική αλλά μερική, και κυρίως δεν είναι σύγκλιση δεδομένων. Οι πλατφόρμες μοιράζονται τεχνική υποδομή, όχι κοινό μοντέλο πληροφορίας: κάθε ΚΤΕΛ παραμένει κύριος των δικών του δεδομένων, τα οποία δημοσιεύονται αποκλειστικά μέσω της διεπαφής της αντίστοιχης πλατφόρμας και προορίζονται για ανθρώπινη ανάγνωση ή για την εξυπηρέτηση μιας μεμονωμένης συναλλαγής κράτησης. Δεν προβλέπεται δημόσια διεπαφή προγραμματισμού εφαρμογών, δεν διατίθεται εξαγωγή σε τυποποιημένη μορφή, και οι εσωτερικές υπηρεσίες που τροφοδοτούν τις σελίδες δεν τεκμηριώνονται· σε μία τουλάχιστον περίπτωση οι αποκρίσεις τους μεταδίδονται κρυπτογραφημένες.

Σε αντιδιαστολή, οι αστικές συγκοινωνίες των δύο μεγάλων αστικών κέντρων έχουν προχωρήσει σημαντικά περισσότερο, με τον ΟΑΣΑ να διαθέτει δεδομένα σε μορφή GTFS και να είναι, ως εκ τούτου, ορατός στους διεθνείς σχεδιαστές ταξιδιού [4]. Το υπεραστικό δίκτυο, που είναι ακριβώς εκείνο το οποίο εξυπηρετεί τις μετακινήσεις μεγάλων αποστάσεων και συνδέει την περιφέρεια με

τα αστικά κέντρα, παραμένει εκτός. Η ανομοιογένεια αυτή δεν οφείλεται σε έλλειψη δεδομένων — τα δεδομένα υπάρχουν, είναι ψηφιακά και ενημερώνονται καθημερινά για εμπορικούς σκοπούς — αλλά στην απουσία ενός θεσμικού και τεχνικού πλαισίου που θα επέβαλλε τη δημοσίευσή τους σε κοινή, μηχανικά αναγνώσιμη μορφή.

1.3 Το πρόβλημα: κατακερματισμός δεδομένων και απουσία ενιαίου προτύπου

Από τα προηγούμενα προκύπτει ένα συγκεκριμένο και πρακτικά διατυπώσιμο πρόβλημα: για μια μετακίνηση που απαιτεί συνδυασμό δρομολογίων περισσότερων του ενός ΚΤΕΛ, δεν υπάρχει σήμερα κανένα σύστημα — ούτε δημόσιο ούτε εμπορικό — που να μπορεί να απαντήσει στο ερώτημα «πώς φτάνω από το Α στο Β». Η αδυναμία αυτή δεν οφείλεται στην πολυπλοκότητα του υπολογισμού, ο οποίος είναι θεωρητικά επιλυμένος εδώ και δεκαετίες, αλλά στο ότι δεν υφίσταται ενιαία αναπαράσταση του δικτύου πάνω στην οποία θα μπορούσε να εκτελεστεί ένας τέτοιος υπολογισμός.

Η πρώτη διάσταση του προβλήματος είναι ο κατακερματισμός των πηγών. Τα δεδομένα κατανέμονται σε τέσσερις τεχνολογικά ανεξάρτητες πλατφόρμες, καθεμία με διαφορετική δομή αποκρίσεων, διαφορετική κωδικοποίηση των ημερών εκτέλεσης, διαφορετική ακρίβεια ως προς τους χρόνους διαδρομής και διαφορετικό βαθμό πληρότητας. Ορισμένες πηγές παρέχουν ώρα αναχώρησης και ώρα άφιξης, άλλες μόνο ώρα αναχώρησης· ορισμένες δηλώνουν ρητά τις ημέρες εβδομαδιαίας εκτέλεσης, άλλες τις υπονοούν μέσω της διαθεσιμότητας κράτησης σε συγκεκριμένη ημερομηνία. Η άντληση των δεδομένων απαιτεί, επομένως, χωριστή μελέτη και χωριστή υλοποίηση για κάθε πλατφόρμα.

Η δεύτερη και δυσκολότερη διάσταση αφορά την ταυτότητα των στάσεων. Δεν υπάρχει εθνικό μητρώο σημείων εξυπηρέτησης, ούτε κοινό αναγνωριστικό που να αντιστοιχίζει την ίδια φυσική στάση όπως αυτή εμφανίζεται σε δύο διαφορετικά ΚΤΕΛ. Οι ονομασίες καταχωρίζονται ελεύθερα, με αποτέλεσμα το ίδιο σημείο να εμφανίζεται με πολλαπλές γραφές — με ή χωρίς τόνους, σε κεφαλαία ή πεζά, με ή χωρίς προσδιοριστικά όπως «ΚΤΕΛ», «ΣΤΑΘΜΟΣ» ή το όνομα του νομού. Ταυτόχρονα, και αντίστροφα, το ίδιο ακριβώς όνομα αντιστοιχεί συχνά σε εντελώς διαφορετικά μέρη: τοπωνύμια όπως «Νεοχώρι», «Άγιος Γεώργιος» ή «Καστανιά» απαντούν σε πολλούς νομούς. Η αφελής ενοποίηση βάσει ονόματος θα δημιουργούσε ανύπαρκτες συνδέσεις μεταξύ απομακρυσμένων περιοχών και θα παρήγαγε διαδρομές που δεν είναι απλώς μη βέλτιστες αλλά αδύνατες.

Η τρίτη διάσταση είναι η απουσία πληροφορίας για τις ανταποκρίσεις. Ακόμη και αν ο επιβάτης εντοπίσει και τα δύο σκέλη μιας μετακίνησης, καμία πηγή δεν του δηλώνει αν το πρώτο προλαβαίνει το δεύτερο. Δεν υπάρχει καταγεγραμμένος ελάχιστος χρόνος μετεπιβίβασης, δεν υπάρχει εγγύηση συνέχειας, και σε πολλές περιπτώσεις τα δύο σκέλη εξυπηρετούνται από διαφορετικούς σταθμούς της ίδιας πόλης. Ο υπολογισμός της εφικτότητας μιας σύνδεσης μετατοπίζεται εξ ολοκλήρου στον χρήστη.

Μια τέταρτη διάσταση, λιγότερο εμφανής αλλά εξίσου καθοριστική, αφορά τη μορφή με την οποία δημοσιεύεται η πληροφορία. Τα δεδομένα προορίζονται αποκλειστικά για ανθρώπινη ανάγνωση μέσα από μια συγκεκριμένη διεπαφή και όχι για επαναχρησιμοποίηση. Δεν υπάρχει τεκμηριωμένη

διεπαφή προγραμματισμού, ούτε εξαγωγή σε τυποποιημένη μορφή όπως το GTFS, ούτε δήλωση αδειοδότησης που να ορίζει τους όρους χρήσης. Οι εσωτερικές υπηρεσίες που τροφοδοτούν τις σελίδες αλλάζουν χωρίς προειδοποίηση, καθώς ουδέποτε υπήρξε δέσμευση σταθερότητας απέναντι σε τρίτους. Το αποτέλεσμα είναι ότι κάθε προσπάθεια αξιοποίησης των δεδομένων είναι εξ ορισμού εύθραυστη και απαιτεί διαρκή συντήρηση, ενώ η οποιαδήποτε αναδιαμόρφωση ενός ιστότοπου μπορεί να ακυρώσει το σύνολο της υποδομής συλλογής.

Στα παραπάνω προστίθεται η απουσία ρητής χρονικής ισχύος. Τα δημοσιευμένα χρονοδιαγράμματα δεν συνοδεύονται από ημερομηνία έκδοσης ή περίοδο εγκυρότητας, ούτε τηρείται ιστορικό εκδόσεων. Μια αλλαγή δρομολογίου αντικαθιστά σιωπηρά την προηγούμενη κατάσταση, χωρίς να είναι δυνατόν να διαπιστωθεί πότε επήλθε ή τι ίσχυε πριν. Για τον χρήστη αυτό σημαίνει ότι δεν μπορεί να γνωρίζει αν η πληροφορία που βλέπει είναι επίκαιρη· για τον ερευνητή, ότι κάθε σύνολο δεδομένων αποτελεί στιγμιότυπο συγκεκριμένης χρονικής στιγμής, με περιορισμένη δυνατότητα διαχρονικής σύγκρισης.

Οι δυσκολίες αυτές συνθέτουν, τέλος, ένα πρόβλημα κυκλικής εξάρτησης, το οποίο αποτελεί και τον πυρήνα του ζητήματος. Για να αναζητήσει κανείς δρομολόγια στους ιστότοπους των ΚΤΕΛ, πρέπει να γνωρίζει ήδη ποιο ΚΤΕΛ να ρωτήσει· και για να το γνωρίζει, πρέπει να έχει ήδη υποθέσει μέσω ποιου ενδιάμεσου κόμβου θα περάσει η διαδρομή του. Ακριβώς όμως αυτή η ενδιάμεση πόλη είναι το ζητούμενο. Η υπάρχουσα υποδομή πληροφόρησης απαιτεί από τον χρήστη να γνωρίζει εκ των προτέρων την απάντηση στο ερώτημα που θέτει, και είναι σε θέση να επιβεβαιώσει μια υπόθεση αλλά όχι να παραγάγει μια λύση.

Το πρακτικό αποτέλεσμα είναι μια σιωπηρή υποβάθμιση της προσβασιμότητας. Διαδρομές που είναι απολύτως εφικτές παραμένουν άγνωστες, επειδή κανείς δεν έχει συνθέσει την πληροφορία που τις καθιστά ορατές. Ο επιβάτης που δεν γνωρίζει το δίκτυο καταλήγει είτε να απορρίψει τη μετακίνηση με μέσα μαζικής μεταφοράς ως ανέφικτη είτε να καταφύγει σε ιδιωτικό όχημα. Το πρόβλημα, δηλαδή, δεν είναι ότι το δίκτυο δεν συνδέει τη χώρα· είναι ότι δεν υπάρχει τρόπος να διαπιστωθεί ότι τη συνδέει.

1.4 Κίνητρο

Το κίνητρο της εργασίας προέκυψε από τη διαπίστωση μιας δυσαναλογίας: το πρόβλημα που περιγράφηκε δεν είναι ούτε θεωρητικά ανοικτό ούτε τεχνολογικά απαιτητικό, και ωστόσο παραμένει άλυτο. Οι αλγόριθμοι δρομολόγησης σε δίκτυα χρονοδιαγράμματος αποτελούν ώριμο και εκτενώς μελετημένο πεδίο, τα υπολογιστικά μεγέθη του ελληνικού υπεραστικού δικτύου είναι μικρά με τα σημερινά δεδομένα, και η πληροφορία υπάρχει ήδη σε ψηφιακή μορφή και ενημερώνεται καθημερινά. Το εμπόδιο δεν είναι επιστημονικό αλλά οργανωτικό, και αυτό ακριβώς καθιστά ενδιαφέρον το ερώτημα κατά πόσο μπορεί να παρακαμφθεί με τεχνικά μέσα.

Ταυτόχρονα, το πρόβλημα δεν είναι μόνο τεχνικό ως προς τις συνέπειές του. Η δυνατότητα σχεδιασμού μιας μετακίνησης με μέσα μαζικής μεταφοράς συνιστά προϋπόθεση πρόσβασης σε υπηρεσίες υγείας, εκπαίδευσης και εργασίας, ιδίως για πληθυσμιακές ομάδες που δεν διαθέτουν ιδιωτικό όχημα: φοιτητές, ηλικιωμένοι κάτοικοι της περιφέρειας, νοικοκυριά χαμηλού εισοδήματος. Σε μια χώρα με έντονη γεωγραφική διασπορά και συνεχιζόμενη συρρίκνωση του αγροτικού πληθυσμού, η αοριστία γύρω από το τι είναι εφικτό λειτουργεί ως πρόσθετο εμπόδιο.

Η αδυναμία, επιπλέον, του επισκέπτη να σχεδιάσει μετακινήσεις εκτός των κύριων τουριστικών αξόνων περιορίζει και την οικονομική διάχυση προς τις λιγότερο προβλεβλημένες περιοχές.

Υπάρχει και μια αμιγώς αλγοριθμική πλευρά που κινητοποίησε την ενασχόληση. Το ελληνικό υπεραστικό δίκτυο παρουσιάζει χαρακτηριστικά που το καθιστούν ενδιαφέρουσα περίπτωση μελέτης και όχι απλή εφαρμογή γνωστών τεχνικών: η συχνότητα των δρομολογίων είναι χαμηλή, ώστε ο χρόνος αναμονής να κυριαρχεί συχνά έναντι του χρόνου διαδρομής· οι συνδέσεις μεταξύ απομακρυσμένων περιοχών ενδέχεται να μην ολοκληρώνονται εντός μίας ημέρας, επιβάλλοντας τη ρητή μοντελοποίηση της διανυκτέρευσης· και η εκτέλεση των δρομολογίων μεταβάλλεται ανά ημέρα της εβδομάδας. Τα στοιχεία αυτά διαφοροποιούν το πρόβλημα από τα πυκνά αστικά δίκτυα στα οποία έχει επικεντρωθεί το μεγαλύτερο μέρος της σχετικής βιβλιογραφίας, όπου η υψηλή συχνότητα καθιστά τον χρόνο αναμονής αμελητέο και το ενδιαφέρον μετατοπίζεται στην ταχύτητα απόκρισης για εκατομμύρια ερωτήματα. Εδώ, αντιθέτως, το πλήθος των ερωτημάτων είναι μικρό αλλά η ποιότητα κάθε απάντησης έχει δυσανάλογα μεγάλη σημασία για τον χρήστη.

Η αφορμή, εξάλλου, υπήρξε συγκεκριμένη και εμπειρική. Η προσπάθεια σχεδιασμού μιας μετακίνησης μεταξύ δύο πόλεων που δεν συνδέονται απευθείας ανέδειξε με σαφήνεια το αδιέξοδο που περιγράφηκε: πολλαπλές τηλεφωνικές κλήσεις σε διαφορετικούς σταθμούς, αντικρουόμενες πληροφορίες, και τελικά η διαπίστωση ότι η επιλογή του ενδιάμεσου κόμβου έγινε με κριτήριο την εικασία και όχι τον υπολογισμό. Η εμπειρία αυτή, κοινή σε όποιον έχει επιχειρήσει να μετακινηθεί εκτός των κύριων αξόνων, οδήγησε στο ερώτημα αν η απάντηση θα μπορούσε να παραχθεί αυτόματα — ερώτημα που, μόλις διατυπωθεί με ακρίβεια, αποδεικνύεται ταυτόσημο με ένα κλασικό πρόβλημα δρομολόγησης σε γράφο.

Τέλος, η εργασία εκκινεί από την πεποίθηση ότι η ίδια η κατασκευή ενός λειτουργικού συστήματος αποτελεί μορφή τεκμηρίωσης. Η επίδειξη ότι είναι εφικτό να συγκροτηθεί ένα ενοποιημένο σύνολο δεδομένων και ένας λειτουργικός σχεδιαστής ταξιδιού για το σύνολο του δικτύου, με μέσα διαθέσιμα σε έναν προπτυχιακό φοιτητή, καθιστά σαφές ότι η απουσία μιας τέτοιας υπηρεσίας δεν είναι αναπόφευκτη. Το πρωτότυπο που παρουσιάζεται εδώ δεν φιλοδοξεί να υποκαταστήσει μια θεσμική λύση· επιδιώκει να δείξει με συγκεκριμένο τρόπο τι θα ήταν διαθέσιμο αν τα ήδη υπάρχοντα δεδομένα δημοσιεύονταν σε ενιαία μορφή.

1.5 Στόχοι της εργασίας

Γενικός στόχος της εργασίας είναι η σχεδίαση και υλοποίηση ενός ολοκληρωμένου συστήματος αναζήτησης διαδρομών για το σύνολο του ελληνικού υπεραστικού δικτύου ΚΤΕΛ, από την πρωτογενή συλλογή των δεδομένων έως την παρουσίασή τους σε τελικό χρήστη. Ζητούμενο δεν είναι μόνο η λειτουργικότητα του τελικού προϊόντος, αλλά και η τεκμηρίωση κάθε σταδίου με τρόπο που να επιτρέπει την αναπαραγωγή και την επέκτασή του. Ο γενικός αυτός στόχος αναλύεται στους ακόλουθους επιμέρους στόχους.

Σ1. Συστηματική ανάκτηση των δεδομένων. Καταγραφή του συνόλου των διαδικτυακών πηγών που δημοσιεύουν χρονοδιαγράμματα υπεραστικών δρομολογίων, ανάλυση των μη τεκμηριωμένων διεπαφών τους και ανάπτυξη μηχανισμών αυτοματοποιημένης συλλογής για κάθε τεχνολογική οικογένεια.

Σ2. Ενοποίηση σε ενιαίο μοντέλο. Επίλυση του προβλήματος ταύτισης οντοτήτων για τις στάσεις, με αντιμετώπιση τόσο των συνωνύμων όσο και των ομωνύμων, και κατασκευή μιας σχεσιακής βάσης δεδομένων που αποτελεί ενιαία, συνεπή αναπαράσταση του δικτύου.

Σ3. Τυπική διατύπωση του προβλήματος δρομολόγησης. Μοντελοποίηση του δικτύου ως χρονοεξαρτώμενου γράφου και διατύπωση του προβλήματος πρωιμότερης άφιξης με περιορισμούς ελάχιστου χρόνου ανταπόκρισης, εβδομαδιαίας εκτέλεσης και ορίζοντα πολλαπλών ημερών.

Σ4. Ανάπτυξη και τεκμηρίωση του αλγορίθμου. Υλοποίηση αλγορίθμου αναζήτησης βασισμένου στην αρχή του Dijkstra για χρονοεξαρτώμενα δίκτυα, με τεκμηρίωση της ορθότητας και της πολυπλοκότητάς του, καθώς και μηχανισμού μετα-βελτιστοποίησης για την ελαχιστοποίηση των χρόνων αναμονής.

Σ5. Υλοποίηση προσβάσιμης διεπαφής. Ανάπτυξη διαδικτυακής εφαρμογής με διαδραστικό χάρτη, αναζήτηση με αυτόματη συμπλήρωση και παραμετροποίηση των κριτηρίων, καθώς και ισοδύναμης διεπαφής γραμμής εντολών.

Σ6. Αξιολόγηση. Αποτίμηση της απόδοσης και της ορθότητας των αποτελεσμάτων έναντι πραγματικών δρομολογίων και σύγκριση με τις υφιστάμενες διαθέσιμες λύσεις.

Παράλληλα με τους στόχους, κρίνεται σκόπιμο να οριοθετηθεί ρητά και το πεδίο της εργασίας. Η μελέτη περιορίζεται στις υπεραστικές οδικές συγκοινωνίες που εκτελούνται από τα ΚΤΕΛ και δεν καλύπτει τις αστικές συγκοινωνίες, τον σιδηρόδρομο, την ακτοπλοΐα ή τις αεροπορικές συνδέσεις, παρότι η αρχιτεκτονική που προτείνεται σχεδιάστηκε ώστε να επιδέχεται μελλοντική επέκταση προς την κατεύθυνση αυτή. Η δρομολόγηση βασίζεται αποκλειστικά στο προγραμματισμένο χρονοδιάγραμμα και δεν ενσωματώνει δεδομένα πραγματικού χρόνου, καθώς τέτοια δεδομένα δεν δημοσιεύονται. Τέλος, το κόστος του εισιτηρίου δεν αποτελεί κριτήριο βελτιστοποίησης, αφενός επειδή δεν είναι διαθέσιμο ομοιόμορφα σε όλες τις πηγές και αφετέρου επειδή, ελλείψει ενιαίου κομίστρου, δεν είναι συγκρίσιμο μεταξύ διαφορετικών φορέων.

1.6 Συνεισφορά

Η εργασία συνεισφέρει σε τέσσερα επίπεδα: στα δεδομένα, στη μεθοδολογία, στον αλγόριθμο και στην τεκμηρίωση της υφιστάμενης κατάστασης.

Ενοποιημένο σύνολο δεδομένων. Συγκροτήθηκε, εξ όσων γνωρίζουμε για πρώτη φορά, ένα ενιαίο μηχανικά αναγνώσιμο σύνολο δεδομένων για τις ελληνικές υπεραστικές συγκοινωνίες. Το τελικό σύνολο περιλαμβάνει 106.380 εγγραφές δρομολογίων από 35 ΚΤΕΛ και τέσσερις ανεξάρτητες τεχνολογικές πλατφόρμες, οι οποίες αντιστοιχίστηκαν σε 353 κανονικοποιημένες οντότητες στάσεων προερχόμενες από 1.157 επιμέρους καταχωρίσεις, με πλήρη γεωγραφική πληροφορία και κωδικοποίηση των ημερών εβδομαδιαίας εκτέλεσης. Το σύνολο αυτό καλύπτει το μεγαλύτερο μέρος της ηπειρωτικής χώρας και αποτελεί, στην παρούσα μορφή του, την πληρέστερη ενιαία αποτύπωση του υπεραστικού δικτύου που έχει συγκροτηθεί εκτός των ίδιων των φορέων εκμετάλλευσης.

Μεθοδολογία ανάκτησης και ενοποίησης. Τεκμηριώνεται μια πλήρης μεθοδολογία αντίστροφης μηχανίκευσης μη τεκμηριωμένων διαδικτυακών υπηρεσιών, η οποία περιλαμβάνει τον εντοπισμό

των εσωτερικών σημείων πρόσβασης μέσω ανάλυσης της κίνησης δικτύου και την αποκρυπτογράφηση των αποκρίσεων μίας εξ αυτών. Παράλληλα προτείνεται μια διαδικασία ενοποίησης οντοτήτων που συνδυάζει λεξιλογική ομοιότητα, γεωκωδικοποίηση μέσω Nominatim, χωρική ομαδοποίηση γειτονικών σημείων και ημιαυτόματη επικύρωση μέσω διαδραστικού εργαλείου χάρτη. Κρίσιμο στοιχείο της είναι η αντιστοίχιση με επίγνωση του συμφραζομένου, η οποία αίρει την αμφισημία των ομώνυμων τοπωνυμίων αξιοποιώντας το ΚΤΕΛ και την πηγή προέλευσης κάθε καταχώρισης.

Αλγοριθμική συνεισφορά. Διατυπώνεται και υλοποιείται μια παραλλαγή αναζήτησης πρωιμότερης άφιξης σε χρονοεξαρτώμενο γράφο, η οποία εκφράζει τον χρόνο σε απόλυτα λεπτά από τη στιγμή εκκίνησης και υποστηρίζει έτσι ομοιόμορφα διαδρομές που εκτείνονται σε περισσότερες από μία ημερολογιακές ημέρες, συμπεριλαμβανομένων των αναγκαστικών διανυκτερεύσεων. Κεντρικό στοιχείο της υλοποίησης είναι ένα στάδιο οπισθοδρομικής βελτιστοποίησης, το οποίο μετατοπίζει κάθε σκέλος της διαδρομής όσο το δυνατόν αργότερα διατηρώντας αμετάβλητη την ώρα τελικής άφιξης, εξαλείφοντας τις περιττές αναμονές που ο κλασικός αλγόριθμος πρωιμότερης άφιξης παράγει σε αραιά δίκτυα. Η αρχή στην οποία στηρίζεται το στάδιο αυτό δεν είναι νέα: ανάγεται στο γνωστό πρόβλημα της τελευταίας αναχώρησης και στη σχέση κυριαρχίας των ερωτημάτων εύρους [9, 10, 13]. Η συνεισφορά της εργασίας έγκειται στην προσαρμογή της σε ελαφρύ στάδιο μετα-επεξεργασίας μίας ήδη υπολογισμένης διαδρομής, στη ρητή εξαίρεση των διανυκτερεύσεων και στην ποσοτική αποτίμηση του οφέλους της σε πραγματικά δεδομένα αραιού υπεραστικού δικτύου. Η βελτίωση αυτή δεν μεταβάλλει τη βελτιστότητα της λύσης ως προς την ώρα άφιξης, αλλά αναβαθμίζει ουσιαστικά την ποιότητά της από τη σκοπιά του επιβάτη, ο οποίος δεν έχει κανένα όφελος από το να αναχωρήσει τρεις ώρες νωρίτερα προκειμένου να προλάβει την ίδια ακριβώς ανταπόκριση. Παρουσιάζεται επίσης η αντιμετώπιση των αναγκαστικών διανυκτερεύσεων ως ρητά αναγνωρίσιμου συστατικού της διαδρομής, συνοδευόμενου από τις εναλλακτικές αναχωρήσεις της επόμενης ημέρας.

Λειτουργικό σύστημα. Παραδίδεται πλήρως λειτουργικό σύστημα, όχι απλή προσομοίωση ή μερικό πρωτότυπο, το οποίο απαντά σε ερωτήματα δρομολόγησης για ολόκληρη τη χώρα σε διαδραστικό χρόνο. Το σύστημα ονομάστηκε ΜΙΤΟΣ, κατά τον μίτο της Αριάδνης: το δίκτυο υπάρχει και συνδέει τη χώρα, εκείνο που λείπει είναι το νήμα που επιτρέπει τη διέξοδο από τον λαβύρινθο των εξήντα αυτοτελών χρονοδιαγραμμάτων. Περιλαμβάνει διαδικτυακή εφαρμογή με διαδραστικό χάρτη, αναζήτηση στάσεων με ανοχή σε ορθογραφικές αποκλίσεις και τονισμό, παραμετροποίηση των κριτηρίων αναζήτησης από τον χρήστη, εβδομαδιαία σύνοψη διαθεσιμότητας, καθώς και ισοδύναμη διεπαφή γραμμής εντολών για χρήση σε σεσάκια αυτοματοποίησης.

Τεκμηρίωση της υφιστάμενης κατάστασης. Τέλος, η εργασία αποτυπώνει εμπειρικά και με μετρήσιμο τρόπο την κατάσταση των ανοικτών συγκοινωνιακών δεδομένων στην Ελλάδα: ποιες πηγές υπάρχουν, σε ποια μορφή δημοσιεύονται, ποια πληροφορία λείπει και με τι κόστος μπορεί να ανακτηθεί. Καταγράφοντας αναλυτικά τα εμπόδια που συναντήθηκαν και τον τρόπο με τον οποίο ξεπεράστηκαν, καταδεικνύει ότι το εμπόδιο για την ύπαρξη μιας εθνικής υπηρεσίας πληροφόρησης δεν είναι τεχνικό, και συνεισφέρει έτσι στη συζήτηση για την ανάγκη ενός εθνικού προτύπου δημοσίευσης συγκοινωνιακών δεδομένων.

1.7 Οργάνωση της εργασίας

Η εργασία οργανώνεται σε επτά κεφάλαια, τα οποία ακολουθούν τη φυσική ροή από τη θεωρία προς την υλοποίηση και την αξιολόγηση. Η διάρθρωση αυτή επιτρέπει στον αναγνώστη να παρακολουθήσει τη διαδρομή από το αφηρημένο πρόβλημα δρομολόγησης έως το συγκεκριμένο λειτουργικό σύστημα, χωρίς να απαιτείται προηγούμενη εξοικείωση με το πεδίο.

Το **Κεφάλαιο 2** αναπτύσσει το θεωρητικό υπόβαθρο. Ξεκινά από τους βασικούς ορισμούς της θεωρίας γράφων και το πρόβλημα συντομότερης διαδρομής, παρουσιάζει τον αλγόριθμο Dijkstra με απόδειξη ορθότητας και ανάλυση πολυπλοκότητας, και προχωρά στα χρονοεξαρτώμενα δίκτυα και στη σημασία της ιδιότητας FIFO. Στη συνέχεια εξετάζει τα καθιερωμένα μοντέλα αναπαράστασης δικτύων χρονοδιαγράμματος, το πρόβλημα πρωιμότερης άφιξης, τη πολυκριτηριακή δρομολόγηση και τους σύγχρονους αλγορίθμους του πεδίου, και ολοκληρώνεται με την τυπική διατύπωση και τεκμηρίωση του προτεινόμενου αλγορίθμου.

Το **Κεφάλαιο 3** παρουσιάζει τις γλώσσες, τις βιβλιοθήκες και τα πρότυπα που χρησιμοποιήθηκαν. Καλύπτει τις τεχνολογίες ανάκτησης δεδομένων και ασύγχρονου προγραμματισμού, τη σχεσιακή αποθήκευση και τη σειριοποίηση, το πλαίσιο ανάπτυξης της διαδικτυακής εφαρμογής, τις γεωχωρικές υπηρεσίες που αξιοποιήθηκαν, το πρότυπο GTFS και τα κρυπτογραφικά εργαλεία που απαιτήθηκαν για την επικοινωνία με μία από τις πηγές.

Το **Κεφάλαιο 4** περιγράφει αναλυτικά τη διαδικασία ανάκτησης και ενοποίησης των δεδομένων. Περιλαμβάνει την καταγραφή και ταξινόμηση των πηγών, τη μεθοδολογία αντίστροφης μηχανίκευσης των μη τεκμηριωμένων υπηρεσιών, τη σχεδίαση των μηχανισμών συλλογής με έλεγχο ταυτοχρονισμού και πολιτική επαναλήψεων, τη στρατηγική εξαγωγής του εβδομαδιαίου προτύπου εκτέλεσης, την επίλυση του προβλήματος ταύτισης των στάσεων και, τέλος, τη σχεδίαση και φόρτωση της σχεσιακής βάσης δεδομένων.

Το **Κεφάλαιο 5** αφορά την ανάπτυξη του ΜΙΤΟΥ. Ξεκινά από τις λειτουργικές και μη λειτουργικές απαιτήσεις και τις περιπτώσεις χρήσης, περιγράφει την αρχιτεκτονική και τον διαχωρισμό των επιπέδων, και προχωρά στην υλοποίηση του υπολογιστικού πυρήνα και της διεπαφής χρήστη.

Το **Κεφάλαιο 6** παρουσιάζει τον ΜΙΤΟ σε λειτουργία μέσα από αντιπροσωπευτικά σενάρια χρήσης, από την απλή απευθείας σύνδεση έως τη σύνθετη διαδρομή με διανυκτέρευση, και αποτιμά την απόδοση και την ορθότητά του έναντι πραγματικών δρομολογίων.

Το **Κεφάλαιο 7** συνοψίζει τα συμπεράσματα, αναγνωρίζει τους περιορισμούς της προσέγγισης και διατυπώνει προτάσεις για μελλοντικές επεκτάσεις, τεχνικές αλλά και θεσμικές.

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Το κεφάλαιο αυτό συγκροτεί το εννοιολογικό και μαθηματικό υπόβαθρο πάνω στο οποίο στηρίζεται η υλοποίηση που παρουσιάζεται στη συνέχεια της εργασίας. Η ανάπτυξη ακολουθεί μια πορεία από το γενικό προς το ειδικό: ξεκινά από τους θεμελιώδεις ορισμούς της θεωρίας γράφων και το κλασικό πρόβλημα συντομότερης διαδρομής, εισάγει στη συνέχεια τη χρονική διάσταση που μετατρέπει το πρόβλημα σε χρονοεξαρτώμενο, εξετάζει τα καθιερωμένα μοντέλα αναπαράστασης δικτύων χρονοδιαγράμματος και τους σύγχρονους εξειδικευμένους αλγορίθμους του πεδίου, και καταλήγει στην τυπική διατύπωση και τεκμηρίωση του αλγορίθμου που αναπτύχθηκε για τις ανάγκες της παρούσας εργασίας. Στόχος δεν είναι η εξαντλητική επισκόπηση ενός εκτενούς ερευνητικού πεδίου, αλλά η ακριβής θεμελίωση των επιλογών που έγιναν και η αιτιολόγησή τους με βάση τα ιδιαίτερα χαρακτηριστικά του ελληνικού υπεραστικού δικτύου. Για τον λόγο αυτό, κάθε θεωρητική έννοια που εισάγεται συνοδεύεται από την πρακτική της συνέπεια για το συγκεκριμένο πρόβλημα, ενώ οι αποδείξεις ορθότητας παρατίθενται αναλυτικά όπου αποτελούν προϋπόθεση για την εγκυρότητα της υλοποίησης.

2.1 Θεωρία γράφων: βασικοί ορισμοί

2.1.1 Γράφοι, κατευθυνόμενοι γράφοι, βάρη ακμών

Ένας γράφος ορίζεται ως ένα διατεταγμένο ζεύγος $G = (V, E)$, όπου V είναι ένα πεπερασμένο σύνολο κόμβων ή κορυφών και E ένα σύνολο ακμών που συνδέουν ζεύγη κόμβων. Όταν οι ακμές δεν έχουν κατεύθυνση, δηλαδή όταν κάθε ακμή είναι ένα μη διατεταγμένο ζεύγος $\{u, v\}$, ο γράφος καλείται μη κατευθυνόμενος. Στην αντίθετη περίπτωση, όταν κάθε ακμή είναι ένα διατεταγμένο ζεύγος (u, v) που δηλώνει μετάβαση από τον u στον v και όχι απαραίτητα το αντίστροφο, ο γράφος καλείται κατευθυνόμενος.

Η διάκριση αυτή δεν είναι τυπολογική αλλά ουσιώδης για τη μοντελοποίηση συγκοινωνιακών δικτύων. Η ύπαρξη ενός δρομολογίου από την πόλη A προς την πόλη B δεν συνεπάγεται την ύπαρξη αντίστοιχου δρομολογίου από τη B προς την A την ίδια ημέρα ή την ίδια ώρα· στο ελληνικό υπεραστικό δίκτυο, όπως περιγράφηκε στο προηγούμενο κεφάλαιο, οι δύο κατευθύνσεις

μιας γραμμής εκτελούνται συχνά από διαφορετικούς φορείς με ανεξάρτητα χρονοδιαγράμματα. Το δίκτυο μοντελοποιείται επομένως αναγκαστικά ως κατευθυνόμενος γράφος.

Σε πολλές εφαρμογές κάθε ακμή συνοδεύεται από ένα βάρος, δηλαδή μια συνάρτηση $w: E \rightarrow \mathbb{R}$ που εκφράζει το κόστος διάσχισής της — απόσταση, χρόνο, χρηματικό αντίτιμο ή οποιοδήποτε άλλο μετρήσιμο μέγεθος. Στα προβλήματα που εξετάζονται εδώ το βάρος αντιστοιχεί σε χρόνο και είναι, ως εκ τούτου, πάντοτε μη αρνητικό — παρατήρηση καθοριστική για την εφαρμοσιμότητα του αλγορίθμου Dijkstra.

2.1.2 Μονοπάτια, κύκλοι, συνεκτικότητα

Μονοπάτι από έναν κόμβο v_0 σε έναν κόμβο v_k ονομάζεται μια ακολουθία κόμβων $\square v_0, v_1, \dots, v_k \square$ τέτοια ώστε για κάθε i να ισχύει $(v_{i-1}, v_i) \in E$. Το μήκος του μονοπατιού ορίζεται ως το άθροισμα των βαρών των ακμών που το απαρτίζουν. Ένα μονοπάτι καλείται απλό όταν δεν επαναλαμβάνει κόμβο, ενώ κύκλος ονομάζεται ένα μονοπάτι στο οποίο ο αρχικός και ο τελικός κόμβος ταυτίζονται.

Η έννοια του κύκλου έχει άμεση σημασία για τη δρομολόγηση. Σε δίκτυο με μη αρνητικά βάρη κανένα βέλτιστο μονοπάτι δεν χρειάζεται να περιέχει κύκλο, καθώς η απαλοιφή του δεν αυξάνει το κόστος· η αναζήτηση περιορίζεται έτσι στα απλά μονοπάτια και ο χώρος των υποψήφιων λύσεων παραμένει πεπερασμένος. Αντιθέτως, οι κύκλοι αρνητικού βάρους καθιστούν το πρόβλημα μη καλά ορισμένο, αφού το κόστος μειώνεται απεριόριστα με επαναλαμβανόμενη διάσχισή τους.

Ένας γράφος καλείται συνεκτικός όταν υπάρχει μονοπάτι μεταξύ κάθε ζεύγους κόμβων του· στους κατευθυνόμενους γράφους διακρίνεται περαιτέρω η ισχυρή συνεκτικότητα, που απαιτεί μονοπάτι και προς τις δύο κατευθύνσεις, από την ασθενή, που αρκείται στη συνεκτικότητα του υποκείμενου μη κατευθυνόμενου γράφου. Σε ένα δίκτυο συγκοινωνιών η συνεκτικότητα δεν είναι δεδομένη: στάσεις που εξυπηρετούνται μόνο ως προορισμοί και ποτέ ως αφετηρίες σχηματίζουν συνιστώσες από τις οποίες δεν υπάρχει έξοδος, γεγονός με πρακτικές συνέπειες στην ανάλυση προσβασιμότητας.

2.1.3 Αναπαράσταση γράφων και υπολογιστικές επιπτώσεις

Δύο είναι οι καθιερωμένοι τρόποι αναπαράστασης ενός γράφου. Ο πίνακας γειτνίασης είναι δισδιάστατος πίνακας μεγέθους $|V| \times |V|$, στη θέση (i, j) του οποίου καταχωρίζεται το βάρος της ακμής από τον i στον j ή ειδική τιμή που δηλώνει την απουσία της· επιτρέπει έλεγχο ύπαρξης ακμής σε σταθερό χρόνο, απαιτεί όμως χώρο $\Theta(|V|^2)$ ανεξάρτητα από το πλήθος των ακμών. Οι λίστες γειτνίασης αποθηκεύουν, για κάθε κόμβο, τις εξερχόμενες ακμές του: ο χώρος περιορίζεται σε $\Theta(|V| + |E|)$ και η απαρίθμηση των γειτόνων κοστίζει όσο ο βαθμός του κόμβου, ενώ ο έλεγχος ύπαρξης συγκεκριμένης ακμής απαιτεί γραμμική σάρωση.

Η επιλογή καθορίζεται από την πυκνότητα του γράφου. Τα συγκοινωνιακά δίκτυα είναι χαρακτηριστικά αραιά: κάθε στάση συνδέεται απευθείας με λίγες μόνο άλλες, ώστε το $|E|$ να είναι της τάξης του $|V|$. Στο σύνολο δεδομένων της εργασίας, οι 2.310 στάσεις εμφανίζουν 460 μόνο αφετηρίες με εξερχόμενα δρομολόγια. Η υλοποίηση χρησιμοποιεί, συνεπώς, λίστες γειτνίασης υπό

τη μορφή λεξικού που αντιστοιχίζει κάθε στάση στον κατάλογο των αναχωρήσεων της, ταξινομημένο κατά ώρα — διάταξη που αξιοποιείται από τον αλγόριθμο αναζήτησης.

2.2 Το πρόβλημα συντομότερης διαδρομής

2.2.1 Ορισμός και παραλλαγές

Δοθέντος ενός κατευθυνόμενου γράφου $G = (V, E)$ με συνάρτηση βαρών $w: E \rightarrow \mathbb{R}_{\geq 0}$, το πρόβλημα της συντομότερης διαδρομής συνίσταται στην εύρεση, για ένα δεδομένο ζεύγος κόμβων, εκείνου του μονοπατιού που ελαχιστοποιεί το άθροισμα των βαρών των ακμών του. Η απόσταση $d(u, v)$ ορίζεται ως το ελάχιστο αυτό κόστος, ή ως $+\infty$ όταν δεν υπάρχει μονοπάτι από τον u στον v .

Το πρόβλημα εμφανίζεται σε τέσσερις βασικές παραλλαγές. Στην παραλλαγή μοναδικής πηγής (single-source) ζητούνται οι αποστάσεις από έναν κόμβο προς όλους τους υπόλοιπους. Στην παραλλαγή σημείου προς σημείο (point-to-point) ενδιαφέρει ένα μόνο ζεύγος αφετηρίας και προορισμού. Η παραλλαγή μοναδικού προορισμού (single-destination) ανάγεται στην πρώτη με αντιστροφή της κατεύθυνσης των ακμών, ενώ η παραλλαγή όλων των ζευγών (all-pairs) απαιτεί τον υπολογισμό ολόκληρου του πίνακα αποστάσεων.

Η διάκριση έχει πρακτική σημασία, διότι δεν είναι γνωστός αλγόριθμος που να επιλύει την παραλλαγή σημείου προς σημείο ασυμπτωτικά ταχύτερα από την παραλλαγή μοναδικής πηγής στη χειρότερη περίπτωση. Στην πράξη, ωστόσο, η αναζήτηση μπορεί να τερματιστεί μόλις οριστικοποιηθεί ο ζητούμενος προορισμός, γεγονός που περιορίζει σημαντικά τον όγκο της εργασίας. Η παρούσα εργασία αντιμετωπίζει πρωτίστως ερωτήματα σημείου προς σημείο, καθώς ο χρήστης δηλώνει συγκεκριμένη αφετηρία και συγκεκριμένο προορισμό.

2.2.2 Η αρχή της βελτιστότητας και η υποδομή βέλτιστης λύσης

Η αλγοριθμική επίλυση του προβλήματος στηρίζεται σε μια θεμελιώδη δομική ιδιότητα, γνωστή ως υποδομή βέλτιστης λύσης: κάθε υπομονοπάτι ενός βέλτιστου μονοπατιού είναι το ίδιο βέλτιστο. Τυπικά, αν το $p = \langle v_0, \dots, v_k \rangle$ είναι συντομότερο μονοπάτι από τον v_0 στον v_k , τότε για κάθε $0 \leq i \leq j \leq k$ το υπομονοπάτι $\langle v_i, \dots, v_j \rangle$ είναι συντομότερο μονοπάτι από τον v_i στον v_j .

Η απόδειξη γίνεται με απαγωγή σε άτοπο. Αν υπήρχε μονοπάτι p' από τον v_i στον v_j με κόστος αυστηρά μικρότερο από εκείνο του υπομονοπατιού, η αντικατάσταση του τελευταίου από το p' θα παρήγε μονοπάτι από τον v_0 στον v_k με συνολικό κόστος μικρότερο του $w(p)$, αντιφάσκοντας προς τη βελτιστότητα του p .

Η ιδιότητα αυτή είναι εκείνη που καθιστά εφικτή τόσο τη δυναμική προγραμματιστική προσέγγιση όσο και την άπληστη στρατηγική του Dijkstra. Επιτρέπει τη σταδιακή οικοδόμηση της λύσης μέσω της λεγόμενης χαλάρωσης (relaxation): διατηρείται για κάθε κόμβο v μια εκτίμηση $d[v]$ της απόστασής του από την πηγή, η οποία αποτελεί πάντοτε άνω φράγμα της πραγματικής απόστασης, και για κάθε ακμή (u, v) ελέγχεται αν η διαδρομή μέσω του u βελτιώνει την τρέχουσα εκτίμηση, οπότε $d[v] \leftarrow d[u] + w(u, v)$.

2.2.3 Ο αλγόριθμος Dijkstra: περιγραφή και ψευδοκώδικας

Ο αλγόριθμος που διατύπωσε ο Edsger W. Dijkstra το 1959 επιλύει το πρόβλημα μοναδικής πηγής σε γράφους με μη αρνητικά βάρη [5]. Η στρατηγική του είναι άπληστη: διατηρεί ένα σύνολο S κόμβων των οποίων η απόσταση έχει οριστικοποιηθεί και, σε κάθε επανάληψη, επιλέγει από τους υπόλοιπους εκείνον με τη μικρότερη τρέχουσα εκτίμηση, τον εντάσσει στο S και χαλαρώνει τις εξερχόμενες ακμές του.

Η κεντρική διαίσθηση είναι ότι, εφόσον όλα τα βάρη είναι μη αρνητικά, ο κόμβος με τη μικρότερη εκτίμηση δεν μπορεί να βελτιωθεί περαιτέρω: οποιοδήποτε εναλλακτικό μονοπάτι προς αυτόν θα έπρεπε να περάσει από κάποιον κόμβο εκτός του S , του οποίου η εκτίμηση είναι ήδη μεγαλύτερη ή ίση, και κάθε επιπλέον ακμή δεν μπορεί παρά να αυξήσει το κόστος.

```
Dijkstra(G, w, s):
  για κάθε  $v \in V$ :  $d[v] \leftarrow \infty$ ,  $\pi[v] \leftarrow \text{NIL}$ 
   $d[s] \leftarrow 0$ 
   $S \leftarrow \emptyset$ ;  $Q \leftarrow$  ουρά προτεραιότητας με όλους τους  $v \in V$ , κλειδί  $d[v]$ 
  όσο  $Q \neq \emptyset$ :
     $u \leftarrow \text{EXTRACT-MIN}(Q)$ 
     $S \leftarrow S \cup \{u\}$ 
    για κάθε ακμή  $(u, v) \in E$ :
      αν  $d[u] + w(u, v) < d[v]$ :
         $d[v] \leftarrow d[u] + w(u, v)$ 
         $\pi[v] \leftarrow u$ 
         $\text{DECREASE-KEY}(Q, v, d[v])$ 
  επίστρεψε  $(d, \pi)$ 
```

Ο πίνακας π των προκατόχων επιτρέπει την ανασύσταση του ίδιου του μονοπατιού και όχι μόνο του κόστους του, μέσω οπισθοδρομικής διάσχισης από τον προορισμό προς την πηγή. Η δομή αυτή σχηματίζει το λεγόμενο δέντρο συντομότερων μονοπατιών με ρίζα την πηγή.

2.2.4 Απόδειξη ορθότητας

Θα αποδειχθεί ότι κατά τον τερματισμό ισχύει $d[v] = \delta(s, v)$ για κάθε $v \in V$. Η απόδειξη στηρίζεται στην ακόλουθη αναλλοίωτη: κατά την εξαγωγή κάθε κόμβου u από την ουρά προτεραιότητας ισχύει $d[u] = \delta(s, u)$.

Έστω, προς άτοπο, ότι ο u είναι ο πρώτος κόμβος που εξάγεται με $d[u] > \delta(s, u)$. Προφανώς $u \neq s$, διότι $d[s] = 0 = \delta(s, s)$ κατά την αρχικοποίηση. Επίσης πρέπει να υπάρχει μονοπάτι από τον s στον u , αλλιώς $\delta(s, u) = \infty \geq d[u]$. Έστω p ένα συντομότερο μονοπάτι από τον s στον u . Τη στιγμή της εξαγωγής του u , το p ξεκινά από κόμβο εντός του S και καταλήγει σε κόμβο εκτός αυτού· έστω y ο πρώτος κόμβος του p που δεν ανήκει στο S και x ο προκατόχός του, ο οποίος ανήκει στο S .

Επειδή ο x εξήχθη πριν από τον u και ο u είναι εξ υποθέσεως ο πρώτος που παραβιάζει την αναλλοίωτη, ισχύει $d[x] = \delta(s, x)$. Κατά την εξαγωγή του x χαλαρώθηκε η ακμή (x, y) , οπότε $d[y] \leq d[x] + w(x, y) = \delta(s, x) + w(x, y) = \delta(s, y)$, όπου η τελευταία ισότητα προκύπτει από την υποδομή βέλτιστης λύσης, αφού ο y βρίσκεται πάνω σε συντομότερο μονοπάτι προς τον u . Καθώς η d αποτελεί πάντοτε άνω φράγμα της δ , συμπεραίνουμε $d[y] = \delta(s, y)$.

Ο y προηγείται του u στο μονοπάτι p και όλα τα βάρη είναι μη αρνητικά, άρα $\delta(s, y) \leq \delta(s, u)$. Συνδυάζοντας, προκύπτει $d[y] = \delta(s, y) \leq \delta(s, u) < d[u]$. Όμως και οι δύο κόμβοι y και u βρίσκονταν στην ουρά τη στιγμή της εξαγωγής, και ο αλγόριθμος επέλεξε τον u ως ελάχιστο, δηλαδή $d[u] \leq d[y]$. Οι δύο σχέσεις είναι αντιφατικές, οπότε η αρχική υπόθεση απορρίπτεται και η αναλλοίωτη ισχύει.

Είναι σημαντικό να επισημανθεί σε ποιο ακριβώς σημείο χρησιμοποιήθηκε η μη αρνητικότητα των βαρών: στο βήμα $\delta(s, y) \leq \delta(s, u)$. Αν επιτρέπονταν αρνητικά βάρη, το υπόλοιπο τμήμα του μονοπατιού από τον y στον u θα μπορούσε να έχει αρνητικό κόστος και η ανισότητα να αντιστραφεί, ακυρώνοντας το επιχείρημα.

2.2.5 Ανάλυση πολυπλοκότητας

Η πολυπλοκότητα εξαρτάται από την υλοποίηση της ουράς προτεραιότητας. Ο αλγόριθμος εκτελεί $|V|$ πράξεις εξαγωγής ελαχίστου και το πολύ $|E|$ πράξεις μείωσης κλειδιού, καθώς κάθε ακμή χαλαρώνεται ακριβώς μία φορά.

Με απλό πίνακα, όπου η εξαγωγή ελαχίστου απαιτεί γραμμική σάρωση $O(|V|)$ και η μείωση κλειδιού $O(1)$, ο συνολικός χρόνος είναι $O(|V|^2 + |E|) = O(|V|^2)$. Με δυαδικό σωρό, όπου αμφότερες οι πράξεις κοστίζουν $O(\log |V|)$, προκύπτει $O((|V| + |E|) \cdot \log |V|)$, το οποίο για αραιούς γράφους είναι σαφώς προτιμότερο. Με σωρό Fibonacci η μείωση κλειδιού εκτελείται σε επιμερισμένο σταθερό χρόνο και η πολυπλοκότητα βελτιώνεται θεωρητικά σε $O(|V| \cdot \log |V| + |E|)$, πλην όμως οι σταθεροί παράγοντες της δομής την καθιστούν σπάνια συμφέρουσα στην πράξη [6].

Στην υλοποίηση της παρούσας εργασίας χρησιμοποιείται δυαδικός σωρός μέσω της τυπικής βιβλιοθήκης `heapq` της Python, χωρίς πράξη μείωσης κλειδιού: αντί να ενημερώνεται το κλειδί ενός υπάρχοντος στοιχείου, εισάγεται νέα καταχώριση και οι παρωχημένες αγνοούνται κατά την εξαγωγή μέσω ελέγχου έναντι της καλύτερης γνωστής τιμής. Η τεχνική αυτή, γνωστή ως τεμπέλικη διαγραφή, αυξάνει το μέγεθος του σωρού έως $O(|E|)$ αλλά διατηρεί την ίδια ασυμπτωτική συμπεριφορά και είναι σημαντικά απλούστερη στην υλοποίηση.

2.2.6 Προϋποθέσεις εφαρμογής και όρια του κλασικού Dijkstra

Ο αλγόριθμος στη μορφή που παρουσιάστηκε προϋποθέτει τρία πράγματα: μη αρνητικά βάρη, στατική συνάρτηση κόστους και ένα μόνο κριτήριο βελτιστοποίησης. Η πρώτη προϋπόθεση ικανοποιείται τετριμμένα στα συγκοινωνιακά δίκτυα, όπου το κόστος είναι χρόνος. Οι άλλες δύο, όμως, παραβιάζονται συστηματικά.

Η υπόθεση της στατικότητας καταρρέει διότι το κόστος μετάβασης από μια στάση σε μια άλλη δεν είναι σταθερό: εξαρτάται από τη στιγμή άφιξης του επιβάτη στην πρώτη στάση, η οποία καθορίζει πόσο θα αναμείνει έως την επόμενη διαθέσιμη αναχώρηση. Δύο επιβάτες που φτάνουν στον ίδιο σταθμό με διαφορά λίγων λεπτών ενδέχεται να αντιμετωπίσουν εντελώς διαφορετικό κόστος, αν ο ένας προλάβει και ο άλλος χάσει το ίδιο δρομολόγιο.

Η υπόθεση του μοναδικού κριτηρίου καταρρέει διότι, όπως αναπτύχθηκε στο πρώτο κεφάλαιο, ο επιβάτης σταθμίζει ταυτόχρονα την ώρα άφιξης, το πλήθος των μετεπιβιβάσεων και τον χρόνο αναμονής, μεγέθη μεταξύ των οποίων δεν υπάρχει φυσική διάταξη.

Οι δύο αυτές αποκλίσεις οριοθετούν τη συνέχεια του κεφαλαίου: η ενότητα 2.3 πραγματεύεται την άρση της πρώτης μέσω των χρονοεξαρτώμενων δικτύων, ενώ η ενότητα 2.6 της δεύτερης μέσω της πολυκριτηριακής δρομολόγησης.

2.3 Χρονοεξαρτώμενα δίκτυα

2.3.1 Η έννοια της χρονοεξαρτώμενης συνάρτησης κόστους

Σε ένα χρονοεξαρτώμενο δίκτυο το βάρος μιας ακμής δεν είναι σταθερός αριθμός αλλά συνάρτηση του χρόνου. Για κάθε ακμή $(u, v) \in E$ ορίζεται μια συνάρτηση $f_{uv}: T \rightarrow \mathbb{R}_{\geq 0}$, όπου T το πεδίο του χρόνου, τέτοια ώστε το $f_{uv}(t)$ να εκφράζει τον χρόνο που απαιτείται για τη διάσχιση της ακμής εφόσον η αναχώρηση από τον u γίνει τη στιγμή t . Ισοδύναμα και συχνά ευκολότερα, ορίζεται η συνάρτηση άφιξης $a_{uv}(t) = t + f_{uv}(t)$, η οποία επιστρέφει απευθείας τη στιγμή άφιξης στον v .

Η διατύπωση καλύπτει δύο διαφορετικές καταστάσεις. Στην οδική κυκλοφορία η συνάρτηση είναι συνεχής και εκφράζει τη μεταβολή του χρόνου διαδρομής ανάλογα με τον φόρτο· στα δίκτυα χρονοδιαγράμματος, που ενδιαφέρουν εδώ, είναι βηματική και εκφράζει τη διακριτότητα των αναχωρήσεων.

Πράγματι, όταν από τη στάση u προς τη στάση v εκτελούνται δρομολόγια σε προκαθορισμένες στιγμές $t_1 < t_2 < \dots < t_m$ με αντίστοιχες διάρκειες $c_1, \dots, c_m$, η συνάρτηση άφιξης παίρνει τη μορφή

$$a_{uv}(t) = \min \{ t_i + c_i : t_i \geq t \}.$$

Ο επιβάτης που φτάνει τη στιγμή t αναμένει έως την πρώτη διαθέσιμη αναχώρηση, ώστε το κόστος να περιλαμβάνει και την αναμονή και τη διαδρομή. Το κόστος μπορεί να διαφέρει δραματικά για γειτονικές τιμές του t : σε αραιό δίκτυο όπως το ελληνικό υπεραστικό, όπου μεταξύ διαδοχικών αναχωρήσεων μεσολαβούν συχνά ώρες, μια καθυστέρηση ενός λεπτού μεταφράζεται σε τετράωρη αναμονή.

2.3.2 Η ιδιότητα FIFO (non-overtaking) και η σημασία της για την ορθότητα

Μια χρονοεξαρτώμενη ακμή ικανοποιεί την ιδιότητα FIFO, ή ιδιότητα μη προσπέρασης, όταν η συνάρτηση άφιξης της είναι μη φθίνουσα, δηλαδή όταν για κάθε $t_1 \leq t_2$ ισχύει $a_{uv}(t_1) \leq a_{uv}(t_2)$. Διαισθητικά, η καθυστερημένη αναχώρηση δεν μπορεί ποτέ να οδηγήσει σε πρωιμότερη άφιξη: κανείς δεν κερδίζει περιμένοντας.

Η σημασία της ιδιότητας είναι θεμελιώδης. Αν παραβιαζόταν, η αναμονή στην αφετηρία θα αποτελούσε ενδεχομένως συμφέρουσα στρατηγική και η κατάσταση της αναζήτησης δεν θα μπορούσε πλέον να συνοψιστεί στον πρωιμότερο χρόνο άφιξης ανά κόμβο: θα έπρεπε να διατηρείται ολόκληρο το σύνολο των μη κυριαρχούμενων ζευγών χρόνου και κόστους, μετατρέποντας το πρόβλημα σε ουσιαστικά πολυκριτηριακό. Είναι γνωστό ότι η συντομότερη διαδρομή σε γενικά χρονοεξαρτώμενα δίκτυα χωρίς την ιδιότητα FIFO και χωρίς δυνατότητα αναμονής είναι NP-δύσκολο πρόβλημα [7].

Η βηματική συνάρτηση της προηγούμενης ενότητας ικανοποιεί την ιδιότητα εξ ορισμού: καθώς το $a_{uv}(t)$ ορίστηκε ως το ελάχιστο πάνω στο σύνολο των αναχωρήσεων με $t_i \geq t$, η αύξηση του t μπορεί μόνο να συρρικνώσει το σύνολο αυτό, οπότε το ελάχιστο δεν μπορεί να μειωθεί. Η ιδιότητα εξασφαλίζεται, δηλαδή, από την ίδια τη λήψη του ελαχίστου και όχι από κάποια πρόσθετη υπόθεση για τα δεδομένα.

Πρέπει ωστόσο να διευκρινιστεί ότι η ιδιότητα αφορά τη συνάρτηση κόστους της ακμής και όχι τα επιμέρους δρομολόγια. Είναι απολύτως δυνατό ένα ταχύτερο λεωφορείο που αναχωρεί αργότερα να προσπεράσει ένα βραδύτερο, ώστε $t_2 + c_2 < t_1 + c_1$. Η λήψη του ελαχίστου απορροφά ακριβώς αυτή την περίπτωση: ο επιβάτης που φτάνει πριν από την t_1 μπορεί να επιλέξει το ταχύτερο δρομολόγιο, οπότε η άφιξή του δεν είναι ποτέ χειρότερη από εκείνου που φτάνει αργότερα.

2.3.3 Επέκταση του Dijkstra σε χρονοεξαρτώμενα δίκτυα

Η προσαρμογή του αλγορίθμου στο χρονοεξαρτώμενο πλαίσιο είναι εντυπωσιακά περιορισμένη σε έκταση. Η ετικέτα $d[v]$ επανερμηνεύεται: δεν εκφράζει πλέον αφηρημένο κόστος αλλά τον πρωιμότερο χρόνο κατά τον οποίο μπορεί να βρίσκεται κανείς στον κόμβο v , δεδομένης αναχώρησης από την πηγή s τη στιγμή t_0 . Η αρχικοποίηση γίνεται με $d[s] = t_0$ αντί για μηδέν, και η πράξη της χαλάρωσης αντικαθίσταται από τη σχέση

$$d[v] \leftarrow \min\{d[v], a_{uv}(d[u])\}.$$

Η δομή της άπληστης επιλογής παραμένει αναλλοίωτη: εξάγεται πάντοτε ο κόμβος με τον μικρότερο χρόνο και οι εξερχόμενες ακμές του αποτιμώνται στη στιγμή που αντιστοιχεί σε αυτόν. Η ουσιώδης διαφορά είναι ότι η αποτίμηση της a_{uv} δεν μπορεί να προηγηθεί της οριστικοποίησης του $d[u]$. Η σειρά επεξεργασίας που επιβάλλει ο Dijkstra —κατά αύξοντα χρόνο— εξασφαλίζει ακριβώς ότι, όταν χαλαρώνεται μια ακμή, ο χρόνος στον οποίο αποτιμάται είναι ήδη οριστικός.

Στην πράξη η αποτίμηση δεν γίνεται συμβολικά αλλά με σάρωση του ταξινομημένου καταλόγου των δρομολογίων που ξεκινούν από τη στάση u : αναζητείται το πρώτο του οποίου η ώρα αναχώρησης δεν προηγείται του $d[u]$ προσαυξημένου κατά τον ελάχιστο χρόνο ανταπόκρισης και το οποίο εκτελείται την τρέχουσα ημέρα.

2.3.4 Απόδειξη ορθότητας υπό την υπόθεση FIFO

Η απόδειξη της ενότητας 2.2.4 μεταφέρεται σχεδόν αυτούσια, αρκεί να αντικατασταθεί η προσθετικότητα του κόστους από τη μονοτονία των συναρτήσεων άφιξης. Έστω, προς άτοπο, ότι ο u είναι ο πρώτος κόμβος που εξάγεται με $d[u] > \delta_0(s, u)$, όπου δ_0 συμβολίζει τον πραγματικό πρωιμότερο χρόνο άφιξης για αναχώρηση τη στιγμή t_0 . Έστω p ένα βέλτιστο ως προς τον χρόνο μονοπάτι προς τον u , έστω y ο πρώτος κόμβος του εκτός του συνόλου S των οριστικοποιημένων και x ο προκάτοχός του εντός αυτού.

Για τον x ισχύει $d[x] = \delta_0(s, x)$, αφού εξήχθη πριν από τον u . Κατά την εξαγωγή του χαλαρώθηκε η ακμή (x, y) , οπότε $d[y] \leq a_{xy}(d[x]) = a_{xy}(\delta_0(s, x)) = \delta_0(s, y)$, όπου η τελευταία ισότητα προκύπτει από την υποδομή βέλτιστης λύσης, η οποία εξακολουθεί να ισχύει: λόγω της ιδιότητας

FIFO, μια πρωιμότερη άφιξη στον y δεν θα μπορούσε να οδηγήσει σε χειρότερη συνέχεια, οπότε θα βελτιώνει και τη συνολική άφιξη στον u .

Επειδή ο y προηγείται του u στο μονοπάτι και ο χρόνος είναι μονότονα αύξων κατά μήκος κάθε διαδρομής —καμία ακμή δεν έχει αρνητική διάρκεια, ώστε $a_{uv}(t) \geq t$ — ισχύει $\delta_\theta(s, y) \leq \delta_\theta(s, u)$. Συνδυάζοντας προκύπτει $d[y] \leq \delta_\theta(s, u) < d[u]$, το οποίο αντιφάσκει προς την επιλογή του u ως ελαχίστου της ουράς. Η αναλλοίωτη ισχύει, επομένως, και ο αλγόριθμος υπολογίζει ορθά τους πρωιμότερους χρόνους άφιξης.

Οι δύο υποθέσεις διαδραματίζουν διακριτό ρόλο. Η μη αρνητικότητα των διαρκειών εξασφαλίζει τη μονοτονία του χρόνου κατά μήκος του μονοπατιού και άρα τη νομιμότητα της άπληστης επιλογής, ενώ η ιδιότητα FIFO εξασφαλίζει την υποδομή βέλτιστης λύσης: χωρίς αυτήν, η πρωιμότερη άφιξη σε ενδιαμέσο κόμβο δεν θα ήταν κατ' ανάγκη προτιμότερη και η άπληστη οριστικοποίηση θα κατέρρεε.

2.4 Μοντελοποίηση δικτύων συγκοινωνιών βάσει χρονοδιαγράμματος

2.4.1 Ιδιαιτερότητες: διακριτές αναχωρήσεις, περιοδικότητα, ανταποκρίσεις

Πριν από την παρουσίαση των μοντέλων, είναι σκόπιμο να καταγραφούν τα χαρακτηριστικά που πρέπει να αποτυπώσουν. Το πρώτο είναι η διακριτότητα: η μετάβαση μεταξύ δύο στάσεων είναι δυνατή μόνο σε συγκεκριμένες στιγμές, τις προγραμματισμένες αναχωρήσεις, και όχι σε οποιοδήποτε σημείο του χρονικού άξονα. Το δεύτερο είναι η περιοδικότητα: το χρονοδιάγραμμα επαναλαμβάνεται με εβδομαδιαία περίοδο, με τη διαφοροποίηση ότι δεν εκτελούνται όλα τα δρομολόγια όλες τις ημέρες. Το τρίτο είναι ο περιορισμός ανταπόκρισης: η μετεπιβίβαση από ένα όχημα σε άλλο απαιτεί έναν ελάχιστο χρόνο, ώστε ο επιβάτης να μεταβεί φυσικά από την αποβίβαση στην επιβίβαση.

Το τρίτο χαρακτηριστικό είναι και το πλέον απαιτητικό από μοντελοποιητική άποψη, διότι εισάγει εξάρτηση από την ιστορία της διαδρομής: το αν μια αναχώρηση είναι προσπελάσιμη εξαρτάται όχι μόνο από τη στιγμή άφιξης στη στάση αλλά και από το αν ο επιβάτης έφτασε εκεί με άλλο όχημα ή βρισκόταν ήδη στη στάση. Η αντιμετώπιση του ζητήματος διαφοροποιεί τα δύο καθιερωμένα μοντέλα αναπαράστασης, τα οποία παρουσιάζονται στη συνέχεια.

2.4.2 Το time-expanded μοντέλο

Στο χρονικά αναπτυγμένο μοντέλο ο χρόνος ενσωματώνεται στην ίδια τη δομή του γράφου. Κάθε κόμβος αντιστοιχεί όχι σε στάση αλλά σε ένα ζεύγος στάσης και χρονικής στιγμής: υπάρχει, δηλαδή, χωριστός κόμβος για κάθε αναχώρηση και κάθε άφιξη που καταγράφεται στο χρονοδιάγραμμα. Οι ακμές είναι δύο ειδών. Οι ακμές δρομολογίου συνδέουν τον κόμβο αναχώρησης ενός σκέλους με τον κόμβο άφιξής του, με βάρος τη διάρκεια της διαδρομής. Οι ακμές αναμονής συνδέουν χρονικά διαδοχικούς κόμβους της ίδιας στάσης, εκφράζοντας τη δυνατότητα παραμονής του επιβάτη σε αυτήν.

Το κύριο πλεονέκτημα του μοντέλου είναι η εννοιολογική του απλότητα: ο γράφος που προκύπτει είναι στατικός, με σταθερά μη αρνητικά βάρη, οπότε ο κλασικός Dijkstra εφαρμόζεται αυτούσιος και χωρίς καμία τροποποίηση. Οι περιορισμοί ανταπόκρισης ενσωματώνονται φυσικά, μέσω κατάλληλης διάσπασης των κόμβων άφιξης και αναχώρησης και προσθήκης ακμών μετεπιβίβασης με βάρος τον ελάχιστο χρόνο ανταπόκρισης· η διαδρομή που παραμένει στο ίδιο όχημα δεν διέρχεται από τις ακμές αυτές και δεν επιβαρύνεται.

Το μειονέκτημα είναι το μέγεθος. Ο αριθμός των κόμβων είναι ανάλογος του συνολικού πλήθους των γεγονότων του χρονοδιαγράμματος και όχι των στάσεων, ενώ η επέκταση του χρονικού ορίζοντα σε περισσότερες ημέρες πολλαπλασιάζει αντίστοιχα τη δομή. Για ένα σύνολο δεδομένων της τάξης των εκατό χιλιάδων δρομολογίων και ορίζοντα δύο ημερών, ο γράφος θα περιλάμβανε εκατοντάδες χιλιάδες κόμβους — μέγεθος διαχειρίσιμο αλλά αισθητά σπάταλο σε μνήμη.

2.4.3 Το time-dependent μοντέλο

Στο χρονοεξαρτώμενο μοντέλο ο γράφος παραμένει μικρός και ο χρόνος μετατοπίζεται από τη δομή στη συνάρτηση κόστους. Κάθε κόμβος αντιστοιχεί σε μία στάση και κάθε ακμή σε ένα ζεύγος στάσεων που συνδέονται με τουλάχιστον ένα δρομολόγιο. Το βάρος της ακμής δεν είναι αριθμός αλλά η βηματική συνάρτηση της ενότητας 2.3.1, η οποία περικλείει το σύνολο των αναχωρήσεων του συγκεκριμένου τμήματος.

Το μέγεθος του γράφου γίνεται έτσι ανάλογο του πλήθους των στάσεων και των διακριτών συνδέσεων, ανεξάρτητο από τον χρονικό ορίζοντα: η επέκταση σε περισσότερες ημέρες δεν προσθέτει κόμβους αλλά μεταβάλλει μόνο τον τρόπο αποτίμησης των συναρτήσεων. Το κόστος μετατοπίζεται στην αποτίμηση, η οποία απαιτεί αναζήτηση της κατάλληλης αναχώρησης εντός του καταλόγου κάθε ακμής — πράξη λογαριθμική με δυαδική αναζήτηση σε ταξινομημένη δομή.

Το βασικό μειονέκτημα του μοντέλου είναι η δυσκολία ενσωμάτωσης των περιορισμών ανταπόκρισης, ακριβώς επειδή αυτοί εξαρτώνται από την προϊστορία της διαδρομής, την οποία ο συμπυκνμένος γράφος δεν καταγράφει. Η αντιμετώπιση απαιτεί είτε εμπλουτισμό της κατάστασης της αναζήτησης με πληροφορία για το τρέχον όχημα, είτε την υιοθέτηση ενός σταθερού ελάχιστου χρόνου ανταπόκρισης που εφαρμόζεται ομοιόμορφα σε κάθε μετεπιβίβαση.

2.4.4 Συγκριτική αποτίμηση των δύο μοντέλων

Η βιβλιογραφία έχει συγκρίνει συστηματικά τα δύο μοντέλα, με κυριότερη αναφορά την εργασία των Pyrga και συνεργατών, η οποία καταλήγει ότι το χρονοεξαρτώμενο μοντέλο υπερτερεί σαφώς σε χρόνο απόκρισης και κατανάλωση μνήμης για τα κλασικά προβλήματα πρωιμότερης άφιξης, ενώ το χρονικά αναπτυγμένο προσφέρει μεγαλύτερη ευελιξία στην ενσωμάτωση σύνθετων περιορισμών και στην επέκταση προς πολυκριτηριακά ερωτήματα [8].

Η αποτίμηση μπορεί να αναλυθεί σε τρεις άξονες. Ο πρώτος είναι η κατανάλωση μνήμης, όπου η υπεροχή του χρονοεξαρτώμενου μοντέλου είναι αδιαμφισβήτητη, καθώς το μέγεθός του καθορίζεται από την τοπολογία του δικτύου και όχι από το πλήθος των γεγονότων του χρονοδιαγράμματος. Ο δεύτερος είναι η ταχύτητα απόκρισης, όπου η υπεροχή είναι μεν υπαρκτή αλλά μικρότερη, διότι μέρος του κόστους απλώς μετατοπίζεται από τη διάσχιση του γράφου στην αποτίμηση των συναρτήσεων. Ο τρίτος είναι η εκφραστικότητα, όπου η σχέση αντιστρέφεται: το

χρονικά αναπτυγμένο μοντέλο αποτυπώνει φυσικά κάθε περιορισμό που αφορά μεμονωμένο γεγονός, ενώ το χρονοεξαρτώμενο μόνο όσους εκφράζονται ομοιόμορφα σε επίπεδο ακμής.

Η επιλογή είναι, εν τέλει, ζήτημα ισορροπίας ανάμεσα στην πολυπλοκότητα της δομής και στην πολυπλοκότητα του αλγορίθμου: το πρώτο μοντέλο επενδύει σε μεγάλο γράφο ώστε να χρησιμοποιεί απλό αλγόριθμο, το δεύτερο σε μικρό γράφο με τίμημα έναν πιο σύνθετο αλγόριθμο. Για τις ανάγκες της παρούσας εργασίας επιλέχθηκε το χρονοεξαρτώμενο μοντέλο, με την αιτιολόγηση της επιλογής να αναπτύσσεται στην ενότητα 2.9.2.

2.5 Το πρόβλημα πρωιμότερης άφιξης (Earliest Arrival Problem)

2.5.1 Τυπικός ορισμός

Το πρόβλημα πρωιμότερης άφιξης αποτελεί τη βασικότερη διατύπωση της δρομολόγησης σε δίκτυα χρονοδιαγράμματος. Δίνεται ένα σύνολο συνδέσεων C , όπου κάθε σύνδεση c περιγράφεται από την πλειάδα $(p_dep, p_arr, \tau_dep, \tau_arr)$ που δηλώνει τη στάση αναχώρησης, τη στάση άφιξης και τις αντίστοιχες χρονικές στιγμές. Δίνονται επιπλέον μια στάση αφετηρίας s , μια στάση προορισμού z και μια στιγμή εκκίνησης t_0 .

Ζητείται μια ακολουθία συνδέσεων $\langle c_1, \dots, c_k \rangle$ που ικανοποιεί τρεις συνθήκες: η πρώτη αναχωρεί από την s όχι νωρίτερα από την t_0 , η τελευταία καταλήγει στην z , και κάθε διαδοχικό ζεύγος είναι συμβατό, δηλαδή η στάση άφιξης της μίας ταυτίζεται με τη στάση αναχώρησης της επόμενης και η χρονική τους διαδοχή είναι εφικτή. Μεταξύ όλων αυτών των ακολουθιών, βέλτιστη είναι εκείνη που ελαχιστοποιεί τη στιγμή άφιξης $\tau_arr(c_k)$.

Η αντικειμενική συνάρτηση αφορά, επομένως, αποκλειστικά τη στιγμή άφιξης και όχι τη διάρκεια του ταξιδιού — ουσιώδης διαφορά με πρακτικές συνέπειες, στην οποία επανερχόμαστε στην ενότητα 2.6.4.

2.5.2 Περιορισμοί ελάχιστου χρόνου ανταπόκρισης

Η συμβατότητα δύο διαδοχικών συνδέσεων απαιτεί ακριβέστερη διατύπωση. Όταν και οι δύο ανήκουν στο ίδιο δρομολόγιο, δηλαδή ο επιβάτης δεν αλλάζει όχημα, αρκεί η χρονική διαδοχή $\tau_arr(c_i) \leq \tau_dep(c_{i+1})$. Όταν όμως μεσολαβεί αλλαγή οχήματος, απαιτείται

$$\tau_dep(c_{i+1}) \geq \tau_arr(c_i) + \Delta(p),$$

όπου $\Delta(p)$ ο ελάχιστος χρόνος ανταπόκρισης στη στάση p .

Το μέγεθος αυτό εξαρτάται από τη διαρρύθμιση του σταθμού και σε ώριμα συστήματα καταγράφεται ανά στάση ή ακόμη και ανά ζεύγος αποβάθρων. Ελλείπει τέτοιας πληροφορίας — όπως στα ελληνικά δεδομένα — υιοθετείται σταθερή τιμή, η οποία λειτουργεί ως συντηρητικό περιθώριο ασφαλείας: υπερεκτιμά ελαφρώς τον απαιτούμενο χρόνο σε μικρούς σταθμούς, αποτρέποντας όμως την πρόταση ανταποκρίσεων που στην πράξη δεν προλαβαίνονται.

2.5.3 Επέκταση σε ορίζοντα πολλαπλών ημερών

Η κλασική διατύπωση του προβλήματος υποθέτει σιωπηρά ότι η λύση εντοπίζεται εντός ενός χρονικού ορίζοντα μίας ημέρας. Η υπόθεση αυτή είναι εύλογη σε πυκνά αστικά δίκτυα, όπου η υψηλή συχνότητα καθιστά σχεδόν βέβαιη την ύπαρξη σύνδεσης την ίδια ημέρα, καταρρέει όμως σε αραιά υπεραστικά δίκτυα.

Στο ελληνικό δίκτυο, μια διαδρομή με δύο ανταποκρίσεις μεταξύ απομακρυσμένων περιφερειών ενδέχεται να μην ολοκληρώνεται εντός της ίδιας ημέρας: το δρομολόγιο του δεύτερου σκέλους μπορεί να έχει ήδη αναχωρήσει ή να μην εκτελείται καθόλου τη συγκεκριμένη ημέρα. Η αναγκαστική διανυκτέρευση σε ενδιάμεσο κόμβο δεν αποτελεί, επομένως, παθολογία των δεδομένων αλλά νόμιμο συστατικό της λύσης, το οποίο πρέπει να μοντελοποιείται ρητά.

Η επέκταση απαιτεί δύο τροποποιήσεις: τη διεύρυνση του ορίζοντα ώστε να καλύπτει D διαδοχικές ημέρες, και την επέκταση της αναζήτησης της επόμενης έγκυρης αναχώρησης πέραν της τρέχουσας ημέρας, με δεδομένο ότι το σύνολο των εκτελούμενων δρομολογίων μεταβάλλεται ανά ημέρα της εβδομάδας.

2.5.4 Η έννοια του απόλυτου χρόνου ως ενοποιητική αναπαράσταση

Η επέκταση σε πολλαπλές ημέρες εισάγει ένα πρακτικό ζήτημα αναπαράστασης. Αν ο χρόνος εκφράζεται ως ώρα της ημέρας στο διάστημα $[0, 1440)$ λεπτών, τότε η σύγκριση δύο χρονικών στιγμών παύει να είναι απλή αριθμητική σύγκριση: η ώρα 06:00 της επόμενης ημέρας είναι μεταγενέστερη της 23:00 της τρέχουσας, παρότι η αριθμητική της τιμή είναι μικρότερη. Ο αλγόριθμος θα έπρεπε να συγκρίνει ζεύγη ημέρας και ώρας, με αντίστοιχο κίνδυνο σφαλμάτων.

Η λύση που υιοθετείται είναι η αναπαράσταση κάθε χρονικής στιγμής σε απόλυτα λεπτά από τη στιγμή εκκίνησης, μέσω της αντιστοίχισης

$$t_{\text{abs}} = 1440 \cdot d + t_{\text{tod}},$$

όπου d ο δείκτης της ημέρας σε σχέση με την ημέρα εκκίνησης και t_{tod} η ώρα της ημέρας σε λεπτά.

Η αναπαράσταση αυτή αποκαθιστά τη γραμμική διάταξη του χρόνου και επιτρέπει στον αλγόριθμο να λειτουργεί ομοιόμορφα, χωρίς καμία διάκριση ανάμεσα σε διαδρομές μίας και πολλών ημερών. Οι διανυκτερεύσεις παύουν να αποτελούν ειδική περίπτωση και προκύπτουν αυτομάτως ως αναμονές μεγάλης διάρκειας, ενώ ο δείκτης της ημέρας ανακτάται όποτε χρειάζεται —για τον έλεγχο εκτέλεσης ή την παρουσίαση— με ακέραια διαίρεση.

2.6 Πολυκριτηριακή δρομολόγηση

2.6.1 Κυριαρχία και βελτιστότητα κατά Pareto

Όταν η αξιολόγηση μιας διαδρομής γίνεται ως προς περισσότερα του ενός κριτήρια, η έννοια της βέλτιστης λύσης παύει να είναι μονοσήμαντη. Αν κάθε διαδρομή p χαρακτηρίζεται από ένα διάνυσμα κόστους $c(p) = (c_1(p), \dots, c_m(p))$, λέμε ότι η p κυριαρχεί της q όταν $c_i(p) \leq c_i(q)$ για κάθε

i και υπάρχει τουλάχιστον ένα j για το οποίο η ανισότητα είναι αυστηρή. Διαισθητικά, η p είναι τουλάχιστον εξίσου καλή σε όλα τα κριτήρια και αυστηρά καλύτερη σε ένα.

Μια διαδρομή καλείται βέλτιστη κατά Pareto όταν καμία άλλη δεν την κυριαρχεί. Το σύνολο των μη κυριαρχούμενων διαδρομών περιέχει όλες τις λύσεις που θα μπορούσαν εύλογα να επιλεγούν, ανάλογα με τις προτιμήσεις του χρήστη. Ζητούμενο σε ένα αυστηρά πολυκριτηριακό πρόβλημα είναι ο υπολογισμός ολόκληρου του συνόλου αυτού, το μέγεθος του οποίου μπορεί ωστόσο να αυξάνεται εκθετικά ως προς το πλήθος των κριτηρίων.

2.6.2 Τυπικά αντικρουόμενα κριτήρια

Στη δρομολόγηση με μέσα συγκοινωνιών τα κριτήρια που εξετάζονται συνηθέστερα είναι η στιγμή άφιξης στον προορισμό, το πλήθος των μετεπιβιβάσεων, ο συνολικός χρόνος αναμονής, η στιγμή αναχώρησης και το κόστος του εισιτηρίου. Ο ανταγωνισμός μεταξύ τους είναι δομικός και όχι συμπτωματικός: η μείωση των μετεπιβιβάσεων συνεπάγεται συχνά αναμονή για κάποια απευθείας σύνδεση και άρα μεταγενέστερη άφιξη, ενώ η πρωιμότερη δυνατή άφιξη επιτυγχάνεται τυπικά με περισσότερες αλλαγές οχήματος.

Στην παρούσα εργασία το κόστος εξαιρέθηκε των κριτηρίων για τους λόγους που εκτέθηκαν στην ενότητα 1.5. Προστίθεται όμως ένα κριτήριο ιδιαίτερης σημασίας για το ελληνικό δίκτυο, το πλήθος των αναγκαστικών διανυκτερεύσεων, το οποίο υπερισχύει όλων των υπολοίπων: μια διαδρομή που ολοκληρώνεται αυθημερόν είναι, από τη σκοπιά του επιβάτη, ποιοτικά διαφορετική από μια διαδρομή που απαιτεί διανυκτέρευση, ανεξαρτήτως των υπόλοιπων μεγεθών.

2.6.3 Λεξικογραφική διάταξη ως πρακτική προσέγγιση

Ο πλήρης υπολογισμός του συνόλου Pareto είναι υπολογιστικά δαπανηρός και, το σημαντικότερο, παράγει αποτέλεσμα δύσχρηστο για τον τελικό χρήστη, ο οποίος δεν επιθυμεί δεκάδες μη συγκρίσιμες εναλλακτικές αλλά λίγες σαφείς προτάσεις. Δύο πρακτικές προσεγγίσεις έχουν επικρατήσει.

Η πρώτη είναι η σταθμισμένη σύνθεση, όπου τα κριτήρια συνδυάζονται σε μια βαθμωτή συνάρτηση κόστους με συντελεστές βαρύτητας. Το πλεονέκτημά της είναι ότι επαναφέρει το πρόβλημα στη μονοκριτηριακή μορφή· το μειονέκτημα ότι οι συντελεστές είναι αυθαίρετοι και ότι ορισμένες βέλτιστες κατά Pareto λύσεις παραμένουν αθέατες για κάθε επιλογή τους.

Η δεύτερη είναι η λεξικογραφική διάταξη, όπου τα κριτήρια ιεραρχούνται και η σύγκριση δύο λύσεων γίνεται διαδοχικά: το δεύτερο κριτήριο εξετάζεται μόνο σε περίπτωση ισοπαλίας στο πρώτο και ούτω καθεξής. Η προσέγγιση αυτή δεν απαιτεί βαθμονόμηση παραμέτρων, παράγει ντετερμινιστική και πλήρη διάταξη και είναι ερμηνεύσιμη, καθώς η ιεραρχία μπορεί να διατυπωθεί σε φυσική γλώσσα. Η παρούσα εργασία υιοθετεί λεξικογραφική διάταξη με σειρά προτεραιότητας τις διανυκτερεύσεις, το πλήθος των σκελών και τη στιγμή άφιξης.

2.6.4 Ελαχιστοποίηση αναμονής και αναζήτηση τελευταίας αναχώρησης

Υπάρχει ένα κριτήριο που δεν εντάσσεται εύκολα στην παραπάνω διάταξη: ο χρόνος αναμονής. Ο αλγόριθμος πρωιμότερης άφιξης αδιαφορεί πλήρως για το πότε αναχωρεί ο επιβάτης, με

αποτέλεσμα να παράγει συστηματικά λύσεις που ξεκινούν όσο το δυνατόν νωρίτερα και αναμένουν στους ενδιάμεσους κόμβους. Σε πυκνά δίκτυα το φαινόμενο είναι αμελητέο· σε αραιά, όπου οι αναμονές μετρώνται σε ώρες, γίνεται εξόφθαλμο.

Η αντιμετώπιση περνά από το συμμετρικό πρόβλημα της τελευταίας αναχώρησης, το οποίο ζητά τη μέγιστη δυνατή στιγμή αναχώρησης που εξασφαλίζει άφιξη έως ένα δεδομένο χρονικό όριο. Επιλύεται με αναζήτηση επί του αντεστραμμένου γράφου, με τον χρόνο να διατρέχεται προς τα πίσω [13]. Ο συνδυασμός των δύο —πρωιμότερη άφιξη για τον προσδιορισμό του ορίου και τελευταία αναχώρηση για τη συμπίεση των αναμονών— παράγει διαδρομές που είναι ταυτόχρονα βέλτιστες ως προς την άφιξη και ελάχιστες ως προς τη συνολική διάρκεια. Η προσέγγιση αυτή είναι καθιερωμένη στη βιβλιογραφία: στα ερωτήματα εύρους (range ή profile queries) μία διαδρομή κυριαρχεί μιας άλλης όταν αναχωρεί αργότερα χωρίς να φτάνει αργότερα, οπότε η μεγιστοποίηση της ώρας αναχώρησης υπό σταθερή ώρα άφιξης αποτελεί ακριβώς ειδική περίπτωση της σχέσης κυριαρχίας [9, 10, 13]. Η ιδέα αυτή αποτελεί τη βάση του σταδίου οπισθοδρομικής βελτιστοποίησης της ενότητας 2.9.4.

2.7 Αλγόριθμοι δρομολόγησης σε δίκτυα συγκοινωνιών

2.7.1 Connection Scan Algorithm (CSA)

Ο αλγόριθμος σάρωσης συνδέσεων, που προτάθηκε από τους Dijkstra και συνεργάτες, εγκαταλείπει εντελώς την έννοια του γράφου [9]. Τα δεδομένα οργανώνονται σε έναν ενιαίο πίνακα όλων των στοιχειωδών συνδέσεων του δικτύου, ταξινομημένο κατά αύξουσα ώρα αναχώρησης. Η αναζήτηση συνίσταται σε μία και μόνη γραμμική σάρωση του πίνακα: για κάθε σύνδεση ελέγχεται αν είναι προσπελάσιμη με βάση τους τρέχοντες χρόνους άφιξης και, αν ναι, ενημερώνεται ο χρόνος άφιξης στη στάση προορισμού της.

Η απουσία ουράς προτεραιότητας και η αυστηρά σειριακή προσπέλαση της μνήμης καθιστούν τον αλγόριθμο εξαιρετικά ταχύ στην πράξη, παρά τη φαινομενικά αφελή γραμμική του φύση, καθώς εκμεταλλεύεται πλήρως τις κρυφές μνήμες του επεξεργαστή. Η υλοποίησή του είναι επιπλέον απλούστατη και επιδέχεται επεκτάσεις, όπως η παραλλαγή προφίλ που υπολογίζει με μία σάρωση τις βέλτιστες διαδρομές για ολόκληρο το εύρος των πιθανών ωρών αναχώρησης. Το μειονέκτημα είναι ότι κάθε ερώτημα σαρώνει καταρχήν ολόκληρο τον πίνακα συνδέσεων, γεγονός που καθιστά τον αλγόριθμο ευαίσθητο στο συνολικό μέγεθος του χρονοδιαγράμματος και όχι στην απόσταση αφετηρίας και προορισμού· η πρόωρη διακοπή της σάρωσης μόλις οριστικοποιηθεί ο προορισμός περιορίζει μεν το πρόβλημα, δεν το εξαλείφει όμως.

2.7.2 RAPTOR

Ο αλγόριθμος RAPTOR, από τους Dellinger, Pajor και Werneck, οργανώνει την αναζήτηση κατά γύρους, με τον k -οστό γύρο να υπολογίζει τους βέλτιστους χρόνους άφιξης για διαδρομές που χρησιμοποιούν το πολύ k οχήματα [10]. Σε κάθε γύρο σαρώνονται οι γραμμές που διέρχονται από τις στάσεις οι οποίες βελτιώθηκαν στον προηγούμενο, και για καθεμία εντοπίζεται το πρωιμότερο όχημα που μπορεί να ληφθεί.

Η προσέγγιση δεν χρησιμοποιεί ούτε αυτή ουρά προτεραιότητας και έχει δύο σημαντικά πλεονεκτήματα. Πρώτον, παράγει εγγενώς το πλήρες σύνολο Pareto ως προς τα δύο κριτήρια της άφιξης και του πλήθους μετεπιβιβάσεων, αφού κάθε γύρος αντιστοιχεί σε διαφορετική τιμή του δεύτερου κριτηρίου. Δεύτερον, οι γύροι παραλληλοποιούνται εύκολα. Το πλήθος των γύρων φράσσεται στην πράξη από έναν μικρό αριθμό, καθώς οι ρεαλιστικές διαδρομές σπάνια απαιτούν περισσότερες από τέσσερις ή πέντε μετεπιβιβάσεις. Ο αλγόριθμος ενσωματώνει επίσης τις πεζές μεταβάσεις μεταξύ γειτονικών στάσεων ως χωριστό στάδιο κάθε γύρου, ενώ οι παραλλαγές του καλύπτουν τόσο ερωτήματα προφίλ για ολόκληρο χρονικό διάστημα όσο και πρόσθετα κριτήρια βελτιστοποίησης πέραν των δύο βασικών.

2.7.3 Transfer Patterns

Η προσέγγιση των προτύπων μετεπιβίβασης, που αναπτύχθηκε από τους Bast και συνεργάτες, στηρίζεται σε εκτεταμένη προεπεξεργασία [11]. Η κεντρική παρατήρηση είναι ότι, για ένα δεδομένο ζεύγος στάσεων, το σύνολο των βέλτιστων διαδρομών σε όλη τη διάρκεια της ημέρας χρησιμοποιεί ελάχιστες μόνο διαφορετικές ακολουθίες σταθμών μετεπιβίβασης. Αν οι ακολουθίες αυτές υπολογιστούν και αποθηκευτούν εκ των προτέρων, η απάντηση σε ένα ερώτημα ανάγεται σε αναζήτηση εντός ενός ελάχιστου βοηθητικού γράφου.

Το αποτέλεσμα είναι χρόνοι απόκρισης της τάξης του χιλιοστού του δευτερολέπτου, ικανοί να εξυπηρετήσουν υπηρεσίες παγκόσμιας κλίμακας. Το τίμημα είναι η προεπεξεργασία, η οποία απαιτεί ώρες υπολογισμού σε κατανομημένες υποδομές, και η ακαμψία: κάθε μεταβολή του χρονοδιαγράμματος επιβάλλει επανάληψή της, γεγονός που καθιστά την προσέγγιση δυσχερή για δεδομένα που ενημερώνονται συχνά. Μεταγενέστερες παραλλαγές περιορίζουν το κόστος αυτό μέσω κόμβων-διαμετακομιστών, οι οποίοι επιτρέπουν τον υπολογισμό τοπικών μόνο προτύπων και τη σύνθεσή τους κατά τον χρόνο του ερωτήματος.

2.7.4 Συγκριτική αξιολόγηση και κριτήρια επιλογής

Η σύγκριση των τριών προσεγγίσεων αναδεικνύει μια σαφή ανταλλαγή μεταξύ κόστους προεπεξεργασίας και χρόνου απόκρισης. Ο CSA δεν απαιτεί ουσιαστικά προεπεξεργασία πέραν της ταξινόμησης, ο RAPTOR μια απλή οργάνωση των γραμμών ανά στάση, ενώ τα Transfer Patterns επενδύουν δυσανάλογα στην προεπεξεργασία για να ελαχιστοποιήσουν τον χρόνο ερωτήματος. Κοινό γνώρισμα και των τριών είναι η εγκατάλειψη της ουράς προτεραιότητας, η οποία αποδεικνύεται περιττή όταν η δομή του χρονοδιαγράμματος επιβάλλει ήδη μια φυσική σειρά επεξεργασίας.

Το κρίσιμο ερώτημα για την παρούσα εργασία είναι ποιο από τα μεγέθη αυτά αποτελεί πραγματικό περιορισμό. Οι αλγόριθμοι αυτοί σχεδιάστηκαν για δίκτυα με εκατομμύρια συνδέσεις που εξυπηρετούν εκατομμύρια ερωτήματα ημερησίως, όπου η βελτίωση κατά μία τάξη μεγέθους στον χρόνο απόκρισης έχει άμεσο οικονομικό αντίκρισμα. Το σύνολο δεδομένων της παρούσας εργασίας περιλαμβάνει 106.380 δρομολόγια και 460 στάσεις με αναχωρήσεις, μεγέθη μικρότερα κατά δύο τουλάχιστον τάξεις. Υπό αυτές τις συνθήκες, ένας ορθά υλοποιημένος χρονοεξαρτώμενος Dijkstra αποκρίνεται ήδη σε χρόνο μη αντιληπτό από τον χρήστη, και η πολυπλοκότητα των εξειδικευμένων προσεγγίσεων δεν αντισταθμίζεται από κάποιο μετρήσιμο

όφελος. Η επιλογή, επομένως, υπαγορεύεται από τη διαφάνεια και την ευκολία επέκτασης προς τα ιδιαίτερα χαρακτηριστικά του προβλήματος και όχι από την ταχύτητα.

2.8 Ενοποίηση ετερογενών πηγών δεδομένων

2.8.1 Το πρόβλημα ταύτισης οντοτήτων (entity resolution)

Η ταύτιση οντοτήτων είναι το πρόβλημα του εντοπισμού εκείνων των εγγραφών που, ενώ προέρχονται από διαφορετικές πηγές και διαφέρουν ως προς τη μορφή τους, αναφέρονται στην ίδια οντότητα του πραγματικού κόσμου. Πρόκειται για κλασικό ζήτημα της ολοκλήρωσης δεδομένων, το οποίο εμφανίζεται σε κάθε προσπάθεια συνένωσης ανεξάρτητα τηρούμενων καταλόγων [12].

Στην περίπτωση των συγκοινωνιακών δεδομένων το πρόβλημα λαμβάνει διπλή μορφή. Το πρόβλημα των συνωνύμων αφορά την ίδια φυσική στάση που εμφανίζεται με διαφορετικές γραφές σε διαφορετικές πηγές, λόγω τονισμού, κεφαλαιοποίησης ή προσδιοριστικών όπως «ΚΤΕΛ» και «ΣΤΑΘΜΟΣ». Το πρόβλημα των ομωνύμων είναι το αντίστροφο και δυσκολότερο: διαφορετικές φυσικές στάσεις φέρουν πανομοιότυπο όνομα, όπως συμβαίνει με τοπωνύμια που απαντούν σε πολλούς νομούς. Η αστοχία στην πρώτη περίπτωση διασπά το δίκτυο σε ασύνδετα τμήματα και αποκρύπτει υπαρκτές ανταποκρίσεις· η αστοχία στη δεύτερη είναι σοβαρότερη, καθώς κατασκευάζει ανύπαρκτες συνδέσεις και παράγει διαδρομές αδύνατες.

2.8.2 Μέτρα ομοιότητας και χωρική ομαδοποίηση

Η λεξιλογική σύγκριση συμβολοσειρών στηρίζεται σε καθιερωμένα μέτρα, με συνηθέστερη την απόσταση επεξεργασίας κατά Levenshtein, η οποία μετρά το ελάχιστο πλήθος στοιχειωδών μετασχηματισμών χαρακτήρων που μετατρέπουν τη μία συμβολοσειρά στην άλλη. Απαραίτητο προηγούμενο βήμα είναι η κανονικοποίηση, δηλαδή η αφαίρεση διακριτικών σημείων μέσω αποσύνθεσης Unicode, η ενοποίηση της πεζότητας και η απαλοιφή των τυποποιημένων προσδιοριστικών.

Επειδή η σύγκριση όλων των δυνατών ζευγών είναι τετραγωνικής πολυπλοκότητας ως προς το πλήθος των καταχωρίσεων, προηγείται συνήθως ένα στάδιο ομαδοποίησης, το οποίο περιορίζει τις συγκρίσεις εντός υποψήφιων ομάδων που μοιράζονται κάποιο φθηνά υπολογίσιμο γνώρισμα, όπως τα αρχικά γράμματα του κανονικοποιημένου ονόματος ή τον φορέα προέλευσης.

Η λεξιλογική ομοιότητα από μόνη της, ωστόσο, δεν επαρκεί, ακριβώς λόγω των ομωνύμων: δύο πανομοιότυπες συμβολοσειρές έχουν μηδενική απόσταση ανεξαρτήτως του αν αναφέρονται στο ίδιο μέρος. Απαιτείται, επομένως, ένα ανεξάρτητο κριτήριο, το οποίο προσφέρει η γεωγραφική θέση: αν οι υποψήφιες εγγραφές γεωκωδικοποιηθούν και οι προκύπτουσες συντεταγμένες απέχουν πέραν ενός κατωφλίου, η ταύτιση απορρίπτεται ακόμη και επί απόλυτης λεξιλογικής ταυτότητας. Αντιστρόφως, η χωρική ομαδοποίηση σημείων που απέχουν ελάχιστα επιτρέπει τη συγχώνευση στάσεων με ανόμοια ονόματα οι οποίες εξυπηρετούν το ίδιο σημείο.

2.8.3 Αμφισημία τοπωνυμίων και ο ρόλος του γεωγραφικού συμφραζομένου

Η γεωκωδικοποίηση μετατοπίζει, χωρίς να εξαλείφει, το πρόβλημα: η υποβολή ενός αμφίσημου τοπωνυμίου σε μια υπηρεσία γεωκωδικοποίησης επιστρέφει την πιθανότερη ερμηνεία, η οποία δεν είναι κατ' ανάγκη η ορθή. Η άρση της αμφισημίας απαιτεί συμφραζόμενα, και εδώ η ίδια η δομή των δεδομένων προσφέρει μια ισχυρή ένδειξη.

Καθώς κάθε καταχώριση προέρχεται από συγκεκριμένο φορέα με σαφώς οριοθετημένη γεωγραφική περιοχή ευθύνης, η ταυτότητα του φορέα λειτουργεί ως γεωγραφικό συμφραζόμενο που περιορίζει δραστικά τον χώρο των πιθανών ερμηνειών. Η αντιστοίχιση παύει έτσι να γίνεται με κλειδί το όνομα και γίνεται με κλειδί την τριάδα πηγής, φορέα και ονόματος, ώστε δύο ομώνυμες στάσεις διαφορετικών φορέων να μην συγχέονται ποτέ. Η προσέγγιση αυτή, που περιγράφεται αναλυτικά στο τέταρτο κεφάλαιο, αποδείχθηκε το αποφασιστικό στοιχείο για την ορθότητα του τελικού συνόλου δεδομένων.

2.9 Ο προτεινόμενος αλγόριθμος

2.9.1 Τυπική διατύπωση του προβλήματος της εργασίας

Έστω P το σύνολο των κανονικοποιημένων στάσεων και R το σύνολο των δρομολογίων, όπου κάθε δρομολόγιο $r \in R$ περιγράφεται από την πλειάδα $(p(r), q(r), \tau(r), c(r), D(r), \kappa(r))$: στάση αναχώρησης, στάση άφιξης, ώρα αναχώρησης ως λεπτά της ημέρας, διάρκεια διαδρομής, υποσύνολο $D(r) \subseteq \{0, \dots, 6\}$ των ημερών της εβδομάδας κατά τις οποίες εκτελείται, και αναγνωριστικό φορέα.

Δίνονται αφετηρία s , προορισμός z , ημέρα εκκίνησης d_0 και ώρα εκκίνησης t_0 . Ζητείται ακολουθία σκελών $\square \ell_1, \dots, \ell_k \square$ με $\ell_i \square R$, τέτοια ώστε $p(\ell_1) = s$, $q(\ell_k) = z$, και $q(\ell_i) = p(\ell_{i+1})$ για κάθε i . Κάθε σκέλος εκτελείται σε συγκεκριμένη ημέρα d_i , με τον απόλυτο χρόνο αναχώρησης να ορίζεται ως $T_{\text{dep}}(i) = 1440 \cdot d_i + \tau(\ell_i)$ και τον απόλυτο χρόνο άφιξης ως $T_{\text{arr}}(i) = T_{\text{dep}}(i) + c(\ell_i)$.

Οι περιορισμοί είναι τέσσερις: εγκυρότητα εκτέλεσης, $(d_0 + d_i) \bmod 7 \square D(\ell_i)$: ελάχιστος χρόνος ανταπόκρισης, $T_{\text{dep}}(i+1) \geq T_{\text{arr}}(i) + \Delta$: μέγιστη αναμονή αυθημερόν, $T_{\text{dep}}(i+1) - T_{\text{arr}}(i) \leq W$ όταν τα δύο σκέλη εκτελούνται την ίδια ημέρα: και άνω φράγμα σκελών, $k \leq K$. Ζητούμενο είναι η ελαχιστοποίηση του $T_{\text{arr}}(k)$ και, δευτερευόντως, του πλήθους των διανυκτερεύσεων και των σκελών.

2.9.2 Το επιλεγμένο μοντέλο και η αιτιολόγησή του

Επιλέχθηκε το χρονοεξαρτώμενο μοντέλο της ενότητας 2.4.3, με τους κόμβους να αντιστοιχούν σε κανονικοποιημένες στάσεις και τις εξερχόμενες αναχωρήσεις κάθε στάσης να αποθηκεύονται σε λίστα ταξινομημένη κατά ώρα.

Τρεις λόγοι υπαγόρευαν την επιλογή. Πρώτον, η ανεξαρτησία του μεγέθους του γράφου από τον χρονικό ορίζοντα: εφόσον η υποστήριξη πολλαπλών ημερών αποτελεί λειτουργική απαίτηση, το χρονικά αναπτυγμένο μοντέλο θα απαιτούσε επτά αντίγραφα της δομής για την κάλυψη μιας πλήρους εβδομάδας. Δεύτερον, η φυσική ενσωμάτωση του εβδομαδιαίου προτύπου εκτέλεσης: το

σύνολο $D(r)$ ελέγχεται κατά την αποτίμηση της συνάρτησης άφιξης, ενώ στο εναλλακτικό μοντέλο θα έπρεπε να αποτυπωθεί στην ίδια τη δομή. Τρίτον, το βασικό μειονέκτημα του μοντέλου —η δυσκολία ενσωμάτωσης λεπτομερών περιορισμών ανταπόκρισης— δεν ισχύει εδώ, διότι τα διαθέσιμα δεδομένα δεν περιλαμβάνουν καθόλου πληροφορία ανά αποβάθρα και επιβάλλουν έτσι ούτως ή άλλως τη χρήση ενιαίας τιμής Δ .

2.9.3 Περιγραφή και ψευδοκώδικας

Ο αλγόριθμος αποτελεί χρονοεξαρτώμενο Dijkstra με εμπλουτισμένη κατάσταση. Η κατάσταση της αναζήτησης δεν είναι η στάση αλλά το ζεύγος στάσης και πλήθους σκελών, ώστε να διατηρούνται και διαδρομές που φτάνουν αργότερα με λιγότερες μετεπιβιβάσεις. Η ουρά προτεραιότητας διατάσσεται κατά απόλυτο χρόνο άφιξης.

```
FindRoutes(G, s, z, d0, t0):
  Q ← {(t0, 0, s, {})}           # (απόλυτος χρόνος, πλήθος σκελών, στάση, μονοπάτι)
  best ← ∅ ; results ← {}
  όσο Q ≠ ∅:
    (t, k, u, path) ← EXTRACT-MIN(Q)
    αν best[(u,k)] ≤ t: συνέχισε           # παρωχημένη καταχώριση
    best[(u,k)] ← t ; αν k ≥ K: συνέχισε
    για κάθε r ∈ G[u]:
      # ταξινομημένα κατά ώρα αναχώρησης
      earliest ← t + (Δ αν path ≠ {} αλλιώς 0)
      βρες ελάχιστη ημέρα δ με 1440·δ + τr ≥ earliest και (d0+δ) mod 7 ∈ Dr
      αν δεν υπάρχει: συνέχισε
      Tdep ← 1440·δ + τr ; wait ← Tdep - t
      αν path ≠ {} και ίδια ημέρα και wait > W: συνέχισε
      Tarr ← Tdep + cr ; path' ← path ⊕ leg(r, Tdep, Tarr)
      αν qr = z: results ← results ⊕ path'
      αλλιώς: INSERT(Q, (Tarr, k+1, qr, path'))
  results ← [TightenPath(p) για p ∈ results]
  ταξινόμησε κατά (διανυκτερεύσεις, πλήθος σκελών, χρόνο άφιξης)· απάλειψε διπλότυπα
  επέστρεψε τα πρώτα M
```

Η αναζήτηση δεν τερματίζεται στην πρώτη άφιξη στον προορισμό αλλά συνεχίζεται έως τη συλλογή επαρκούς πλήθους υποψηφίων, ώστε να προταθούν στον χρήστη εναλλακτικές με διαφορετικά χαρακτηριστικά. Οι διανυκτερεύσεις δεν αντιμετωπίζονται ως ειδική περίπτωση: προκύπτουν αυτομάτως όταν η πρωιμότερη έγκυρη επιβίβαση εντοπίζεται σε επόμενη ημέρα, οπότε το όριο μέγιστης αναμονής W δεν εφαρμόζεται.

2.9.4 Το στάδιο οπισθοδρομικής βελτιστοποίησης

Αφορμή για το στάδιο αυτό αποτελεί η αδυναμία που επισημάνθηκε στην ενότητα 2.6.4: ο αλγόριθμος πρωιμότερης άφιξης επιλέγει πάντοτε το πρωιμότερο δρομολόγιο κάθε σκέλους, ακόμη και όταν αυτό οδηγεί σε πολύωρη αναμονή στον επόμενο κόμβο χωρίς κανένα όφελος στην τελική άφιξη.

Δεδομένης μιας βέλτιστης διαδρομής $\square l_1, \dots, l_k \square$, το στάδιο διατρέχει τα σκέλη από το προτελευταίο προς το πρώτο. Για κάθε σκέλος l_i ορίζεται η προθεσμία $deadline(i) = T_dep(i+1) -$

Δ και αναζητείται, μεταξύ όλων των δρομολογίων που συνδέουν τις ίδιες δύο στάσεις την ίδια ημέρα, εκείνο με τη μέγιστη ώρα αναχώρησης του οποίου η άφιξη δεν υπερβαίνει την προθεσμία. Για το πρώτο σκέλος τίθεται επιπλέον κάτω φράγμα η ώρα εκκίνησης t_0 .

Δύο ιδιότητες του σταδίου αξίζει να τονιστούν. Πρώτον, δεν μεταβάλλει τον χρόνο τελικής άφιξης: το τελευταίο σκέλος παραμένει αμετάβλητο και κάθε αντικατάσταση διατηρεί εξ ορισμού την εφικτότητα της επόμενης ανταπόκρισης. Η λύση παραμένει, επομένως, βέλτιστη ως προς το αρχικό κριτήριο, ενώ ταυτόχρονα ελαχιστοποιείται η συνολική διάρκεια για τη δεδομένη ακολουθία σκελών. Δεύτερον, το σφίξιμο δεν εφαρμόζεται πάνω από διανυκτέρευση. Η επιλογή αυτή είναι συνειδητή και όχι τεχνικός περιορισμός: όταν μεσολαβεί διανυκτέρευση, η πρωιμότερη άφιξη στον ενδιαμέσο κόμβο είναι προτιμότερη, καθώς αφήνει στον επιβάτη περιθώριο για τη διευθέτηση της διαμονής του. Αντ' αυτού, το σύστημα παρουσιάζει ρητά το σύνολο των μεταγενέστερων αναχωρήσεων του ίδιου σκέλους, μεταθέτοντας την απόφαση στον χρήστη.

Οφείλει, τέλος, να προσδιοριστεί η θέση του σταδίου ως προς τη βιβλιογραφία. Το κριτήριο που βελτιστοποιεί —μέγιστη ώρα αναχώρησης υπό δεδομένη ώρα άφιξης— δεν είναι νέο: ταυτίζεται με τη σχέση κυριαρχίας των ερωτημάτων εύρους, στα οποία μία διαδρομή θεωρείται καλύτερη όταν αναχωρεί αργότερα χωρίς να φτάνει αργότερα, και επιλύεται στη γενική του μορφή με αναζήτηση τελευταίας αναχώρησης επί του αντεστραμμένου δικτύου [9, 10, 13]. Οι αλγόριθμοι αυτοί υπολογίζουν, ωστόσο, ολόκληρο το σύνολο Pareto των ζευγών αναχώρησης και άφιξης, με δεύτερη διάσχιση του γράφου και κόστος δυσανάλογο προς το ζητούμενο εδώ. Το στάδιο που περιγράφηκε δεν διεκδικεί, συνεπώς, αλγοριθμική πρωτοτυπία: αποτελεί προσαρμογή γνωστής αρχής σε μορφή μετα-επεξεργασίας μίας ήδη υπολογισμένης διαδρομής, με κόστος γραμμικό ως προς το πλήθος των σκελών και χωρίς επανάληψη της αναζήτησης. Ο χαρακτήρας του είναι, αντιστοίχως, ευρετικός ως προς το σύνολο των εναλλακτικών, καθώς η ακολουθία των κόμβων παραμένει δεσμευμένη από την αρχική λύση· διατηρεί όμως αποδεδειγμένα την ώρα άφιξης και αποδίδει, όπως δείχνει η ενότητα 6.8, μετρήσιμο όφελος στο συγκεκριμένο δίκτυο. Στοιχείο που δεν εντοπίστηκε υπό αυτή τη μορφή στη βιβλιογραφία είναι η συνειδητή εξαίρεση των διανυκτερεύσεων από το σφίξιμο και η παρουσίαση, στη θέση του, των μεταγενέστερων αναχωρήσεων ως ρητών εναλλακτικών — επιλογή που πηγάζει από τις συνθήκες του υπεραστικού δικτύου και όχι από αλγοριθμική ανάγκη.

2.9.5 Ορθότητα, τερματισμός και πολυπλοκότητα

Η ορθότητα ως προς την πρωιμότερη άφιξη προκύπτει από την απόδειξη της ενότητας 2.3.4, εφόσον επαληθευτούν οι προϋποθέσεις της. Οι διάρκειες είναι μη αρνητικές, η δε ιδιότητα FIFO ισχύει διότι η επιλογή της πρωιμότερης έγκυρης επιβίβασης πάνω σε ταξινομημένο κατάλογο υλοποιεί ακριβώς τη συνάρτηση ελαχίστου της ενότητας 2.3.1. Η επέκταση της κατάστασης με το πλήθος σκελών δεν θίγει την ορθότητα, καθώς ο έλεγχος κυριαρχίας γίνεται εντός κάθε κλάσης (u, k) χωριστά, ενώ η αναπαράσταση σε απόλυτα λεπτά διατηρεί τη γραμμική διάταξη του χρόνου και πέραν των μεσανυχτίων.

Πρέπει, ωστόσο, να διευκρινιστεί η εμβέλεια της εγγύησης: η βελτιστότητα κατοχυρώνεται για την πρωιμότερη άφιξη εντός κάθε κλάσης πλήθους σκελών, όχι για το σύνολο των εναλλακτικών που τελικά παρουσιάζονται. Οι υπόλοιπες συλλέγονται ως έγκυρες μεν αλλά όχι κατ' ανάγκη βέλτιστες προτάσεις, με μοναδικό σκοπό να προσφέρουν στον χρήστη εύλογο εύρος επιλογών.

Ο τερματισμός εξασφαλίζεται από τρεις ανεξάρτητους μηχανισμούς: ο απόλυτος χρόνος αυξάνει γνησίως κατά μήκος κάθε μονοπατιού, το πλήθος σκελών φράσσεται από K , και ο ορίζοντας ημερών από D . Ο χώρος καταστάσεων είναι, επομένως, πεπερασμένος και κάθε κατάσταση οριστικοποιείται το πολύ μία φορά.

Ως προς την πολυπλοκότητα, ο χώρος καταστάσεων έχει μέγεθος $O(|P| \cdot K)$ και κάθε εξαγωγή κοστίζει $O(\log(|R| \cdot K))$. Η αποτίμηση της συνάρτησης άφιξης απαιτεί σάρωση των αναχωρήσεων μιας στάσης επί το πολύ $D+1$ ημέρες, οπότε ο συνολικός χρόνος φράσσεται από $O(|R| \cdot K \cdot D \cdot \log(|R| \cdot K))$. Το στάδιο οπισθοδρομικής βελτιστοποίησης προσθέτει $O(k \cdot \text{deg_max})$ ανά διαδρομή, μέγεθος αμελητέο έναντι της κύριας αναζήτησης.

2.9.6 Παραδοχές και θεωρητικοί περιορισμοί

Η ορθότητα του αλγορίθμου τελεί υπό συγκεκριμένες παραδοχές, οι οποίες οφείλουν να δηλωθούν ρητά. Η πρώτη είναι ο ενιαίος χρόνος ανταπόκρισης: υιοθετείται μία σταθερή τιμή για κάθε στάση, ελλείψει δεδομένων ανά σταθμό, με συνέπεια την ελαφρά υπερεκτίμηση σε μικρές στάσεις και ενδεχόμενη υποεκτίμηση σε μεγάλους σταθμούς με χωριστά κτίρια αναχωρήσεων.

Η δεύτερη είναι η ακρίβεια εκτέλεσης: η αναζήτηση στηρίζεται αποκλειστικά στο προγραμματισμένο χρονοδιάγραμμα και αγνοεί καθυστερήσεις και ματαιώσεις, καθώς δεδομένα πραγματικού χρόνου δεν δημοσιεύονται από καμία πηγή. Η τρίτη είναι η πληρότητα των διαρκειών: όπου η διάρκεια διαδρομής λείπει από την πηγή, το σκέλος δεν μπορεί να χρησιμοποιηθεί ως ενδιάμεσο, διότι ο χρόνος άφιξης παραμένει άγνωστος και η αναζήτηση δεν μπορεί να συνεχιστεί από αυτόν τον κόμβο.

Η τέταρτη είναι το άνω φράγμα σκελών και ο ορίζοντας ημερών, τα οποία περιορίζουν τον χώρο αναζήτησης εξ ορισμού. Πρόκειται για συνειδητή αποκοπή και όχι για ατέλεια: διαδρομές με περισσότερες από τρεις μετεπιβιβάσεις είναι πρακτικά μη χρησιμοποιήσιμες, ενώ η ελαχιστοποίηση των διανυκτερεύσεων προϋποθέτει ούτως ή άλλως περιορισμένο ορίζοντα. Ο αλγόριθμος είναι, επομένως, πλήρης εντός των ορίων που ο ίδιος θέτει, χωρίς να διεκδικεί καθολικότητα πέραν αυτών.

2.10 Σχετικές εργασίες

Το πεδίο της δρομολόγησης σε δίκτυα συγκοινωνιών έχει γνωρίσει έντονη ερευνητική δραστηριότητα τις τελευταίες δύο δεκαετίες. Η περιεκτικότερη επισκόπηση παραμένει εκείνη των Bast και συνεργατών, η οποία συστηματοποιεί το σύνολο των τεχνικών από τον κλασικό Dijkstra έως τις σύγχρονες μεθόδους προεπεξεργασίας και αποτελεί σημείο αναφοράς για το πεδίο [13]. Στα θεμέλια της μοντελοποίησης βρίσκεται η εργασία των Pyrga και συνεργατών, η οποία τυποποίησε τη διάκριση των δύο μοντέλων και προσέφερε την πρώτη συστηματική πειραματική σύγκρισή τους [8].

Η πολυκριτηριακή διάσταση αναπτύχθηκε από τους Müller-Hannemann και Schnee, οι οποίοι μελέτησαν τον υπολογισμό συνόλων Pareto σε ρεαλιστικά σιδηροδρομικά δίκτυα, ενώ ο Dissert και οι συνεργάτες του διερεύνησαν τεχνικές επιτάχυνσης της πολυκριτηριακής αναζήτησης [14, 15]. Οι σύγχρονοι αλγόριθμοι που παρουσιάστηκαν στην ενότητα 2.7 —RAPTOR, CSA και

Transfer Patterns— συνιστούν το κυρίαρχο σήμερα τεχνολογικό υπόβαθρο των εμπορικών συστημάτων.

Σε ό,τι αφορά την ενοποίηση ετερογενών πηγών, η βιβλιογραφία της ταύτισης οντοτήτων είναι εκτενής και προηγείται του συγκοινωνιακού πεδίου, με κλασική αναφορά την επισκόπηση των Elmagarmid και συνεργατών [12]. Η εφαρμογή της ειδικά σε δεδομένα στάσεων έχει απασχολήσει κυρίως την κοινότητα του OpenStreetMap και τους φορείς που συγκροτούν εθνικά μητρώα, με τη σχετική εμπειρία να καταγράφεται συνήθως σε τεχνικές αναφορές παρά σε ακαδημαϊκές δημοσιεύσεις.

Ένα δεύτερο ρεύμα εργασιών, λιγότερο αλγοριθμικό αλλά εξίσου κρίσιμο για την πράξη, αφορά την τυποποίηση της δημοσίευσης των δεδομένων. Η καθιέρωση του προτύπου GTFS και η μετέπειτα ευρωπαϊκή τυποποίηση μέσω του NeTEx έχουν αποτελέσει αντικείμενο εκτενούς τεχνικής βιβλιογραφίας, ενώ πλήθος μελετών αξιοποιεί ανοικτά δεδομένα χρονοδιαγράμματος για την ανάλυση της προσβασιμότητας περιοχών και πληθυσμιακών ομάδων [16, 17]. Χαρακτηριστικό της βιβλιογραφίας αυτής είναι ότι θεωρεί τη διαθεσιμότητα των δεδομένων δεδομένη και εστιάζει στην αξιοποίησή τους — υπόθεση που, όπως τεκμηριώθηκε στο πρώτο κεφάλαιο, δεν ισχύει για το ελληνικό υπεραστικό δίκτυο και αντιστρέφει τη σειρά προτεραιοτήτων της παρούσας εργασίας.

Για τον ελληνικό χώρο, η βιβλιογραφία είναι περιορισμένη και επικεντρώνεται σχεδόν αποκλειστικά στις αστικές συγκοινωνίες των δύο μεγάλων κέντρων, όπου η ύπαρξη δεδομένων GTFS έχει επιτρέψει την ανάπτυξη εφαρμογών και μελετών προσβασιμότητας. Δεν εντοπίστηκε εργασία που να αντιμετωπίζει το υπεραστικό δίκτυο ΚΤΕΛ ως ενιαίο σύστημα ούτε δημοσιευμένο σύνολο δεδομένων που να το αποτυπώνει συνολικά. Το κενό αυτό, το οποίο επιβεβαιώνει τη σκοπιμότητα της παρούσας εργασίας, οφείλεται λιγότερο σε έλλειψη ερευνητικού ενδιαφέροντος και περισσότερο στο πρωτογενές εμπόδιο της απουσίας προσβάσιμων δεδομένων — εμπόδιο του οποίου η άρση αποτελεί, όπως περιγράφεται στο επόμενο κεφάλαιο, προϋπόθεση κάθε αλγοριθμικής επεξεργασίας.

Επίλογος κεφαλαίου

Το κεφάλαιο αυτό συγκρότησε το θεωρητικό πλαίσιο της εργασίας ακολουθώντας μια πορεία σταδιακής εξειδίκευσης. Από τους βασικούς ορισμούς της θεωρίας γράφων και τον αλγόριθμο Dijkstra, του οποίου η ορθότητα αποδείχθηκε αναλυτικά, η ανάπτυξη οδηγήθηκε στα χρονοεξαρτώμενα δίκτυα και στην ιδιότητα FIFO, η οποία αναδείχθηκε ως η ακριβής συνθήκη υπό την οποία η άπληστη στρατηγική διατηρεί την ισχύ της παρουσία χρονικής μεταβλητότητας.

Στη συνέχεια εξετάστηκαν τα δύο καθιερωμένα μοντέλα αναπαράστασης δικτύων χρονοδιαγράμματος και αιτιολογήθηκε η επιλογή του χρονοεξαρτώμενου, το πρόβλημα πρωιμότερης άφιξης επεκτάθηκε σε ορίζοντα πολλαπλών ημερών μέσω της αναπαράστασης του χρόνου σε απόλυτα λεπτά, και η πολυκριτηριακή διάσταση αντιμετωπίστηκε με λεξικογραφική διάταξη. Η επισκόπηση των σύγχρονων αλγορίθμων του πεδίου τεκμηρίωσε γιατί η πολυπλοκότητά τους δεν δικαιολογείται στην κλίμακα του συγκεκριμένου προβλήματος.

Το κεφάλαιο ολοκληρώθηκε με την τυπική διατύπωση του προτεινόμενου αλγορίθμου, τη θεμελίωση της ορθότητας και της πολυπλοκότητάς του, την περιγραφή του σταδίου

οπισθοδρομικής βελτιστοποίησης και τον προσδιορισμό της θέσης του ως προς τη σχετική βιβλιογραφία, και τη ρητή δήλωση των παραδοχών και των ορίων εντός των οποίων ισχύουν οι εγγυήσεις του. Η θεωρητική αυτή κατασκευή προϋποθέτει, ωστόσο, την ύπαρξη ενός συνεπούς συνόλου δεδομένων — προϋπόθεση που, στην περίπτωση του ελληνικού υπεραστικού δικτύου, απέχει πολύ από το να θεωρείται δεδομένη και αποτελεί το αντικείμενο των δύο επόμενων κεφαλαίων.

Κεφάλαιο 3

Γλώσσες και Τεχνολογίες

Το προηγούμενο κεφάλαιο συγκρότησε το θεωρητικό πλαίσιο της εργασίας και κατέληξε στη διαπίστωση ότι η αλγοριθμική επίλυση προϋποθέτει ένα συνεπές σύνολο δεδομένων, το οποίο δεν υπάρχει και πρέπει να κατασκευαστεί. Η κατασκευή αυτή, μαζί με την υλοποίηση του υπολογιστικού πυρήνα και της διεπαφής χρήστη, απαιτήσε ένα σύνολο γλωσσών, βιβλιοθηκών, προτύπων και εξωτερικών υπηρεσιών. Το παρόν κεφάλαιο τις παρουσιάζει, όχι ως απλή απαρίθμηση αλλά ως σειρά τεκμηριωμένων επιλογών: για κάθε τεχνολογία εξηγείται τι πρόβλημα κλήθηκε να λύσει, γιατί προτιμήθηκε έναντι των εναλλακτικών και ποια όρια θέτει η χρήση της. Η σειρά παρουσίασης ακολουθεί τη ροή των δεδομένων μέσα στο σύστημα, από την ανάκτηση και την αποθήκευση έως τον υπολογισμό και την παρουσίαση, ώστε να προετοιμάσει την αναλυτική περιγραφή της υλοποίησης στα κεφάλαια 4 και 5.

3.1 Python και το οικοσύστημα βιβλιοθηκών

Ως κύρια γλώσσα υλοποίησης επιλέχθηκε η Python, στην έκδοση 3.10 ή νεότερη, και χρησιμοποιήθηκε σε ολόκληρη την αλυσίδα επεξεργασίας: στους μηχανισμούς συλλογής, στη φόρτωση της βάσης, στην κατασκευή του γράφου, στον αλγόριθμο αναζήτησης και στον εξυπηρετητή της διαδικτυακής εφαρμογής. Η ομοιογένεια αυτή δεν ήταν αυτονόητη — θα ήταν εξίσου εύλογο η συλλογή και ο υπολογιστικός πυρήνας να αναπτυχθούν σε διαφορετικά περιβάλλοντα — αλλά προτιμήθηκε συνειδητά, διότι επιτρέπει την επαναχρησιμοποίηση κοινών συναρτήσεων, ιδίως εκείνων που αφορούν τη μετατροπή του χρόνου, την κανονικοποίηση των ονομάτων και την αντιστοίχιση των στάσεων, χωρίς κίνδυνο απόκλισης μεταξύ των σταδίων.

Τρεις λόγοι υπαγόρευαν την επιλογή. Ο πρώτος είναι η ωριμότητα του οικοσυστήματος βιβλιοθηκών στους ακριβώς τομείς που απαιτούσε η εργασία: ανάκτηση δεδομένων από το διαδίκτυο, ανάλυση HTML, ασύγχρονη δικτυακή επικοινωνία, κρυπτογραφία και ανάπτυξη διαδικτυακών υπηρεσιών. Ο δεύτερος είναι η ταχύτητα ανάπτυξης και δοκιμής, κρίσιμη σε ένα έργο όπου η μορφή των δεδομένων ήταν άγνωστη εκ των προτέρων και ανακαλύφθηκε σταδιακά, μέσω επαναλαμβανόμενων κύκλων πειραματισμού. Ο τρίτος είναι η πλούσια πρότυπη βιβλιοθήκη, από την οποία αξιοποιήθηκαν άμεσα τα `json`, `sqlite3`, `heapq`, `pathlib`, `datetime`, `unicodedata`,

`argparse`, `asyncio` και `concurrent.futures`, περιορίζοντας τις εξωτερικές εξαρτήσεις σε ελάχιστες.

Το γνωστό μειονέκτημα της γλώσσας, η χαμηλότερη ταχύτητα εκτέλεσης έναντι μεταγλωττιζόμενων εναλλακτικών, αποδείχθηκε στην πράξη ακίνδυνο. Το μέγεθος του προβλήματος είναι μικρό με σύγχρονα δεδομένα: ο γράφος περιλαμβάνει μερικές εκατοντάδες κόμβους και περίπου εκατό χιλιάδες αναχωρήσεις, ενώ ο αλγόριθμος εκτελείται σε χώρο καταστάσεων φραγμένο τόσο ως προς το πλήθος σκελών όσο και ως προς τον χρονικό ορίζοντα. Όπως τεκμηριώνεται στο κεφάλαιο 6, ο χρόνος απόκρισης παραμένει εντός των ορίων της διαδραστικότητας χωρίς καμία βελτιστοποίηση χαμηλού επιπέδου. Στα στάδια όπου το κόστος ήταν πραγματικά αισθητό — στη μαζική συλλογή δεδομένων — αυτό οφειλόταν στην αναμονή του δικτύου και όχι στην υπολογιστική ισχύ, οπότε αντιμετωπίστηκε με τεχνικές ταυτοχρονισμού που περιγράφονται αμέσως παρακάτω.

3.2 Τεχνολογίες ανάκτησης δεδομένων: `requests`, `BeautifulSoup`, `aiohttp` / `asyncio`

Η ανάκτηση των χρονοδιαγραμμάτων απαιτήσε δύο διαφορετικές τεχνικές προσεγγίσεις, ανάλογα με τη φύση κάθε πηγής. Όπου η πληροφορία εκτίθεται μέσω εσωτερικών υπηρεσιών που επιστρέφουν δομημένες αποκρίσεις JSON, η συλλογή περιορίζεται σε κλήσεις προς αυτές· όπου, αντιθέτως, η πληροφορία υπάρχει μόνο ενσωματωμένη στον κώδικα των σελίδων, απαιτείται συντακτική ανάλυση του HTML και εξαγωγή των πεδίων από τη δομή του εγγράφου.

Για την πρώτη περίπτωση χρησιμοποιήθηκε η βιβλιοθήκη `requests`, καθιερωμένος συγχρονισμένος πελάτης HTTP για την Python, ο οποίος αφαιρεί τη διαχείριση συνδέσεων, κεφαλίδων, ανακατευθύνσεων και κωδικοποιήσεων πίσω από μια ιδιαίτερα απλή διεπαφή [18]. Αξιοποιήθηκε τόσο στους μηχανισμούς συλλογής όσο και στην επικοινωνία με τις γεωχωρικές υπηρεσίες της ενότητας 3.7, με ρητό ορισμό κεφαλίδας ταυτοποίησης του πράκτορα χρήστη και χρονικού ορίου απόκρισης σε κάθε κλήση.

Για τη δεύτερη περίπτωση χρησιμοποιήθηκε η `BeautifulSoup`, βιβλιοθήκη ανάλυσης HTML που κατασκευάζει από το έγγραφο ένα δέντρο διάσχισης και επιτρέπει την αναζήτηση κόμβων βάσει ετικέτας, κλάσης ή ιεραρχικής θέσης [19]. Το βασικό της πλεονέκτημα έναντι της αναζήτησης με κανονικές εκφράσεις είναι η ανθεκτικότητα σε συντακτικά ατελές σήμανση, συνήθη σε σελίδες που παράγονται δυναμικά, καθώς και η αναγνωσιμότητα του κώδικα εξαγωγής, ο οποίος εκφράζεται σε όρους δομής και όχι σε όρους χαρακτήρων.

Η κρισιμότερη επιλογή του σταδίου αυτού αφορούσε, ωστόσο, τον ταυτοχρονισμό. Η πλήρης συλλογή απαιτεί δεκάδες χιλιάδες αιτήματα, ένα για κάθε συνδυασμό αφετηρίας, προορισμού και ημερομηνίας δειγματοληψίας. Εκτελούμενα σειριακά, με τυπικό χρόνο απόκρισης της τάξης του ενός δευτερολέπτου, τα αιτήματα αυτά θα απαιτούσαν χρόνο μετρήσιμο σε ημέρες, με το επεξεργαστικό φορτίο να παραμένει ουσιαστικά μηδενικό: πρόκειται για φόρτο δεσμευμένο από την είσοδο-έξοδο και όχι από τον επεξεργαστή. Η λύση αναζητήθηκε επομένως στην επικάλυψη των αναμονών.

Για την πλατφόρμα με τον μεγαλύτερο όγκο δεδομένων επιλέχθηκε το ασύγχρονο υπόδειγμα της `asyncio` σε συνδυασμό με τον πελάτη `aiohttp`. Το υπόδειγμα αυτό εκτελεί χιλιάδες συνυπάρχουσες συνρουτίνες σε ένα μοναδικό νήμα, παραχωρώντας τον έλεγχο στον βρόχο συμβάντων σε κάθε σημείο αναμονής, και αποφεύγει έτσι τόσο το κόστος δημιουργίας νημάτων όσο και τους κινδύνους του κοινόχρηστου μεταβλητού κράτους [20]. Ο έλεγχος του ρυθμού επιτεύχθηκε με σημαφόρους δύο επιπέδων: έναν καθολικό, που περιορίζει το συνολικό πλήθος ταυτόχρονων αιτημάτων, και έναν ανά φορέα, ώστε καμία μεμονωμένη υπηρεσία να μην επιβαρύνεται δυσανάλογα. Για τις υπόλοιπες πηγές, όπου ο όγκος ήταν μικρότερος και η ασύγχρονη αναδιατύπωση δεν δικαιολογούνταν, προτιμήθηκε η κλασική πολυνηματική εκτέλεση μέσω `ThreadPoolExecutor` της `concurrent.futures`. Η επιλογή αυτή είναι εξίσου αποτελεσματική για φόρτο δεσμευμένο από την είσοδο-έξοδο, καθώς το καθολικό κλείδωμα διερμηνέα της CPython απελευθερώνεται κατά τη διάρκεια των δικτυακών λειτουργιών, και έχει το πλεονέκτημα ότι επιτρέπει την επαναχρησιμοποίηση της συγχρονισμένης `requests` χωρίς τροποποίηση.

3.3 SQLite και σχεσιακή αποθήκευση

Ως μόνιμος χώρος αποθήκευσης του ενοποιημένου συνόλου δεδομένων επιλέχθηκε η SQLite, ενσωματωμένο σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων που λειτουργεί ως βιβλιοθήκη εντός της ίδιας της διεργασίας και αποθηκεύει ολόκληρη τη βάση σε ένα αρχείο [21]. Η υποστήριξή της περιλαμβάνεται στην πρότυπη βιβλιοθήκη της Python μέσω του αρθρώματος `sqlite3`, οπότε δεν εισάγει καμία εξωτερική εξάρτηση.

Η επιλογή έναντι ενός εξυπηρετητή βάσης δεδομένων, όπως η PostgreSQL ή η MySQL, στηρίχθηκε στα χαρακτηριστικά του συγκεκριμένου φόρτου. Οι εγγραφές γίνονται μαζικά και σπάνια, κατά τη φόρτωση των συλλεγμένων δεδομένων, ενώ οι αναγνώσεις εκτελούνται από μία μόνο διεργασία και μάλιστα άπαξ, κατά την κατασκευή του γράφου. Δεν υπάρχει ταυτόχρονη πρόσβαση πολλών χρηστών, ούτε ανάγκη ελέγχου δικαιωμάτων ή απομακρυσμένης σύνδεσης — δηλαδή κανένα από τα προβλήματα για τα οποία σχεδιάζονται τα συστήματα πελάτη-εξυπηρετητή. Αντιθέτως, το γεγονός ότι ολόκληρη η βάση συνιστά ένα μεταφερόμενο αρχείο αποτελεί ουσιώδες πλεονέκτημα για την αναπαραγωγικότητα της εργασίας, καθώς επιτρέπει τη διανομή του συνόλου δεδομένων χωρίς καμία διαδικασία εγκατάστασης.

Το σχήμα, το οποίο περιγράφεται αναλυτικά στην ενότητα 4.8, οργανώνεται γύρω από έναν κεντρικό πίνακα δρομολογίων με ευρετήρια στη στάση αναχώρησης, στη στάση άφιξης και στο αρχείο προέλευσης. Η φόρτωση εκτελείται ανά αρχείο με πλήρη αντικατάσταση των προηγούμενων εγγραφών του, εξασφαλίζοντας την ιδιότητα της ταυτοδυναμίας: η επανεκτέλεση της διαδικασίας επί των ίδιων δεδομένων παράγει πανομοιότυπη βάση, χωρίς διπλοεγγραφές. Αξιοποιήθηκαν, τέλος, δύο ρυθμίσεις που επηρεάζουν αισθητά την ταχύτητα μαζικής εισαγωγής: το ημερολόγιο εγγραφής προτεταγμένης καταγραφής και η χαλάρωση του συγχρονισμού με τον δίσκο, επιλογές αποδεκτές εδώ διότι η βάση είναι αναπαραγώγιμη από τα πρωτογενή αρχεία σε περίπτωση απώλειας.

Πέρα από τον ρόλο της ως χώρου αποθήκευσης, η βάση αποδείχθηκε πολύτιμο εργαλείο ανάλυσης. Η δηλωτική φύση της SQL επέτρεψε τη διατύπωση ερωτημάτων συγκέντρωσης —

πλήθος δρομολογίων ανά φορέα και πηγή, στάσεις χωρίς καμία αναχώρηση, εγγραφές με ελλιπή διάρκεια— χωρίς συγγραφή κώδικα διάσχισης, και από αυτά προέκυψε το μεγαλύτερο μέρος των ποσοτικών στοιχείων που παρουσιάζονται στην ενότητα 4.9. Οι γνωστοί περιορισμοί του συστήματος, δηλαδή η χαλαρή τήρηση των τύπων και η υποστήριξη ενός μόνο συγγραφέα τη φορά, δεν επηρεάζουν τη συγκεκριμένη χρήση.

3.4 Σειριοποίηση δεδομένων: JSON και pickle

Στο σύστημα χρησιμοποιούνται δύο μορφές σειριοποίησης, με σαφώς διακριτούς ρόλους. Η μορφή JSON αποτελεί την ενδιάμεση αναπαράσταση μεταξύ της συλλογής και της βάσης: κάθε φορέας αποτυπώνεται σε ένα αρχείο που περιέχει τον κατάλογο των δρομολογίων του, ενώ ένα χωριστό αρχείο ανά πηγή καταγράφει τις στάσεις και τις συντεταγμένες τους. Η επιλογή αυτή δεν είναι διακοσμητική. Τα αρχεία JSON είναι αναγνώσιμα από άνθρωπο, γεγονός που επέτρεψε τον οπτικό έλεγχο των αποτελεσμάτων κάθε μηχανισμού συλλογής κατά την ανάπτυξη, και ταυτόχρονα λειτουργούν ως αμετάβλητο στιγμιότυπο των πρωτογενών δεδομένων, από το οποίο η βάση μπορεί να ανακατασκευαστεί οποτεδήποτε. Η σειριοποίηση εκτελείται με ρητή απενεργοποίηση της διαφυγής μη λατινικών χαρακτήρων, ώστε τα ελληνικά τοπωνύμια να παραμένουν αναγνώσιμα, και με κωδικοποίηση UTF-8 σε όλα τα στάδια ανάγνωσης και εγγραφής.

Η δεύτερη μορφή, το άρθρωμα `pickle` της Python, χρησιμοποιείται για εντελώς διαφορετικό σκοπό: την προσωρινή αποθήκευση του έτοιμου γράφου. Η κατασκευή του γράφου από τη βάση απαιτεί ανάγνωση όλων των δρομολογίων, αντιστοίχιση των στάσεων στις κανονικές οντότητές τους, μετατροπή των χρόνων σε λεπτά και ταξινόμηση των αναχωρήσεων κάθε στάσης. Η επανάληψη της διαδικασίας σε κάθε εκκίνηση του εξυπηρετητή θα ήταν σπατάλη, καθώς το αποτέλεσμα μεταβάλλεται μόνο όταν αλλάζουν τα δεδομένα. Το `pickle` σειριοποιεί απευθείας τις δομές της Python, διατηρώντας τους τύπους τους χωρίς ενδιάμεση μετατροπή, και επιτρέπει την ανάκτηση του γράφου σε χρόνο τάξεις μεγέθους μικρότερο από την ανακατασκευή. Το τίμημα είναι διπλό και δηλώνεται ρητά: η μορφή δεν είναι διαλειτουργική, ενώ η αποσειριοποίηση αναξιόπιστου αρχείου συνιστά κίνδυνο ασφαλείας, καθώς μπορεί να προκαλέσει εκτέλεση αυθαίρετου κώδικα [22]. Στην προκειμένη περίπτωση το αρχείο παράγεται τοπικά από το ίδιο το σύστημα και ουδέποτε λαμβάνεται από εξωτερική πηγή, οπότε ο κίνδυνος δεν υφίσταται.

3.5 Flask και αρχιτεκτονική REST

Η διαδικτυακή διεπαφή στηρίχθηκε στο Flask, ελαφρύ πλαίσιο ανάπτυξης εφαρμογών ιστού για την Python, το οποίο παρέχει δρομολόγηση αιτημάτων, μηχανισμό προτύπων και βοηθητικές συναρτήσεις για αποκρίσεις JSON, αφήνοντας κάθε άλλη απόφαση στον προγραμματιστή [23]. Η εναλλακτική επιλογή ενός πλήρους πλαισίου, όπως το Django, θα προσέφερε αντιστοίχιση αντικειμένων-σχέσεων, σύστημα διαχείρισης χρηστών και πίνακα διαχείρισης — δηλαδή υποδομή για την οποία η εφαρμογή δεν έχει καμία χρήση, αφού δεν διαθέτει λογαριασμούς, δεν γράφει στη βάση και δεν διατηρεί κατάσταση συνεδρίας.

Η οργάνωση της υπηρεσίας ακολουθεί τις αρχές της αρχιτεκτονικής REST: ο εξυπηρετητής εκθέτει πόρους μέσω σταθερών διευθύνσεων, χρησιμοποιεί τις μεθόδους του πρωτοκόλλου HTTP

σύμφωνα με τη σημασιολογία τους και απαντά χωρίς να διατηρεί κατάσταση μεταξύ διαδοχικών αιτημάτων, καθιστώντας κάθε αίτημα αυτοτελές [24]. Υλοποιήθηκαν τρία σημεία πρόσβασης: ένα που επιστρέφει τον πλήρη κατάλογο των κανονικοποιημένων στάσεων με τις συντεταγμένες τους για την τροφοδότηση του χάρτη, ένα που εξυπηρετεί την αναζήτηση στάσης με αυτόματη συμπλήρωση, και ένα που δέχεται τις παραμέτρους αναζήτησης και επιστρέφει τις υπολογισμένες διαδρομές. Ο σαφής διαχωρισμός μεταξύ του εξυπηρετητή, ο οποίος παράγει αποκλειστικά δεδομένα, και του πελάτη, ο οποίος αναλαμβάνει εξ ολοκλήρου την παρουσίαση, έχει και μια πρακτική συνέπεια που υπερβαίνει την καθαρότητα του σχεδιασμού: η ίδια υπηρεσία μπορεί να καταναλωθεί από οποιονδήποτε άλλο πελάτη — εφαρμογή κινητού, σενάριο αυτοματοποίησης ή τρίτο σύστημα — χωρίς καμία τροποποίηση, ακριβώς δηλαδή με τον τρόπο που θα όφειλαν να εκθέτουν τα δεδομένα τους οι ίδιες οι πηγές. Ο γράφος φορτώνεται μία φορά κατά την εκκίνηση της εφαρμογής και παραμένει στη μνήμη, ώστε κάθε αίτημα να εξυπηρετείται χωρίς πρόσβαση στον δίσκο. Η μοναδική σελίδα της εφαρμογής αποδίδεται μέσω της μηχανής προτύπων Jinja2, την οποία ενσωματώνει το Flask, ενώ ο ενσωματωμένος εξυπηρετητής ανάπτυξης επαρκεί για τους σκοπούς της παρουσίασης και θα αντικαθίστατο από τυπικό εξυπηρετητή WSGI σε συνθήκες πραγματικής λειτουργίας.

3.6 Τεχνολογίες frontend: HTML, CSS, JavaScript

Η διεπαφή χρήστη αναπτύχθηκε αποκλειστικά με τις θεμελιώδεις τεχνολογίες του ιστού, χωρίς πλαίσιο διεπαφής και χωρίς στάδιο μεταγλώττισης. Η δομή περιγράφεται σε HTML5 με σημασιολογικά στοιχεία, η εμφάνιση σε CSS με χρήση των μηχανισμών διάταξης Flexbox και Grid και μεταβλητών για τη συγκέντρωση της χρωματικής παλέτας σε ένα σημείο, ενώ η συμπεριφορά υλοποιείται σε JavaScript που εκτελείται απευθείας στον περιηγητή.

Η αποφυγή ενός πλαισίου όπως το React ή το Vue υπήρξε συνειδητή επιλογή και όχι παράλειψη. Η εφαρμογή διαθέτει μία μόνο οθόνη και περιορισμένη κατάσταση — τις δύο επιλεγμένες στάσεις, τις παραμέτρους αναζήτησης και το τελευταίο αποτέλεσμα — μέγεθος για το οποίο η υποδομή ενός πλαισίου εισάγει περισσότερη πολυπλοκότητα από όση αφαιρεί. Η απουσία σταδίου μεταγλώττισης σημαίνει επιπλέον ότι το σύστημα εκτελείται με μοναδική προϋπόθεση την ύπαρξη διερμηνέα Python, χωρίς εγκατάσταση περιβάλλοντος Node.js και χωρίς διαχειριστή πακέτων για την πλευρά του πελάτη, γεγονός που απλοποιεί ουσιαστικά την αναπαραγωγή της εργασίας.

Η επικοινωνία με το backend γίνεται μέσω της διεπαφής fetch και ασύγχρονων συναρτήσεων, με τις αποκρίσεις JSON να μετατρέπονται δυναμικά σε στοιχεία του εγγράφου. Δύο σημεία απαίτησαν ιδιαίτερη προσοχή: η αναζήτηση με αυτόματη συμπλήρωση, όπου η αποστολή αιτήματος σε κάθε πληκτρολόγηση αντιμετωπίστηκε με χρονική απόσβεση των γεγονότων, και η αποκρίσιμη συμπεριφορά της διάταξης, η οποία επιτεύχθηκε με ερωτήματα μέσων και σχετικές μονάδες μέτρησης ώστε η εφαρμογή να παραμένει λειτουργική σε οθόνες κινητών συσκευών. Ολόκληρη η διεπαφή αποδίδεται σε ελληνικά, με ρητή δήλωση της γλώσσας του εγγράφου και κωδικοποίηση UTF-8, ώστε τα τοπωνύμια να εμφανίζονται και να συγκρίνονται σωστά ανεξαρτήτως τονισμού — ζήτημα που, όπως θα φανεί στην ενότητα 5.5.7, απαιτεί ιδιαίτερη μεταχείριση και στην πλευρά του εξυπηρετητή.

3.7 Γεωχωρικές υπηρεσίες: OpenStreetMap, Nominatim, OSRM, Leaflet

Η γεωγραφική διάσταση του προβλήματος απαίτησε τέσσερις συνιστώσες, οι οποίες προέρχονται όλες από το οικοσύστημα του OpenStreetMap. Το ίδιο το OpenStreetMap είναι συνεργατική βάση γεωχωρικών δεδομένων παγκόσμιας κάλυψης, διαθέσιμη υπό ανοικτή άδεια που επιτρέπει την ελεύθερη χρήση και αναδιανομή με μοναδική ουσιαστική υποχρέωση την αναφορά της προέλευσης [25]. Η προϋπόθεση αυτή ήταν καθοριστική: η χρήση εμπορικών χαρτογραφικών υπηρεσιών θα εισήγαγε περιορισμούς αδειοδότησης και εξάρτηση από κλειδιά πρόσβασης, ασύμβατα με τον χαρακτήρα μιας ακαδημαϊκής εργασίας που επιδιώκει να είναι αναπαραγώγιμη.

Η υπηρεσία Nominatim αξιοποιήθηκε για τη γεωκωδικοποίηση, δηλαδή τη μετατροπή του ονόματος κάθε στάσης σε ζεύγος συντεταγμένων. Πρόκειται για μηχανή αναζήτησης επί των δεδομένων του OpenStreetMap, της οποίας η δημόσια εγκατάσταση διέπεται από αυστηρή πολιτική χρήσης: ένα αίτημα ανά δευτερόλεπτο κατ' ανώτατο όριο και υποχρεωτική δήλωση αναγνωρίσιμου πράκτορα χρήστη [26]. Η πολιτική αυτή τηρήθηκε με ρητή αναμονή μεταξύ διαδοχικών κλήσεων και με περιοδική αποθήκευση της προόδου, ώστε μια διακοπή να μην ακυρώνει τη μέχρι εκείνη τη στιγμή εργασία. Καθώς η αναζήτηση ονόματος επιστρέφει συχνά πολλαπλά ομώνυμα αποτελέσματα σε διαφορετικές περιοχές, τα ερωτήματα περιορίστηκαν γεωγραφικά στην ελληνική επικράτεια και, σε πρώτη φάση, εντός πλαισίου προβολής γύρω από το κέντρο του νομού στον οποίο ανήκει το ΚΤΕΛ της στάσης — μηχανισμός που συνιστά την πρακτική έκφραση της χρήσης του γεωγραφικού συμφραζομένου, όπως αναλύθηκε θεωρητικά στην ενότητα 2.8.3.

Η υπηρεσία OSRM, μηχανή δρομολόγησης σε οδικά δίκτυα βασισμένη επίσης σε δεδομένα του OpenStreetMap, χρησιμοποιήθηκε για τη συμπλήρωση των ελλιπών διαρκειών διαδρομής [27]. Δεδομένου ότι ορισμένες πηγές δημοσιεύουν ώρα αναχώρησης χωρίς ώρα άφιξης, και ότι μια αναχώρηση με άγνωστη διάρκεια δεν μπορεί να χρησιμοποιηθεί ως ενδιάμεσο σκέλος διαδρομής, ο εκτιμώμενος χρόνος οδικής μετακίνησης μεταξύ των συντεταγμένων των δύο στάσεων αξιοποιήθηκε ως προσέγγιση. Η προσέγγιση αυτή είναι ασφαλώς ατελής, καθώς αγνοεί τις ενδιάμεσες στάσεις και τη διαφορά ταχύτητας ενός λεωφορείου από εκείνη ενός επιβατικού οχήματος, και για τον λόγο αυτό τα συμπληρωμένα πεδία επισημαίνονται και αποτιμώνται χωριστά στο κεφάλαιο 4.

Τέλος, η βιβλιοθήκη Leaflet ανέλαβε την οπτικοποίηση. Πρόκειται για ελαφρά βιβλιοθήκη JavaScript διαδραστικών χαρτών, η οποία αποδίδει πλακίδια από εξωτερικό πάροχο και υποστηρίζει δείκτες, γραμμές και συμβάντα αλληλεπίδρασης [28]. Χρησιμοποιήθηκε τόσο στη διεπαφή του τελικού χρήστη, για την επιλογή στάσεων και την προβολή της διαδρομής, όσο και στο εσωτερικό εργαλείο επικύρωσης της ενοποίησης των στάσεων που περιγράφεται στην ενότητα 4.6.4. Λόγω του πλήθους των σημείων, αξιοποιήθηκε επιπλέον η επέκταση ομαδοποίησης δεικτών, η οποία συγχωνεύει γειτονικούς δείκτες σε συσσωματώματα ανάλογα με το επίπεδο μεγέθυνσης και διατηρεί την απόκριση του χάρτη ομαλή.

3.8 Το πρότυπο GTFS

Το GTFS (General Transit Feed Specification) αποτελεί σήμερα το επικρατέστερο πρότυπο δημοσίευσης στατικών δεδομένων συγκοινωνιών. Διαμορφώθηκε αρχικά ως συνεργασία μεταξύ ενός φορέα συγκοινωνιών και μιας εταιρείας χαρτογραφικών υπηρεσιών και εξελίχθηκε σε *de facto* παγκόσμιο πρότυπο, υιοθετημένο από χιλιάδες φορείς [16]. Τεχνικά συνίσταται σε ένα συμπίεσμένο αρχείο που περιέχει πίνακες σε μορφή CSV με τυποποιημένα ονόματα και πεδία: οι στάσεις με τα αναγνωριστικά και τις συντεταγμένες τους, οι γραμμές, τα επιμέρους δρομολόγια κάθε γραμμής, οι διαδοχικές αφίξεις και αναχωρήσεις κάθε δρομολογίου ανά στάση, και το ημερολόγιο που ορίζει τις ημέρες εκτέλεσης, με χωριστό πίνακα εξαιρέσεων για αργίες και έκτακτες μεταβολές.

Η σημασία του προτύπου για την παρούσα εργασία είναι διπλή και, κατά κάποιον τρόπο, αρνητική. Αφενός, το GTFS ορίζει ακριβώς εκείνο που λείπει από το ελληνικό υπεραστικό δίκτυο: μια κοινή, μηχανικά αναγνώσιμη μορφή στην οποία θα μπορούσε κάθε ΚΤΕΛ να εκθέτει τα δεδομένα του, καθιστώντας περιττό το σύνολο της διαδικασίας ανάκτησης του κεφαλαίου 4. Αφετέρου, η δομή του λειτούργησε ως εννοιολογικό πρότυπο για τη σχεδίαση του δικού μας σχήματος, ακόμη και εκεί όπου δεν υιοθετήθηκε αυτούσια. Η απόκλιση ήταν αναπόφευκτη, διότι το GTFS προϋποθέτει πληροφορία που οι πηγές δεν παρέχουν — ιδίως τη σύνθεση κάθε δρομολογίου ως ακολουθίας ενδιάμεσων στάσεων με χρόνους, ενώ τα διαθέσιμα δεδομένα αποτυπώνουν μεμονωμένες συνδέσεις μεταξύ ζευγών στάσεων. Η εξαγωγή του τελικού συνόλου δεδομένων σε μορφή GTFS καταγράφεται συνεπώς ως μελλοντική επέκταση στο κεφάλαιο 7. Αξίζει να σημειωθεί ότι σε ευρωπαϊκό επίπεδο το αντίστοιχο θεσμικό πρότυπο είναι το NeTEx, σημαντικά εκτενέστερο και πλησιέστερο στις απαιτήσεις του κανονιστικού πλαισίου που αναφέρθηκε στην ενότητα 1.1, με το GTFS να υπερτερεί ωστόσο σαφώς σε απλότητα και σε πραγματική διάδοση [17].

3.9 Κρυπτογραφία στην επικοινωνία με APIs (AES-CBC, pycryptodome)

Μία από τις τέσσερις πηγές παρουσίασε ιδιαιτερότητα που απαίτησε κρυπτογραφικά εργαλεία: η εσωτερική της υπηρεσία δεν ανταλλάσσει απλό JSON αλλά συμβολοσειρές κωδικοποιημένες κατά Base64, οι οποίες περιέχουν κρυπτογραφημένο περιεχόμενο. Τόσο το σώμα του αιτήματος όσο και η απόκριση υφίστανται την ίδια επεξεργασία, με αποτέλεσμα η επικοινωνία να είναι αδιαφανής σε οποιονδήποτε παρατηρεί απλώς την κίνηση του δικτύου.

Η ανάλυση του κώδικα που εκτελείται στον περιηγητή —ο οποίος είναι εξ ορισμού δημόσιος, αφού αποστέλλεται αυτούσιος σε κάθε επισκέπτη— αποκάλυψε ότι πρόκειται για τον αλγόριθμο AES σε λειτουργία CBC, με κλειδί 128 ψηφίων και σταθερό διάνυσμα αρχικοποίησης, ενώ η συμπλήρωση των τμημάτων ακολουθεί το σχήμα PKCS#7. Η αναπαραγωγή του μηχανισμού σε Python έγινε με τη βιβλιοθήκη pycryptodome, διάδοχο της παλαιότερης pycrypto, η οποία υλοποιεί τους καθιερωμένους συμμετρικούς αλγορίθμους μαζί με τις αντίστοιχες συναρτήσεις συμπλήρωσης [29]. Η υλοποίηση περιορίζεται σε δύο συναρτήσεις, μία για την κρυπτογράφηση του αιτήματος και μία για την αποκρυπτογράφηση της απόκρισης, οι οποίες παρεμβάλλονται διαφανώς σε κάθε κλήση.

Η τεχνική αυτή διάσταση επιτρέπει μια παρατήρηση με ευρύτερη σημασία. Η ύπαρξη σταθερού κλειδιού ενσωματωμένου στον κώδικα του πελάτη δεν συνιστά μέτρο ασφαλείας αλλά συσκότιση: κάθε πελάτης που μπορεί να εκτελέσει την εφαρμογή διαθέτει εξ ορισμού το κλειδί, και επομένως το σχήμα δεν προσθέτει καμία εμπιστευτικότητα πέραν εκείνης που παρέχει ήδη το πρωτόκολλο HTTPS. Η προστασία που προσφέρει είναι απέναντι στην περιστασιακή επιθεώρηση και όχι απέναντι σε συστηματική ανάλυση. Ας σημειωθεί, τέλος, ότι τα δεδομένα στα οποία δίνει πρόσβαση ο μηχανισμός είναι ακριβώς εκείνα που η ίδια η υπηρεσία εμφανίζει δημόσια σε κάθε επισκέπτη της: δεν παρακάμπτεται μηχανισμός πιστοποίησης ταυτότητας, δεν αποκτάται πρόσβαση σε προσωπικά ή εμπορικά ευαίσθητα στοιχεία, και η συλλογή διενεργήθηκε με ρυθμό που δεν επιβαρύνει τη λειτουργία των εξυπηρετητών.

3.10 Εργαλεία ανάπτυξης και ελέγχου εκδόσεων

Η ανάπτυξη πραγματοποιήθηκε σε περιβάλλον Windows με τον επεξεργαστή Visual Studio Code, ο οποίος προσφέρει ενσωματωμένη υποστήριξη Python, αποσφαλμάτωση με σημεία διακοπής και τερματικό εντός του ίδιου παραθύρου. Κρισιμότερο εργαλείο του σταδίου της αντίστροφης μηχανίκευσης υπήρξαν, ωστόσο, τα εργαλεία ανάπτυξης του περιηγητή και ιδίως η καρτέλα δικτύου: μέσω αυτής καταγράφηκαν τα αιτήματα που εκτελεί κάθε ιστότοπος κατά την αναζήτηση δρομολογίων, εντοπίστηκαν οι διευθύνσεις των εσωτερικών υπηρεσιών, οι απαιτούμενες κεφαλίδες και η δομή των σωμάτων, ενώ η επιθεώρηση του κώδικα JavaScript οδήγησε στην αποκάλυψη του κρυπτογραφικού σχήματος της προηγούμενης ενότητας.

Όλα τα προγράμματα του συστήματος εκτελούνται από τη γραμμή εντολών και δέχονται παραμέτρους μέσω του αρθρώματος `argparse`, γεγονός που επιτρέπει τη μερική επανεκτέλεση σταδίων — για παράδειγμα, τη συλλογή ενός μόνο ΚΤΕΛ ή τη συμπλήρωση διαρκειών σε συγκεκριμένο φάκελο — χωρίς τροποποίηση του κώδικα. Για τις μακροχρόνιες διαδικασίες συλλογής αξιοποιήθηκε η βιβλιοθήκη `tqdm`, που εμφανίζει γραμμή προόδου με εκτίμηση του υπολειπόμενου χρόνου, ενώ τα σφάλματα κάθε εκτέλεσης καταγράφονται σε χωριστά αρχεία, ώστε οι αποτυχημένες κλήσεις να μπορούν να επαναληφθούν στοχευμένα.

Ο έλεγχος εκδόσεων του πηγαίου κώδικα βασίστηκε στο Git, το οποίο επιτρέπει την καταγραφή της εξέλιξης της υλοποίησης και την επιστροφή σε προηγούμενη κατάσταση όταν μια σχεδιαστική επιλογή αποδεικνύεται εσφαλμένη. Ιδιαίτερη μνεία αξίζει η διάκριση μεταξύ κώδικα και δεδομένων: τα συλλεγμένα αρχεία, μεγέθους αρκετών δεκάδων megabytes, δεν αποτελούν πηγαίο κώδικα και η παρακολούθησή τους από ένα σύστημα ελέγχου εκδόσεων θα ήταν αντιπαραγωγική. Ο ρόλος του ιστορικού για τα δεδομένα επιτελείται από τα ίδια τα χρονοσημασμένα στιγμιότυπα των αρχείων JSON, τα οποία, όπως επισημάνθηκε στην ενότητα 1.3, αποκτούν πρόσθετη αξία ακριβώς επειδή οι πηγές δεν τηρούν καμία μορφή ιστορικού.

Επίλογος κεφαλαίου

Το κεφάλαιο παρουσίασε τις τεχνολογικές επιλογές της εργασίας ακολουθώντας τη ροή των δεδομένων: τη γλώσσα υλοποίησης και τις βιβλιοθήκες ανάκτησης με τους δύο μηχανισμούς ταυτοχρονισμού, τη σχεσιακή αποθήκευση και τις δύο μορφές σειριοποίησης, το πλαίσιο της

διαδικτυακής υπηρεσίας και τις τεχνολογίες της διεπαφής, τις γεωχωρικές υπηρεσίες του οικοσυστήματος OpenStreetMap, το πρότυπο GTFS ως μέτρο σύγκρισης και τα κρυπτογραφικά εργαλεία που απαίτησε μία από τις πηγές. Κοινός άξονας των επιλογών υπήρξε η προτίμηση σε ώριμα, ανοικτά και ελαφρά εργαλεία, στο μέτρο ακριβώς που το πρόβλημα απαιτεί, ώστε το σύστημα να παραμένει αναπαραγώγιμο με ελάχιστες προϋποθέσεις εγκατάστασης. Με το τεχνολογικό αυτό υπόβαθρο δεδομένο, το επόμενο κεφάλαιο περιγράφει την εφαρμογή του στο πρώτο και δυσκολότερο στάδιο της εργασίας: τη μετατροπή τεσσάρων ασύνδετων και μη τεκμηριωμένων πηγών σε ένα ενιαίο, συνεπές σύνολο δεδομένων.

Κεφάλαιο 4

Ανάκτηση δεδομένων και δημιουργία σχεσιακής Βάσης Δεδομένων

Το κεφάλαιο αυτό περιγράφει το στάδιο που προηγείται κάθε αλγοριθμικής επεξεργασίας και το οποίο, στην περίπτωση του ελληνικού υπεραστικού δικτύου, αποδείχθηκε το πλέον χρονοβόρο: τη μετατροπή τεσσάρων ασύνδετων, μη τεκμηριωμένων και αμοιβαία ασύμβατων πηγών σε ένα ενιαίο, συνεπές και μηχανικά επεξεργάσιμο σύνολο δεδομένων. Η ανάπτυξη ακολουθεί τη φυσική σειρά της αλυσίδας επεξεργασίας. Καταγράφονται πρώτα οι πηγές και η τεχνολογική τους ταξινόμηση, παρουσιάζεται στη συνέχεια η μεθοδολογία αντίστροφης μηχανίκευσης που επέτρεψε την πρόσβαση στις εσωτερικές τους υπηρεσίες, περιγράφονται οι μηχανισμοί συλλογής και η στρατηγική με την οποία ανακατασκευάστηκε το εβδομαδιαίο πρότυπο εκτέλεσης, αναλύεται το κρισιμότερο βήμα της ενοποίησης των οντοτήτων στάσεων, και τέλος τεκμηριώνονται η συμπλήρωση των ελλিপών διαρκειών, η σχεδίαση της σχεσιακής βάσης και η ποιοτική αποτίμηση του αποτελέσματος. Σε κάθε στάδιο επιχειρείται να καταστεί σαφές όχι μόνο τι έγινε αλλά και γιατί επιλέχθηκε η συγκεκριμένη λύση έναντι των εναλλακτικών, καθώς και ποια ίχνη αφήνει η επιλογή αυτή στην ποιότητα και στα όρια του τελικού συνόλου δεδομένων.

4.1 Καταγραφή και ταξινόμηση των πηγών δεδομένων

Το πρώτο βήμα υπήρξε η συστηματική καταγραφή των διαδικτυακών σημείων στα οποία δημοσιεύονται υπεραστικά χρονοδιαγράμματα. Η καταγραφή ξεκίνησε από τον κατάλογο των εταιρειών ανά νομό και προχώρησε σε επίσκεψη και εξέταση κάθε ιστοτόπου χωριστά. Η διαπίστωση που καθόρισε τη συνέχεια ήταν ότι η φαινομενική πολλαπλότητα των δεκάδων ανεξάρτητων ιστοτόπων είναι επιφανειακή: πίσω από αυτούς λειτουργεί ένας πολύ μικρός αριθμός κοινών τεχνολογικών υποδομών ηλεκτρονικής έκδοσης εισιτηρίων, τις οποίες οι εταιρείες μισθώνουν και προσαρμόζουν οπτικά. Η ταξινόμηση, επομένως, δεν έγινε με κριτήριο τον φορέα αλλά την τεχνική υποδομή που εξυπηρετεί την αναζήτηση δρομολογίων, καθώς αυτή ακριβώς καθορίζει τη μορφή των αποκρίσεων και άρα τον μηχανισμό συλλογής που απαιτείται. Προέκυψαν έτσι τέσσερις οικογένειες, οι οποίες καλύπτουν συνολικά 35 ΚΤΕΛ και παρουσιάζονται στη συνέχεια κατά φθίνουσα συνεισφορά στο τελικό σύνολο δεδομένων.

4.1.1 Πλατφόρμα IME / ktelbus.gr

Η πλατφόρμα IME αποτελεί τη μεγαλύτερη και πλουσιότερη πηγή, εξυπηρετώντας 23 ΚΤΕΛ μέσω ενός κοινού ιστοτόπου στον οποίο κάθε εταιρεία αντιστοιχεί σε ένα τριγράμματο ή τετραγράμματο πρόθεμα διαδρομής και σε ένα αριθμητικό αναγνωριστικό φορέα. Πίσω από τη διεπαφή λειτουργεί μια εσωτερική υπηρεσία με σαφή δομή, της οποίας αξιοποιήθηκαν τρία σημεία πρόσβασης: ένα που επιστρέφει τον κατάλογο των στάσεων του φορέα, ένα που επιστρέφει, για δεδομένη αφετηρία, τους προορισμούς που εξυπηρετούνται από αυτήν, και ένα που επιστρέφει τις αναχωρήσεις ενός ζεύγους στάσεων σε συγκεκριμένη ημερομηνία. Η πηγή είναι η μόνη που παρέχει αυτοτελώς γεωγραφικές συντεταγμένες για κάθε στάση, καθώς και ένδειξη ενεργής προβολής στον ιστό, η οποία επιτρέπει τον αποκλεισμό στάσεων που δεν εξυπηρετούνται πλέον. Παρέχει επιπλέον τη διάρκεια κάθε σκέλους, ώστε ο συνολικός χρόνος διαδρομής να υπολογίζεται από την ώρα αναχώρησης και τη διάρκεια του τελευταίου σκέλους. Το τίμημα του πλούτου αυτού είναι ότι η επικοινωνία διεξάγεται κρυπτογραφημένα, ζήτημα που πραγματεύεται η ενότητα 4.2.2.

4.1.2 Πλατφόρμα e-analysis / e-ticketing.gr

Η δεύτερη οικογένεια εξυπηρετεί οκτώ ΚΤΕΛ της βόρειας και κεντρικής Ελλάδας, καθένα σε δικό του υποτομέα του κοινού τομέα e-ticketing.gr. Τεχνολογικά πρόκειται για εφαρμογή ASP.NET WebForms με βιβλιοθήκη διαδραστικών στοιχείων τρίτου κατασκευαστή, γεγονός με σοβαρές πρακτικές συνέπειες. Δεν υφίσταται καμία υπηρεσία που να επιστρέφει δομημένα δεδομένα: η αναζήτηση εκτελείται ως υποβολή φόρμας και το αποτέλεσμα ενσωματώνεται στον κώδικα της επόμενης σελίδας, οπότε η εξαγωγή απαιτεί συντακτική ανάλυση HTML. Επιπλέον, κάθε υποβολή πρέπει να συνοδεύεται από τα κρυφά πεδία κατάστασης της σελίδας, τα οποία είναι έγκυρα μόνο για την τρέχουσα συνεδρία και επιβάλλουν την ανάκτησή τους πριν από κάθε ερώτημα. Ο κατάλογος στάσεων δεν εκτίθεται χωριστά αλλά εξάγεται από τα ενσωματωμένα στη σελίδα δεδομένα των δύο πτυσσόμενων καταλόγων. Η πηγή δημοσιεύει μόνο ώρα αναχώρησης, χωρίς ώρα άφιξης και χωρίς συντεταγμένες.

4.1.3 Πλατφόρμα 3pi-telematics

Η τρίτη οικογένεια εξυπηρετεί πέντε ΚΤΕΛ, καθένα σε δικό του υποτομέα ηλεκτρονικού εισιτηρίου. Η υποδομή εκθέτει μια απλή υπηρεσία, στην οποία η επιθυμητή ενέργεια δηλώνεται ως παράμετρος της διεύθυνσης και η απόκριση επιστρέφεται σε μορφή JSON. Οι στάσεις οργανώνονται σε ομάδες με δικό τους αναγνωριστικό, ενώ οι αποκρίσεις των δρομολογίων περιλαμβάνουν ώρα αναχώρησης, ώρα άφιξης και διάρκεια σε λεπτά, καθιστώντας την πηγή τη μοναδική που παρέχει και τα τρία μεγέθη ρητά. Η ιδιαιτερότητά της είναι ότι τα πέντε ΚΤΕΛ εκτελούν διαφορετικές εκδόσεις της ίδιας υποδομής: η ενέργεια που επιστρέφει τον κατάλογο των αφετηριών δεν υποστηρίζεται από όλα, με ορισμένα να απαντούν με σφάλμα και να απαιτούν εναλλακτική κλήση. Ο μηχανισμός συλλογής αναγνωρίζει την περίπτωση αυτή και υποχωρεί αυτόματα στην παλαιότερη μορφή.

4.1.4 Αυτόνομα συστήματα (ΚΤΕΛ Ηλείας)

Η τέταρτη κατηγορία περιλαμβάνει συστήματα αναπτυγμένα κατά παραγγελία μεμονωμένων εταιρειών, με μοναδικό εκπρόσωπο στο τελικό σύνολο δεδομένων το ΚΤΕΛ Ηλείας. Πρόκειται για επέκταση συστήματος διαχείρισης περιεχομένου, όπου η αναζήτηση υλοποιείται με συνδυασμό υποβολής φόρμας για τα δρομολόγια και ασύγχρονης κλήσης για τη δυναμική ενημέρωση του καταλόγου προορισμών ανάλογα με την επιλεγμένη αφετηρία. Όπως και στην περίπτωση της e-analysis, η πληροφορία υπάρχει μόνο ενσωματωμένη στη σελίδα και περιορίζεται στην ώρα αναχώρησης. Η κατηγορία αυτή είναι η πλέον εύθραυστη, καθώς κάθε τέτοιο σύστημα αποτελεί μεμονωμένη περίπτωση χωρίς καμία επαναχρησιμοποιήσιμη γνώση, και μια ανασχεδίαση του ιστοτόπου ακυρώνει πλήρως τον αντίστοιχο μηχανισμό συλλογής.

4.2 Μεθοδολογία αντίστροφης μηχανίκευσης των διαδικτυακών υπηρεσιών

Καμία από τις τέσσερις πηγές δεν διαθέτει τεκμηριωμένη διεπαφή προγραμματισμού. Η πρόσβαση στα δεδομένα προϋπέθεσε, επομένως, την ανακάλυψη του τρόπου με τον οποίο η ίδια η διεπαφή του χρήστη επικοινωνεί με τον εξυπηρετητή και την πιστή αναπαραγωγή αυτής της επικοινωνίας από πρόγραμμα. Η μεθοδολογία που ακολουθήθηκε είναι κοινή για όλες τις πηγές και παρουσιάζεται εδώ ως γενική διαδικασία, με τις επιμέρους ιδιαιτερότητες να επισημαίνονται όπου εμφανίστηκαν.

4.2.1 Ανάλυση κίνησης δικτύου και εντοπισμός endpoints

Το βασικό εργαλείο υπήρξε η καρτέλα δικτύου των εργαλείων ανάπτυξης του περιηγητή. Η διαδικασία ξεκινά με μια πλήρη αναζήτηση δρομολογίου εκτελούμενη χειροκίνητα, κατά τη διάρκεια της οποίας καταγράφονται όλα τα αιτήματα που εκπέμπει η σελίδα. Το πλήθος τους είναι μεγάλο, όμως η συντριπτική πλειονότητα αφορά στατικούς πόρους και μπορεί να αποκλειστεί με φιλτράρισμα των ασύγχρονων κλήσεων. Ανάμεσα στις υπόλοιπες εντοπίζεται εκείνη που εκπέμπεται τη στιγμή ακριβώς της υποβολής και της οποίας η απόκριση περιέχει τις ώρες που εμφανίζονται στην οθόνη. Η αντιστοίχιση αυτή, ανάμεσα σε ό,τι βλέπει ο χρήστης και σε ό,τι επιστρέφει ο εξυπηρετητής, αποτελεί το κρίσιμο σημείο της διαδικασίας.

Με το αίτημα εντοπισμένο, καταγράφονται η διεύθυνση, η μέθοδος, οι κεφαλίδες και η δομή του σώματος, και ακολουθεί το στάδιο της αναπαραγωγής. Η κλήση επαναλαμβάνεται από πρόγραμμα και το αποτέλεσμα συγκρίνεται με εκείνο του περιηγητή· μόλις επιτευχθεί ταύτιση, ακολουθεί η σταδιακή απλοποίηση, δηλαδή η αφαίρεση παραμέτρων και κεφαλίδων μία προς μία, ώστε να προσδιοριστεί το ελάχιστο σύνολο που απαιτείται πράγματι. Το βήμα αυτό δεν είναι διακοσμητικό: μειώνει δραστικά την πολυπλοκότητα του κώδικα και αναδεικνύει ποια πεδία είναι σημασιολογικά ουσιώδη και ποια αποτελούν κατάλοιπα της συγκεκριμένης υλοποίησης. Στην περίπτωση της πλατφόρμας ΙΜΕ, για παράδειγμα, το σώμα του αιτήματος περιλαμβάνει πλήθος πεδίων που αφορούν το κανάλι πληρωμής, τον τύπο του εισιτηρίου και το αναγνωριστικό της συσκευής, τα οποία διατηρήθηκαν με σταθερές τιμές επειδή η παράλειψή τους οδηγούσε σε απόρριψη, χωρίς ωστόσο να επηρεάζουν το περιεχόμενο της απόκρισης.

Η ίδια ανάλυση αποκάλυψε και τη δομή του χώρου αναζήτησης, η οποία αποδείχθηκε κοινή και στις τέσσερις πηγές: τα δεδομένα δεν είναι προσπελάσιμα ως ενιαίο χρονοδιάγραμμα αλλά μόνο ως απάντηση σε ερώτημα της μορφής «αφετηρία, προορισμός, ημερομηνία». Η εξαντλητική συλλογή απαιτεί, συνεπώς, τη διάσχιση ολόκληρου του καρτεσιανού γινομένου, και για τον σκοπό αυτό χρειάστηκε να εντοπιστούν επιπλέον δύο βοηθητικά σημεία πρόσβασης ανά πηγή: εκείνο που απαριθμεί τις αφετηρίες και εκείνο που, για δεδομένη αφετηρία, περιορίζει τους προορισμούς σε όσους πράγματι εξυπηρετούνται. Ο περιορισμός αυτός δεν είναι δευτερεύων· χωρίς αυτόν το πλήθος των συνδυασμών θα ήταν τετραγωνικό ως προς τις στάσεις κάθε φορέα και η συλλογή πρακτικά ανέφικτη.

Δύο πηγές παρουσίασαν πρόσθετες απαιτήσεις κατάστασης. Η πλατφόρμα e-analysis απαιτεί ανάκτηση της σελίδας και εξαγωγή των κρυφών πεδίων κατάστασης πριν από κάθε υποβολή, ώστε κάθε ερώτημα να συνίσταται στην πράξη σε δύο διαδοχικά αιτήματα. Η πλατφόρμα 3pi-teleomatics, χωρίς να το απαιτεί ρητά, επιστρέφει αξιόπιστα αποτελέσματα μόνο εφόσον προηγηθεί επίσκεψη στην αρχική σελίδα ώστε να αποδοθούν τα σχετικά cookies, και εφόσον τα αιτήματα φέρουν κεφαλίδες συμβατές με εκείνες ενός περιηγητή. Οι δύο αυτές παρατηρήσεις τεκμηριώνουν με τον πιο άμεσο τρόπο τη διαπίστωση του πρώτου κεφαλαίου: οι υπηρεσίες αυτές δεν σχεδιάστηκαν ποτέ για κατανάλωση από τρίτους.

4.2.2 Αποκρυπτογράφηση των αποκρίσεων του IME API

Η πλατφόρμα IME παρουσίασε ένα πρόσθετο εμπόδιο. Το σώμα του αιτήματος δεν περιέχει τις παραμέτρους της αναζήτησης αλλά ένα μοναδικό πεδίο με μια συμβολοσειρά Base64, ενώ η απόκριση έχει ακριβώς την ίδια μορφή. Η καταγραφή της κίνησης, επομένως, αποκαλύπτει μεν τη διεύθυνση της υπηρεσίας αλλά τίποτε από το περιεχόμενο της επικοινωνίας, και η αποκωδικοποίηση Base64 δίνει δυαδικά δεδομένα χωρίς αναγνωρίσιμη δομή — ένδειξη κρυπτογράφησης και όχι απλής συσκότισης.

Η άρση του εμποδίου στηρίχθηκε σε μια απλή παρατήρηση: εφόσον ο κώδικας που εκτελείται στον περιηγητή κατασκευάζει ο ίδιος τα κρυπτογραφημένα σώματα και αποκρυπτογραφεί τις αποκρίσεις, οφείλει να περιέχει και το κλειδί. Ο κώδικας αυτός αποστέλλεται αυτούσιος σε κάθε επισκέπτη και είναι εξ ορισμού αναγνώσιμος. Η επιθεώρησή του ανέδειξε κλήσεις προς καθιερωμένη κρυπτογραφική βιβλιοθήκη JavaScript και αποκάλυψε το σχήμα: αλγόριθμος AES σε λειτουργία CBC, κλειδί δεκαέξι χαρακτήρων ενσωματωμένο ως σταθερά, διάνυσμα αρχικοποίησης μηδενικό και συμπλήρωση κατά PKCS#7. Η αναπαραγωγή σε Python περιορίστηκε σε δύο συναρτήσεις οκτώ γραμμών συνολικά, οι οποίες παρεμβάλλονται διαφανώς σε κάθε κλήση, ενώ η επιβεβαίωση της ορθότητας ήταν άμεση: η αποκρυπτογράφηση μιας καταγεγραμμένης απόκρισης παρήγαγε έγκυρο JSON με ακριβώς τα δεδομένα που εμφάνιζε η σελίδα.

Για την εξερεύνηση των διαθέσιμων σημείων πρόσβασης αναπτύχθηκε χωριστό βοηθητικό πρόγραμμα, το οποίο εκτυπώνει και αποθηκεύει, για κάθε κλήση, τόσο το αρχικό όσο και το κρυπτογραφημένο σώμα και τόσο την ακατέργαστη όσο και την αποκρυπτογραφημένη απόκριση. Το εργαλείο αυτό λειτούργησε ως εργαστήριο κατανόησης της υπηρεσίας πριν από τη συγγραφή του μηχανισμού συλλογής και τα αποτελέσματά του διατηρήθηκαν ως τεκμηρίωση. Οι ευρύτερες

συνέπειες του ευρήματος —ότι ένα σταθερό κλειδί στην πλευρά του πελάτη συνιστά συσκότιση και όχι ασφάλεια— συζητήθηκαν στην ενότητα 3.9.

4.3 Σχεδίαση και υλοποίηση των scrapers

Παρά τις τεχνολογικές τους διαφορές, οι τέσσερις μηχανισμοί συλλογής μοιράζονται κοινή αρχιτεκτονική τριών φάσεων, η οποία προκύπτει άμεσα από τη δομή του χώρου αναζήτησης που περιγράφηκε παραπάνω. Η πρώτη φάση ανακτά τον κατάλογο των αφετηριών του φορέα. Η δεύτερη ανακτά, για καθεμία από αυτές, τους πράγματι εξυπηρετούμενους προορισμούς, συγκροτώντας το σύνολο των έγκυρων ζευγών. Η τρίτη, η οποία απορροφά πάνω από το ενενήντα εννέα τοις εκατό του χρόνου εκτέλεσης, διατρέχει το γινόμενο των ζευγών επί τις ημερομηνίες δειγματοληψίας και υποβάλλει ένα ερώτημα για κάθε συνδυασμό. Κοινή είναι επίσης η έξοδος, καθώς και τα τέσσερα προγράμματα παράγουν εγγραφές πανομοιότυπου σχήματος, και η διεπαφή γραμμής εντολών, η οποία επιτρέπει τον περιορισμό σε συγκεκριμένα ΚΤΕΛ ή ημερομηνίες και την επιλογή μεταξύ παράκαμψης, συγχώνευσης και πλήρους ανανέωσης των υπαρχόντων αρχείων. Η διαφοροποίηση εντοπίζεται στον τρόπο με τον οποίο επιτυγχάνεται ο ταυτοχρονισμός, ο οποίος αποτελεί και τον καθοριστικό παράγοντα του χρόνου εκτέλεσης, καθώς ολόκληρος ο φόρτος είναι δεσμευμένος από την αναμονή του δικτύου και όχι από τον επεξεργαστή.

4.3.1 Ασύγχρονη συλλογή με `asyncio` και έλεγχος ταυτοχρονισμού

Για την πλατφόρμα IME, όπου το πλήθος των κλήσεων ανέρχεται σε δεκάδες χιλιάδες, υιοθετήθηκε το ασύγχρονο υπόδειγμα. Και οι τρεις φάσεις εκφράζονται ως συνρουτίνες: οι κλήσεις προσπελάσιμων προορισμών όλων των αφετηριών ενός ΚΤΕΛ εκκινούν ταυτόχρονα και συλλέγονται μαζικά, ενώ στην τρίτη φάση κατασκευάζεται μία συνρουτίνα ανά τριάδα αφετηρίας, προορισμού και ημερομηνίας, και τα αποτελέσματα καταναλώνονται με τη σειρά που ολοκληρώνονται, ώστε η ένδειξη προόδου να ενημερώνεται συνεχώς. Επιπλέον, τα ΚΤΕΛ δεν εκτελούνται σειριακά αλλά επίσης ταυτόχρονα, καθώς πρόκειται για ανεξάρτητους εξυπηρετητές.

Ο έλεγχος του ταυτοχρονισμού ασκείται με δύο επίπεδα σημαφόρων. Ο καθολικός σημαφόρος περιορίζει σε πενήντα τα ταυτόχρονα ανοικτά αιτήματα, προστατεύοντας τον τοπικό υπολογιστή από εξάντληση υποδοχών και συνοδεύεται από αντίστοιχο όριο συνδέσεων στον συνδετήρα του πελάτη. Ο σημαφόρος ανά ΚΤΕΛ περιορίζει σε δέκα τα ταυτόχρονα αιτήματα προς τον ίδιο φορέα, ώστε ένας μεγάλος φορέας να μην μονοπωλεί το καθολικό όριο και, κυρίως, ώστε να μην επιβαρύνεται δυσανάλογα μία μεμονωμένη υποδομή. Ορίζεται τέλος συνολικό χρονικό όριο τριάντα δευτερολέπτων ανά αίτημα, ώστε μια κλήση που δεν απαντά ποτέ να μην δεσμεύει επ' αόριστον θέση στον σημαφόρο. Το τελικό αποτέλεσμα είναι ότι ο χρόνος συλλογής μιας πηγής με περισσότερες από ενενήντα χιλιάδες εγγραφές μετρήθηκε σε ώρες αντί για ημέρες, χωρίς αισθητή επιβάρυνση των εξυπηρετητών.

4.3.2 Πολυνηματική συλλογή με `ThreadPoolExecutor`

Για τις υπόλοιπες πηγές, όπου ο όγκος είναι μικρότερος κατά μία τάξη μεγέθους, η ασύγχρονη αναδιατύπωση δεν δικαιολογούνταν και προτιμήθηκε η κλασική πολυνηματική εκτέλεση. Στην

πλατφόρμα 3pi-telematics η μονάδα παραλληλισμού ορίστηκε ως το ζεύγος αφετηρίας και ημέρας: κάθε εργασία αναλαμβάνει μία αφετηρία για μία ημερομηνία και διατρέχει σειριακά όλους τους προορισμούς της. Η επιλογή αυτή έχει ένα σημαντικό πλεονέκτημα έναντι της παραλληλοποίησης ανά μεμονωμένο ερώτημα: κάθε νήμα διατηρεί δική του συνεδρία με δικά της cookies, τα οποία, όπως αναφέρθηκε, απαιτούνται από την πηγή, και η δημιουργία τους αποσβένεται σε δεκάδες ερωτήματα αντί σε ένα.

Το πλήθος των νημάτων ορίστηκε συντηρητικά σε τέσσερα, τιμή που προέκυψε εμπειρικά ως το σημείο πέραν του οποίου εμφανίζονταν σφάλματα από την πλευρά του εξυπηρετητή. Ξεχωριστή μεταχείριση επιφυλάχθηκε στο ΚΤΕΛ Χαλκιδικής, το οποίο διαθέτει δυσανάλογα μεγάλο πλήθος στάσεων και, εκτελούμενο σειριακά μαζί με τα υπόλοιπα, καθόριζε μόνο του τον συνολικό χρόνο· εκκινεί επομένως εξ αρχής σε δικό του νήμα και εξελίσσεται παράλληλα προς τη σειριακή επεξεργασία των υπολοίπων. Αντιθέτως, για τις πηγές e-analysis και ΚΤΕΛ Ηλείας η συλλογή παρέμεινε αυστηρά σειριακή, όχι από παράλειψη αλλά επειδή η εξάρτηση κάθε υποβολής από τα πεδία κατάστασης της αμέσως προηγούμενης ανάκτησης καθιστά τα διαδοχικά ερωτήματα εξαρτημένα μεταξύ τους εντός της ίδιας συνεδρίας.

4.3.3 Πολιτική επαναλήψεων, rate limiting και καταγραφή σφαλμάτων

Σε μια διαδικασία δεκάδων χιλιάδων δικτυακών κλήσεων τα σφάλματα δεν αποτελούν εξαίρεση αλλά βεβαιότητα, και η ανθεκτικότητα απέναντί τους σχεδιάστηκε ρητά. Κάθε κλήση επαναλαμβάνεται έως τρεις φορές με εκθετική υποχώρηση, ώστε ένα παροδικό σφάλμα δικτύου ή μια στιγμιαία υπερφόρτωση του εξυπηρετητή να μην οδηγεί σε απώλεια δεδομένων· η αναμονή διπλασιάζεται σε κάθε αποτυχία, δίνοντας στην απομακρυσμένη υπηρεσία χρόνο να ανακάμψει αντί να επιδεινώνεται η κατάσταση με άμεσες επαναλήψεις. Στους μηχανισμούς που δεν χρησιμοποιούν σημαφόρους, ο ρυθμός ελέγχεται επιπλέον με σταθερή καθυστέρηση τριών δεκάτων του δευτερολέπτου μεταξύ διαδοχικών κλήσεων.

Θεμελιώδης σχεδιαστική αρχή υπήρξε η απομόνωση των αστοχιών. Μια αποτυχία αφορά πάντοτε ένα μεμονωμένο ζεύγος και ημερομηνία και ουδέποτε διακόπτει τη συνολική εκτέλεση: το ζεύγος καταγράφεται και η διαδικασία συνεχίζει. Η καταγραφή γίνεται σε χωριστό αρχείο ανά φορέα και περιλαμβάνει την ημέρα, τα ονόματα των δύο στάσεων και τον τύπο του σφάλματος για καθεμία από τις τρεις προσπάθειες, ώστε η αποτυχημένη περιοχή να μπορεί να επαναληφθεί στοχευμένα με τις παραμέτρους της γραμμής εντολών. Η αποτελεσματικότητα της πολιτικής αποτυπώνεται στο αποτέλεσμα της τελικής εκτέλεσης: από το σύνολο των ΚΤΕΛ, μόνο δύο άφησαν αρχείο σφαλμάτων, καθένα με μία μοναδική εγγραφή που αφορούσε επίμονο εσωτερικό σφάλμα του απομακρυσμένου εξυπηρετητή, επαναλαμβανόμενο και στις τρεις προσπάθειες.

4.4 Στρατηγική δειγματοληψίας χρόνου και εξαγωγή εβδομαδιαίου προτύπου

Η σημαντικότερη εννοιολογική δυσκολία της συλλογής δεν είναι τεχνική αλλά αφορά την ίδια τη φύση των πηγών. Καμία από αυτές δεν δημοσιεύει χρονοδιάγραμμα με την αυστηρή έννοια, δηλαδή κατάλογο δρομολογίων συνοδευόμενο από τους κανόνες εκτέλεσής τους. Όλες λειτουργούν ως μηχανές κράτησης που απαντούν αποκλειστικά σε ερωτήματα συγκεκριμένης

ημερομηνίας. Η πληροφορία «το δρομολόγιο των 07:30 εκτελείται καθημερινά εκτός Κυριακής» δεν είναι πουθενά διαθέσιμη· υπάρχει μόνο έμμεσα, ως το γεγονός ότι η αναχώρηση εμφανίζεται σε έξι από τις επτά ημέρες.

Η στρατηγική που υιοθετήθηκε ανακατασκευάζει το εβδομαδιαίο πρότυπο ακριβώς με αυτόν τον τρόπο: κάθε ζεύγος στάσεων ερωτάται επτά φορές, μία για κάθε ημέρα μιας πλήρους εβδομάδας, και οι επτά απαντήσεις συγκεντρώνονται σε ενιαία εγγραφή. Το κλειδί της συγκέντρωσης είναι η τριάδα αφετηρίας, προορισμού και ώρας αναχώρησης, ενώ οι ημέρες κατά τις οποίες εμφανίστηκε η ίδια τριάδα συλλέγονται σε σύνολο που αποτελεί την κωδικοποίηση της εβδομαδιαίας εκτέλεσης. Η επιλογή της συγκεκριμένης εβδομάδας δεν είναι τυχαία: δειγματοληπτείται εκείνη που ξεκινά δύο Δευτέρες μετά την τρέχουσα. Η απόσταση αυτή είναι επαρκής ώστε να έχουν αποφευχθεί τα φαινόμενα της άμεσης προσεχούς περιόδου, όπου δρομολόγια ενδέχεται να μην εμφανίζονται λόγω εξάντλησης θέσεων ή κλεισίματος των κρατήσεων, και ταυτόχρονα αρκετά μικρή ώστε τα προγράμματα να έχουν ήδη δημοσιευτεί από όλους τους φορείς.

Η προσέγγιση έχει δύο αναγνωρισμένους περιορισμούς. Πρώτον, το πρότυπο που εξάγεται ισχύει για τη συγκεκριμένη περίοδο του έτους και δεν αποτυπώνει τη θερινή ή χειμερινή διαφοροποίηση ούτε τις εξαιρέσεις των αργιών, ζήτημα στο οποίο επανέρχεται το έβδομο κεφάλαιο. Δεύτερον, ορισμένοι φορείς δέχονται κρατήσεις μόνο για λίγες ημέρες μπροστά, οπότε η εβδομάδα-στόχος επιστρέφει κενή. Για τις περιπτώσεις αυτές προβλέπεται λειτουργία συγχώνευσης, στην οποία μια νέα εκτέλεση για συγκεκριμένες ημερομηνίες θεωρείται αυθεντική μόνο για τις ημέρες που πράγματι επέστρεψαν δεδομένα και αντικαθιστά πλήρως τις αντίστοιχες εγγραφές, αφήνοντας τις υπόλοιπες ημέρες ανέπαφες. Η ρητή δικλείδα ότι μια ημέρα με κενή απόκριση δεν θεωρείται ποτέ αυθεντική αποτρέπει το ενδεχόμενο ένα προσωρινό σφάλμα να διαγράψει έγκυρα δεδομένα, ώστε η τμηματική ανανέωση να μπορεί να επαναλαμβάνεται ελεύθερα χωρίς κίνδυνο σταδιακής διάβρωσης του συνόλου.

4.5 Ενδιάμεση μορφή αποθήκευσης και σχήμα των JSON αρχείων

Ανάμεσα στη συλλογή και στη βάση δεδομένων παρεμβάλλεται ένα ενδιάμεσο επίπεδο αρχείων JSON, το οποίο αποτελεί και το σημείο όπου παύει η ετερογένεια των πηγών. Κάθε φορέας αποτυπώνεται σε ένα αρχείο, του οποίου το όνομα φέρει πρόθεμα δηλωτικό της πηγής και επίθημα το όνομα του ΚΤΕΛ σε γενική πτώση, ώστε η προέλευση κάθε εγγραφής να είναι ανακτήσιμη από το ίδιο το όνομα του αρχείου χωρίς πρόσθετα μεταδεδομένα. Το περιεχόμενο είναι πίνακας αντικειμένων με έξι πεδία: τα ονόματα της αφετηρίας και του προορισμού όπως ακριβώς τα δημοσιεύει η πηγή, την ώρα αναχώρησης και τη διάρκεια της διαδρομής σε μορφή ωρών και λεπτών, μια ένδειξη αστικού δρομολογίου και τον πίνακα των ημερών εβδομαδιαίας εκτέλεσης με τα ελληνικά τους ονόματα.

Χωριστό αρχείο ανά πηγή τηρεί τις στάσεις, οργανωμένες σε δύο επίπεδα: πρώτα κατά ΚΤΕΛ και έπειτα κατά όνομα στάσης, με τιμή το ζεύγος των συντεταγμένων ή δύο κενές τιμές όταν αυτές δεν είναι ακόμη γνωστές. Η δομή αυτή δεν είναι απλώς αποθηκευτική· ενσωματώνει από μόνη της την αρχή που θα αποδειχθεί κρίσιμη στην επόμενη ενότητα, ότι δηλαδή η ταυτότητα μιας στάσης δεν είναι το όνομά της αλλά το όνομα σε συνδυασμό με τον φορέα και την πηγή που το δημοσιεύει.

Δύο συμβάσεις τηρούνται καθολικά. Η σειριοποίηση εκτελείται με απενεργοποιημένη τη διαφυγή μη λατινικών χαρακτήρων και με κωδικοποίηση UTF-8, ώστε τα ελληνικά τοπωνύμια να παραμένουν άμεσα αναγνώσιμα και τα αρχεία να μπορούν να ελεγχθούν οπτικά ή να διορθωθούν χειροκίνητα. Τα ονόματα των στάσεων, εξάλλου, διατηρούνται αυτούσια, χωρίς καμία κανονικοποίηση: το επίπεδο αυτό αποτελεί πιστό στιγμιότυπο των πρωτογενών δεδομένων και κάθε μετασχηματισμός τους μετατίθεται σε επόμενα, αναστρέψιμα στάδια. Η αρχή αυτή αποδείχθηκε πολύτιμη στην πράξη, καθώς επέτρεψε την επανειλημμένη αναθεώρηση των κανόνων ενοποίησης χωρίς να απαιτηθεί ποτέ νέα προσφυγή στις πηγές.

4.6 Ενοποίηση οντοτήτων στάσεων

Με την ολοκλήρωση της συλλογής, το σύνολο των δεδομένων περιλαμβάνει περισσότερες από τρεις χιλιάδες καταχωρίσεις στάσεων, καθεμία προσδιοριζόμενη από την τριάδα πηγής, φορέα και ονόματος. Οι καταχωρίσεις αυτές δεν αντιστοιχούν σε ισάριθμα φυσικά σημεία, και η αποκατάσταση της αντιστοιχίας αποτελεί το κρισιμότερο βήμα ολόκληρου του κεφαλαίου: χωρίς αυτήν, το δίκτυο παραμένει ένα σύνολο ασύνδετων νησίδων, μία ανά φορέα, και καμία ανταπόκριση δεν είναι δυνατόν να εντοπιστεί. Το θεωρητικό πλαίσιο του προβλήματος διατυπώθηκε στην ενότητα 2.8· εδώ περιγράφεται η πρακτική του επίλυση, η οποία συνδυάζει τέσσερα διαδοχικά και συμπληρωματικά μέσα.

4.6.1 Το πρόβλημα των συνωνύμων και των ομωνύμων

Η έκταση του προβλήματος γίνεται σαφής μόνο μέσα από τα ίδια τα δεδομένα. Η πόλη της Καρδίτσας εμφανίζεται ως «ΚΑΡΔΙΤΣΑ(ΠΑΛΙΟ Κ.Τ.Ε.Λ)» και «ΚΑΡΔΙΤΣΑ_ΚΤΕΛ» στην πλατφόρμα ΙΜΕ, ως «Καρδίτσα» στην 3pi-telematics και ως «ΚΑΡΔΙΤΣΑ» στην e-analysis, ενώ το ίδιο όνομα με το ίδιο ακριβώς επίθημα καταχωρίζεται ανεξάρτητα από δύο διαφορετικά ΚΤΕΛ. Στην περίπτωση της Κατερίνης συναντώνται επιπλέον η μορφή «Δ.ΚΑΤΕΡΙΝΗΣ», η μορφή «ΚΤΕΛ ΠΙΕΡΙΑΣ_ΚΑΤΕΡΙΝΗ» και δύο εκδοχές που διαφέρουν μόνο κατά έναν τελικό κενό χαρακτήρα. Ο σταθμός διοδίων των Αφιδνών, τέλος, εμφανίζεται άλλοτε στον ενικό και άλλοτε στον πληθυντικό, μία φορά με ορθογραφικό λάθος, μία με το όνομα γειτονικού οικισμού ως πρόθεμα και μία, από συγκεκριμένο φορέα, ως «ΑΘΗΝΑ» — καταχώριση εμπορικά κατανοητή, αφού εκεί καταλήγει το δρομολόγιο προς την πρωτεύουσα, αλλά γεωγραφικά εσφαλμένη. Είναι προφανές ότι καμία λεξιλογική μέθοδος δεν πρόκειται να αντιμετωπίσει επιτυχώς αυτή την ποικιλία.

Ταυτόχρονα ισχύει και το αντίστροφο. Ονόματα όπως «Νεοχώρι», «Άγιος Γεώργιος» ή «Καστανιά» απαντούν αυτούσια σε πολλούς νομούς και αντιστοιχούν σε εντελώς διαφορετικά σημεία. Όπως επισημάνθηκε στην ενότητα 2.8.1, οι δύο αστοχίες δεν είναι ισοδύναμες: η αποτυχία ενοποίησης συνωνύμων απλώς αποκρύπτει υπαρκτές διαδρομές, ενώ η εσφαλμένη ενοποίηση ομωνύμων κατασκευάζει διαδρομές που δεν υπάρχουν και θα οδηγούσαν τον χρήστη σε λάθος άκρη της χώρας. Η ασυμμετρία αυτή υπαγόρευσε ολόκληρη τη σχεδίαση της λύσης, η οποία προτιμά συστηματικά τη διατήρηση μιας διάκρισης έναντι μιας αβέβαιης συγχώνευσης.

4.6.2 Γεωκωδικοποίηση στάσεων μέσω Nominatim

Το ανεξάρτητο κριτήριο που απαιτείται για την άρση της αμφισημίας είναι η γεωγραφική θέση. Η πλατφόρμα IME παρέχει συντεταγμένες μαζί με τον κατάλογο των στάσεων, οπότε το πρόβλημα περιορίζεται στις τρεις υπόλοιπες πηγές, οι οποίες αθροίζουν 345 καταχωρίσεις. Για αυτές αναπτύχθηκε πρόγραμμα γεωκωδικοποίησης που αξιοποιεί την υπηρεσία Nominatim με στρατηγική δύο σταδίων. Στο πρώτο, το ερώτημα περιορίζεται στην ελληνική επικράτεια και ταυτόχρονα εντός πλαισίου προβολής ενάμισι μοίρας γύρω από το κέντρο του νομού στον οποίο ανήκει ο φορέας της στάσης, με ρητή απαίτηση να μην επιστραφούν αποτελέσματα εκτός αυτού. Αν το πρώτο στάδιο αποτύχει, ακολουθεί δεύτερο χωρίς τον αυστηρό περιορισμό, με το πλαίσιο να λειτουργεί πλέον μόνο ως ένδειξη προτίμησης.

Το πρώτο στάδιο συνιστά την άμεση εφαρμογή της αρχής του γεωγραφικού συμφραζομένου της ενότητας 2.8.3: η ταυτότητα του φορέα, σε συνδυασμό με έναν πίνακα κέντρων νομών, μετατρέπει ένα αμφίσημο τοπωνύμιο σε ερώτημα με μονοσήμαντη απάντηση. Δύο επιπλέον φίλτρα ελέγχουν την ποιότητα του αποτελέσματος: γίνονται δεκτά μόνο αποτελέσματα που η υπηρεσία κατατάσσει ως οικισμούς ή διοικητικές ενότητες, και επιβεβαιώνεται ότι οι συντεταγμένες εμπίπτουν στο περίγραμμα της χώρας. Η πολιτική χρήσης της δημόσιας εγκατάστασης τηρείται με αναμονή ενάμισι δευτερολέπτου μεταξύ κλήσεων και αναγνωρίσιμη δήλωση πράκτορα χρήστη, ενώ η πρόοδος αποθηκεύεται ανά είκοσι στάσεις ώστε μια διακοπή να μην ακυρώνει την έως τότε εργασία και η επανεκτέλεση να επεξεργάζεται μόνο τις εκκρεμείς. Όσες στάσεις δεν εντοπίστηκαν αυτόματα απαριθμούνται ρητά στο τέλος της εκτέλεσης και συμπληρώθηκαν χειροκίνητα με το εργαλείο της ενότητας 4.6.4, ώστε στο τελικό σύνολο δεδομένων καμία στάση να μην στερείται συντεταγμένων.

4.6.3 Χωρική ομαδοποίηση γειτονικών στάσεων

Με τις συντεταγμένες διαθέσιμες, οι καταχωρίσεις και των τεσσάρων πηγών συγχωνεύονται σε ενιαίο κατάλογο σημείων και ομαδοποιούνται με κριτήριο απόκλισης απλής σύνδεσης: δύο καταχωρίσεις τοποθετούνται στην ίδια ομάδα όταν απέχουν λιγότερο από ένα χιλιοστό της μοίρας, απόσταση που αντιστοιχεί σε περίπου εκατό μέτρα στα ελληνικά γεωγραφικά πλάτη, και η σχέση αυτή επεκτείνεται μεταβατικά σε ολόκληρη την ομάδα. Το κατώφλι επιλέχθηκε ώστε να συλλαμβάνει τις πραγματικές περιπτώσεις — τον ίδιο σταθμό γεωκωδικοποιημένο με μικρή απόκλιση από διαφορετικές πηγές — χωρίς να συγχωνεύει γειτονικά αλλά διακριτά σημεία εξυπηρέτησης μιας πόλης.

Η μεταβατικότητα της απλής σύνδεσης αποτελεί γνωστό αδύναμο σημείο της μεθόδου, καθώς μια αλυσίδα ενδιάμεσων σημείων μπορεί θεωρητικά να συνενώσει δύο άκρα που απέχουν πολύ περισσότερο από το κατώφλι. Στην πράξη το φαινόμενο περιορίζεται από τη μικρή τιμή του κατωφλίου και, κυρίως, εξουδετερώνεται από το γεγονός ότι η ομαδοποίηση δεν λαμβάνει καμία τελική απόφαση. Παράγει υποψήφιες ομάδες προς επικύρωση, μεταθέτοντας την κρίση στο επόμενο στάδιο, και ακριβώς αυτός ο διαχωρισμός ανάμεσα στην αυτόματη πρόταση και στην ανθρώπινη επιβεβαίωση καθιστά τη διαδικασία ασφαλή.

4.6.4 Διαδραστικό εργαλείο χάρτη για ημιαυτόματη επικύρωση

Για την επικύρωση αναπτύχθηκε εσωτερικό εργαλείο, το οποίο εκκινεί τοπικό εξυπηρετητή και αποδίδει στον περιηγητή διαδραστικό χάρτη με το σύνολο των στάσεων, χρωματισμένων ανά πηγή προέλευσης και ομαδοποιημένων ανάλογα με το επίπεδο μεγέθυνσης, ώστε ο χάρτης να παραμένει αποκρίσιμος παρά το πλήθος των σημείων. Ο χειριστής επιλέγει μια ομάδα, βλέπει συγκεντρωμένες όλες τις καταχωρίσεις που της αντιστοιχούν με την πηγή και τον φορέα καθεμιάς, και αποδίδει σε αυτές ένα κοινό κανονικό όνομα ή, αντιστρόφως, εξαιρεί όσες δεν ανήκουν πράγματι εκεί. Το εργαλείο υποστηρίζει επιπλέον τη διόρθωση εσφαλμένων συντεταγμένων, είτε με επικόλληση συνδέσμου από δημοφιλή χαρτογραφική υπηρεσία, από τον οποίο εξάγονται αυτόματα οι συντεταγμένες, είτε με απευθείας επιλογή σημείου πάνω στον χάρτη· η αλλαγή γράφεται αμέσως πίσω στο αντίστοιχο αρχείο στάσεων.

Η επιλογή της ημιαυτόματης προσέγγισης έναντι μιας πλήρως αυτόματης είναι συνειδητή και στηρίζεται σε δύο επιχειρήματα. Το πρώτο είναι ότι, όπως τεκμηριώθηκε, κανένα κατώφλι δεν είναι ταυτόχρονα ασφαλές και ως προς τα συνώνυμα και ως προς τα ομώνυμα, οπότε η αυτόματη απόφαση θα εισήγαγε αναπόφευκτα σφάλματα ακριβώς του είδους που έχει το υψηλότερο κόστος. Το δεύτερο είναι ζήτημα κλίμακας: το πλήθος των ομάδων που απαιτούν κρίση είναι της τάξης των μερικών εκατοντάδων, μέγεθος απολύτως διαχειρίσιμο από άνθρωπο σε λίγες ώρες εργασίας, ενώ η οπτική επιβεβαίωση πάνω σε χάρτη είναι σχεδόν ακαριαία και εξαιρετικά αξιόπιστη.

4.6.5 Το αρχείο canonical stops και η context-aware αντιστοίχιση

Το προϊόν της διαδικασίας είναι ένα αρχείο που συγκεντρώνει, για κάθε κανονική οντότητα, το όνομά της και τον πλήρη κατάλογο των επιμέρους καταχωρίσεων που αντιστοιχούν σε αυτήν, με καθεμία να προσδιορίζεται από την τριάδα πηγής, φορέα και ονόματος. Η χρήση της τριάδας ως κλειδιού, και όχι του ονόματος, αποτελεί το αποφασιστικό στοιχείο ολόκληρης της προσέγγισης: δύο ομώνυμες στάσεις διαφορετικών φορέων συνιστούν εξ ορισμού διαφορετικές καταχωρίσεις και δεν πρόκειται ποτέ να συγχυστούν, ακόμη και αν τα ονόματά τους ταυτίζονται χαρακτήρα προς χαρακτήρα. Κατά την κατασκευή του γράφου, το αρχείο μετατρέπεται σε λεξικό αντιστοίχισης και κάθε άκρο κάθε δρομολογίου μεταφράζεται στην κανονική του μορφή.

Καθοριστική είναι και η συμπεριφορά σε περίπτωση αστοχίας. Όταν μια τριάδα δεν βρίσκεται στο λεξικό, δεν αγνοείται ούτε αντιστοιχίζεται με βάση την ομοιότητα του ονόματος· λαμβάνει αντ' αυτού ένα σύνθετο αναγνωριστικό που συνενώνει το αρχικό όνομα με τον φορέα προέλευσης. Η στάση παραμένει έτσι πλήρως λειτουργική εντός του δικτύου του φορέα της, χωρίς να συγχωνευτεί ποτέ εσφαλμένα με ομώνυμη άλλου φορέα. Η αστοχία, δηλαδή, οδηγεί πάντοτε στη λιγότερο επιζήμια κατεύθυνση, εκείνη της χαμένης ανταπόκρισης και όχι της ανύπαρκτης σύνδεσης.

Η επιλογή αυτή επέτρεψε και μια ορθολογική ιεράρχηση της χειρωνακτικής προσπάθειας. Η ενοποίηση έχει νόημα μόνο για τις στάσεις που εξυπηρετούνται από περισσότερους του ενός φορείς, δηλαδή ακριβώς για τους δυνητικούς κόμβους μετεπιβίβασης· μια στάση χωριού που εξυπηρετείται από ένα μόνο ΚΤΕΛ δεν χρειάζεται καμία ενοποίηση. Η εργασία επικεντρώθηκε συνεπώς εκεί, με αποτέλεσμα 353 κανονικές οντότητες που συγκεντρώνουν 1.157 επιμέρους καταχωρίσεις. Από αυτές, 239 συνενώνουν στάσεις περισσότερων του ενός ΚΤΕΛ και 93 στάσεις περισσότερων της μίας πηγής, ενώ η μεγαλύτερη οντότητα, η Θεσσαλονίκη, συγκεντρώνει 32

καταχωρίσεις προερχόμενες από 31 διαφορετικά ΚΤΕΛ και από τις τέσσερις πηγές — αριθμός που αποτυπώνει με τον πιο εύγλωττο τρόπο την έκταση του αρχικού κατακερματισμού.

4.7 Συμπλήρωση ελλειπών χρόνων διαδρομής με OSRM

Δύο από τις τέσσερις πηγές δημοσιεύουν αποκλειστικά ώρα αναχώρησης, χωρίς ώρα άφιξης ή διάρκεια. Το κενό αυτό δεν είναι διακοσμητικό: όπως τεκμηριώθηκε στην ενότητα 2.9.6, ένα σκέλος άγνωστης διάρκειας δεν μπορεί να χρησιμοποιηθεί ως ενδιάμεσο σε σύνθετη διαδρομή, διότι ο χρόνος άφιξης παραμένει απροσδιόριστος και η αναζήτηση δεν μπορεί να συνεχιστεί από τον κόμβο αυτόν. Τα δρομολόγια των δύο αυτών πηγών θα ήταν, δηλαδή, αξιοποιήσιμα μόνο ως τελευταία σκέλη, γεγονός που θα ακύρωνε μεγάλο μέρος της αξίας τους.

Η συμπλήρωση στηρίχθηκε στη μηχανή δρομολόγησης OSRM. Το πρόγραμμα εντοπίζει πρώτα τα μοναδικά ζεύγη στάσεων που στερούνται διάρκειας, ώστε να μην επαναλαμβάνεται η ίδια κλήση για δεκάδες αναχωρήσεις της ίδιας σύνδεσης, ανακτά τις συντεταγμένες τους από τα αρχεία στάσεων και ερωτά την υπηρεσία για τον χρόνο οδικής μετακίνησης μεταξύ τους, ζητώντας ρητά μόνο τη διάρκεια και όχι τη γεωμετρία της διαδρομής. Το αποτέλεσμα στρογγυλοποιείται σε λεπτά και μετατρέπεται στη μορφή ωρών και λεπτών που χρησιμοποιεί το υπόλοιπο σύστημα, ενώ προβλέπονται τρεις προσπάθειες ανά κλήση. Συμπληρωματικά, όταν μια σύνδεση έχει έστω και μία αναχώρηση με γνωστή διάρκεια, η τιμή αυτή διαδίδεται στις υπόλοιπες αναχωρήσεις του ίδιου ζεύγους, μηχανισμός που περιορίζει σημαντικά τις κλήσεις προς την εξωτερική υπηρεσία.

Ο συντελεστής προσαρμογής από επιβατικό όχημα σε λεωφορείο διατηρήθηκε ρητά στη μονάδα, δηλαδή δεν εφαρμόζεται καμία διόρθωση. Η επιλογή τεκμηριώνεται ως εξής: η εκτίμηση αφορά διαδρομή από σταθμό σε σταθμό μέσω του οδικού δικτύου, ενώ το υπεραστικό λεωφορείο κινείται κατά κύριο λόγο σε αυτοκινητοδρόμους με ταχύτητα συγκρίσιμη του επιβατικού οχήματος· ενδεχόμενος πολλαπλασιαστής θα ήταν αυθαίρετος και θα εισήγαγε συστηματικό σφάλμα αντί να το διορθώσει. Η προσέγγιση παραμένει ασφαλώς ατελής, καθώς αγνοεί τις ενδιάμεσες στάσεις και τους χρόνους παραμονής σε αυτές, και για τον λόγο αυτό αποτιμάται χωριστά στην αξιολόγηση του έκτου κεφαλαίου. Τα συμπληρωμένα πεδία, εξάλλου, παραμένουν αναγνωρίσιμα εκ των υστέρων, αφού προέρχονται αποκλειστικά από τις δύο πηγές που δεν δημοσιεύουν ώρα άφιξης, και μπορούν έτσι να αποτιμηθούν χωριστά από τα πρωτογενή. Το πρακτικό αποτέλεσμα, πάντως, είναι ότι στο τελικό σύνολο δεδομένων καμία εγγραφή δεν στερείται διάρκειας διαδρομής και κάθε δρομολόγιο είναι αξιοποιήσιμο σε οποιαδήποτε θέση μιας σύνθετης διαδρομής.

4.8 Σχεδίαση της σχεσιακής βάσης δεδομένων

4.8.1 Σχήμα, ευρετήρια και περιορισμοί

Το σχήμα της βάσης οργανώνεται γύρω από έναν και μοναδικό πίνακα δρομολογίων. Κάθε εγγραφή περιλαμβάνει την πηγή και τον φορέα προέλευσης, τα ονόματα της αφετηρίας και του προορισμού όπως τα δημοσιεύει η πηγή, την ώρα αναχώρησης, τη διάρκεια, την ένδειξη αστικού δρομολογίου, τις ημέρες εβδομαδιαίας εκτέλεσης, το αρχείο από το οποίο προήλθε και τη χρονική σήμανση της τελευταίας φόρτωσής της. Η επιλογή ενός επίπεδου, μη κανονικοποιημένου πίνακα

είναι συνειδητή και αντανακλά τη φύση των δεδομένων: το σύνολο δεν είναι ιεραρχία γραμμών και διαδρομών, όπως προϋποθέτει το πρότυπο GTFS, αλλά ένα επίπεδο σύνολο ανεξάρτητων συνδέσεων μεταξύ ζευγών στάσεων. Δεν προβλέπεται, αντιστοίχως, χωριστός πίνακας στάσεων, διότι η ταυτότητα των στάσεων δεν κατοικεί στη βάση αλλά στο αρχείο κανονικών οντοτήτων, το οποίο συντηρείται από το διαδραστικό εργαλείο του χάρτη και θα ήταν άβολο να συγχρονίζεται με σχεσιακή δομή.

δρομολόγια	
πηγή	TEXT
φορέας προέλευσης	TEXT
όνομα αφετηρίας	TEXT
όνομα προορισμού	TEXT
ώρα αναχώρησης	TEXT (ΩΩ:ΛΛ)
διάρκεια	TEXT (ΩΩ:ΛΛ)
ένδειξη αστικού δρομολογίου	INTEGER
ημέρες εβδομαδιαίας εκτέλεσης	TEXT (λίστα)
αρχείο προέλευσης	TEXT
χρονική σήμανση φόρτωσης	TEXT

ευρετήρια:
στάση αναχώρησης,
στάση άφιξης,
όνομα αρχείου

Χωρίς χωριστό πίνακα στάσεων και χωρίς περιορισμό μοναδικότητας:
η ταυτότητα των στάσεων τηρείται στο αρχείο κανονικών οντοτήτων.

Σχήμα 4.1: Το σχήμα της βάσης δεδομένων: ενιαίος, μη κανονικοποιημένος πίνακας δρομολογίων με τα τρία ευρετήρια που υποστηρίζουν τα ερωτήματα του συστήματος.

Τρία ευρετήρια υποστηρίζουν τα πραγματικά ερωτήματα του συστήματος: στη στάση αναχώρησης και στη στάση άφιξης, που εξυπηρετούν τις αναζητήσεις κατά την ανάλυση του δικτύου, και στο όνομα του αρχείου, το οποίο είναι κρίσιμο για τη διαδικασία φόρτωσης της επόμενης υποενότητας. Οι ώρες αποθηκεύονται ως κείμενο σε μορφή ωρών και λεπτών και οι ημέρες ως σειριοποιημένος πίνακας ονομάτων, επιλογές που θυσιάζουν τη δυνατότητα σύγκρισης εντός της SQL προς όφελος της αναγνωσιμότητας και της άμεσης αντιστοιχίας με τα αρχεία JSON· η μετατροπή σε λεπτά και σε σύνολα ημερών εκτελείται ούτως ή άλλως κατά την κατασκευή του γράφου. Δεν ορίζεται, τέλος, περιορισμός μοναδικότητας, καθώς η απαλοιφή διπλοεγγραφών εκτελείται ήδη κατά τη συγκέντρωση του εβδομαδιαίου προτύπου και ο περιορισμός θα απέρριπτε σιωπηλά εγγραφές αντί να αναδείξει το πρόβλημα στην πηγή του.

4.8.2 Διαδικασία φόρτωσης και ιδιότητα idempotency

Η φόρτωση εκτελείται από χωριστό πρόγραμμα που διατρέχει τους τέσσερις φακέλους δεδομένων και επεξεργάζεται κάθε αρχείο ανεξάρτητα, παραλείποντας τα βοηθητικά αρχεία στάσεων. Για κάθε αρχείο εξάγονται από το όνομά του η πηγή και ο φορέας, διαγράφονται πρώτα όλες οι υπάρχουσες εγγραφές που φέρουν το συγκεκριμένο όνομα αρχείου και εισάγονται στη συνέχεια μαζί οι τρέχουσες, με μία μόνο πράξη πολλαπλής εισαγωγής και μία δέσμευση συναλλαγής ανά αρχείο. Εγγραφές που στερούνται κάποιου από τα τρία υποχρεωτικά πεδία απορρίπτονται κατά την προετοιμασία.

Ο συνδυασμός διαγραφής και εισαγωγής ανά αρχείο εξασφαλίζει την ιδιότητα της ταυτοδυναμίας. Το τελικό περιεχόμενο της βάσης εξαρτάται αποκλειστικά από το τρέχον περιεχόμενο των αρχείων και όχι από το πλήθος ή τη σειρά των εκτελέσεων: μια επανάληψη επί αμετάβλητων δεδομένων παράγει πανομοιότυπη βάση, ενώ μια μερική επανεκτέλεση, ύστερα από ανανέωση ενός μόνο φορέα, αντικαθιστά ακριβώς τις εγγραφές του χωρίς να θίγει τους υπόλοιπους. Η ιδιότητα αυτή είναι ουσιώδης σε ένα σύστημα όπου η συλλογή εκτελείται τμηματικά και επανειλημμένα. Ο μοναδικός περιορισμός της είναι ότι η διαγραφή ενός αρχείου από τον δίσκο δεν αφαιρεί τις εγγραφές του από τη βάση, περίπτωση που αντιμετωπίζεται με πλήρη ανακατασκευή. Αξιοποιούνται, τέλος, οι ρυθμίσεις ημερολογίου προτεταγμένης καταγραφής και χαλαρού συγχρονισμού με τον δίσκο που αναφέρθηκαν στην ενότητα 3.3, ενώ χωριστή λειτουργία επιτρέπει τη δημιουργία μόνο του σχήματος, χρήσιμη για την αρχικοποίηση καθαρού αντιγράφου της βάσης πριν από την πρώτη φόρτωση.

4.9 Στατιστικά και ποιοτική αποτίμηση του τελικού συνόλου δεδομένων

Το τελικό σύνολο δεδομένων, όπως αποτυπώθηκε στο στιγμιότυπο του Ιουνίου 2026, περιλαμβάνει 106.380 εγγραφές δρομολογίων προερχόμενες από 35 ΚΤΕΛ. Η κατανομή ανά πηγή είναι έντονα ασύμμετρη: η πλατφόρμα ΙΜΕ συνεισφέρει 93.136 εγγραφές από 21 φορείς, η 3pi-telematics 11.952 από πέντε, το αυτόνομο σύστημα του ΚΤΕΛ Ηλείας 782 και η e-analysis μόλις 510 από οκτώ φορείς, ασυμμετρία που αντανάκλα τόσο το μέγεθος των δικτύων όσο και τη διαφορετική λεπτομέρεια με την οποία κάθε πλατφόρμα εκθέτει τις ενδιαμέσες στάσεις. Οι εγγραφές αντιστοιχούν σε 26.473 διακριτά ζεύγη αφετηρίας και προορισμού και σε 2.310 κόμβους του τελικού γράφου, εκ των οποίων 460 διαθέτουν εξερχόμενα δρομολόγια. Ως προς την εβδομαδιαία εκτέλεση, 37.326 δρομολόγια εκτελούνται και τις επτά ημέρες, 19.693 τις πέντε εργάσιμες και 14.831 μία μόνο ημέρα την εβδομάδα, κατανομή που επιβεβαιώνει την αραιότητα του δικτύου η οποία περιγράφηκε θεωρητικά στο δεύτερο κεφάλαιο. Καμία εγγραφή δεν στερείται διάρκειας διαδρομής και καμία στάση δεν στερείται γεωγραφικών συντεταγμένων. Δύο φορείς της πλατφόρμας ΙΜΕ δεν επέστρεψαν κανένα δρομολόγιο, παρότι εμφανίζονται στον κατάλόγό της, και εξαιρέθηκαν· το γεγονός αυτό, μαζί με την πλήρη απουσία των νησιωτικών ΚΤΕΛ που δεν διαθέτουν καμία από τις τέσσερις υποδομές, οριοθετεί την κάλυψη του συνόλου στην ηπειρωτική χώρα.

Πίνακας 4.1: Σύνθεση του τελικού συνόλου δεδομένων ανά πηγή (στιγμιότυπο Ιουνίου 2026).

Πηγή	Εγγραφές δρομολογίων	Φορείς
ΙΜΕ	93.136	21
3pi-telematics	11.952	5
ΚΤΕΛ Ηλείας (αυτόνομο σύστημα)	782	1
e-analysis	510	8
Σύνολο	106.380	35

Πίνακας 4.2: Κατανομή των δρομολογίων ως προς την εβδομαδιαία εκτέλεση.

Συχνότητα εκτέλεσης	Δρομολόγια
Και τις επτά ημέρες	37.326
Τις πέντε εργάσιμες	19.693
Μία μόνο ημέρα την εβδομάδα	14.831

Επίλογος κεφαλαίου

Το κεφάλαιο διέτρεξε την πλήρη διαδρομή από τέσσερις ασύνδετες και μη τεκμηριωμένες πηγές έως ένα ενιαίο σύνολο δεδομένων: την ταξινόμηση των πηγών κατά τεχνολογική υποδομή, τη μεθοδολογία αντίστροφης μηχανίκευσης και την αποκρυπτογράφηση μίας εξ αυτών, τους μηχανισμούς συλλογής με τους δύο τρόπους ταυτοχρονισμού και την πολιτική ανθεκτικότητας, την ανακατασκευή του εβδομαδιαίου προτύπου μέσω δειγματοληψίας επτά ημερών, την ενοποίηση των οντοτήτων στάσεων με συνδυασμό γεωκωδικοποίησης, χωρικής ομαδοποίησης και ημιαυτόματης επικύρωσης, και τέλος τη φόρτωση σε σχεσιακή βάση με εγγυημένη ταυτοδυναμία. Το αποτέλεσμα είναι η προϋπόθεση που το δεύτερο κεφάλαιο θεωρούσε δεδομένη και το τρίτο ανέδειξε ως ελλείπουσα. Με τη βάση αυτή διαθέσιμη, το επόμενο κεφάλαιο περιγράφει την κατασκευή του γράφου, την υλοποίηση του αλγορίθμου και την ανάπτυξη του ΜΙΤΟΥ, του συστήματος που καθιστά τα δεδομένα αξιοποιήσιμα από τον τελικό χρήστη.

Κεφάλαιο 5

Ανάπτυξη του ΜΙΤΟΥ

Με το ενοποιημένο σύνολο δεδομένων διαθέσιμο και τον αλγόριθμο τυπικά διατυπωμένο, το παρόν κεφάλαιο περιγράφει τη μετατροπή τους σε λειτουργικό σύστημα, τον ΜΙΤΟ. Η ανάπτυξη ακολουθεί τη σειρά με την οποία λαμβάνονται οι σχεδιαστικές αποφάσεις σε ένα έργο λογισμικού: πρώτα καταγράφονται οι απαιτήσεις, λειτουργικές και μη, και οι τυπικές περιπτώσεις χρήσης που το σύστημα οφείλει να καλύψει· έπειτα παρουσιάζεται η αρχιτεκτονική και ο διαχωρισμός των επιπέδων· ακολουθεί η αναλυτική περιγραφή του υπολογιστικού πυρήνα, από την κατασκευή του γράφου έως τη μετα-βελτιστοποίηση των διαδρομών και τη διεπαφή προγραμματισμού· και το κεφάλαιο κλείνει με την ανάπτυξη της διεπαφής χρήστη. Παντού επιδιώκεται να καταστεί σαφής η αντιστοιχία ανάμεσα στις θεωρητικές επιλογές του δεύτερου κεφαλαίου και στον κώδικα που τις υλοποιεί.

5.1 Λειτουργικές απαιτήσεις

Οι λειτουργικές απαιτήσεις προέκυψαν από τους στόχους της ενότητας 1.5 και από την αδυναμία που περιγράφηκε στην ενότητα 1.3: ο χρήστης δεν γνωρίζει ούτε ποιον φορέα να ρωτήσει ούτε από ποιον ενδιάμεσο κόμβο θα περάσει. Το σύστημα οφείλει, επομένως, να απαντά σε ερώτημα διατυπωμένο αποκλειστικά σε όρους αφετηρίας και προορισμού.

ΛΑ1. Αναζήτηση διαδρομής μεταξύ δύο στάσεων. Δοθέντων μιας αφετηρίας, ενός προορισμού, μιας ημέρας της εβδομάδας και μιας πρωιμότερης ώρας αναχώρησης, το σύστημα επιστρέφει τη βέλτιστη διαδρομή ως ακολουθία σκελών, με πλήρη στοιχεία αναχώρησης και άφιξης για καθένα από αυτά.

ΛΑ2. Υποστήριξη ανταποκρίσεων. Η αναζήτηση δεν περιορίζεται στις απευθείας συνδέσεις αλλά συνθέτει διαδρομές με έως τρεις μετεπιβιβάσεις, ενδεχομένως μεταξύ δρομολογίων διαφορετικών φορέων, τηρώντας ελάχιστο χρόνο ανταπόκρισης σε κάθε αλλαγή οχήματος.

ΛΑ3. Διαδρομές πολλαπλών ημερών. Όταν καμία σύνδεση δεν ολοκληρώνεται εντός της ημέρας εκκίνησης, το σύστημα εντοπίζει λύσεις που περιλαμβάνουν διανυκτέρευση σε ενδιάμεσο κόμβο, τη σημαίνει ρητά ως τέτοια και παραθέτει τις εναλλακτικές αναχωρήσεις του αντίστοιχου σκέλους.

ΛΑ4. Εβδομαδιαία σύνοψη διαθεσιμότητας. Όταν ο χρήστης δεν δηλώνει συγκεκριμένη ημέρα, το σύστημα εκτελεί την αναζήτηση για καθεμία από τις επτά ημέρες και επιστρέφει συγκεντρωτική εικόνα του πότε η μετακίνηση είναι εφικτή και ποια είναι η καλύτερη διαδρομή κάθε ημέρας.

ΛΑ5. Αναζήτηση στάσεων με ανοχή. Η επιλογή αφετηρίας και προορισμού γίνεται με πληκτρολόγηση μέρους του ονόματος, ανεξαρτήτως τονισμού και πεζών ή κεφαλαίων, με υποδείξεις που εμφανίζονται κατά την πληκτρολόγηση, καθώς και με απευθείας επιλογή σημείου πάνω σε χάρτη.

ΛΑ6. Παραμετροποίηση των κριτηρίων. Ο χρήστης μπορεί να μεταβάλλει τον ελάχιστο χρόνο ανταπόκρισης, τη μέγιστη ανεκτή αναμονή, το μέγιστο πλήθος σκελών και το μέγιστο πλήθος διανυκτερεύσεων, ώστε η αναζήτηση να προσαρμόζεται στις προτιμήσεις του και όχι το αντίστροφο.

ΛΑ7. Ισοδύναμη πρόσβαση από γραμμή εντολών. Κάθε λειτουργία της διαδικτυακής εφαρμογής είναι διαθέσιμη και μέσω γραμμής εντολών, ώστε το σύστημα να ενσωματώνεται σε σενάρια αυτοματοποίησης ή μαζικών δοκιμών.

5.2 Μη λειτουργικές απαιτήσεις

Πέρα από το τι κάνει, το σύστημα οφείλει να ικανοποιεί ορισμένες ποιοτικές ιδιότητες, οι οποίες επηρέασαν τις αρχιτεκτονικές επιλογές τουλάχιστον όσο και οι λειτουργικές απαιτήσεις.

Η πρώτη είναι η διαδραστικότητα της απόκρισης. Ένας σχεδιαστής ταξιδιού που απαιτεί αισθητή αναμονή παύει να χρησιμοποιείται εξερευνητικά, και ακριβώς η εξερεύνηση —η δοκιμή διαφορετικών ημερών, ωρών και παραμέτρων— αποτελεί τον κύριο τρόπο με τον οποίο ο χρήστης κατανοεί το δίκτυο. Τέθηκε ως στόχος η απόκριση κάθε μεμονωμένου ερωτήματος να παραμένει σαφώς κάτω από το δευτερόλεπτο, όριο πέραν του οποίου η αλληλεπίδραση γίνεται αντιληπτή ως καθυστέρηση· η επίτευξή του αποτιμάται στην ενότητα 6.7.

Η δεύτερη είναι η ορθότητα με την ασύμμετρη έννοια που καθιέρωσε το τέταρτο κεφάλαιο. Μια διαδρομή που δεν προτείνεται ενώ υπάρχει συνιστά χαμένη ευκαιρία· μια διαδρομή που προτείνεται ενώ δεν είναι εφικτή συνιστά σφάλμα με πραγματικό κόστος για τον επιβάτη. Το σύστημα προτιμά συστηματικά τη συντηρητική επιλογή, από τον ελάχιστο χρόνο ανταπόκρισης έως την άρνηση χρήσης ενός σκέλους άγνωστης διάρκειας ως ενδιάμεσου.

Η τρίτη είναι η αναγνωσιμότητα του αποτελέσματος. Δεν αρκεί η επιστροφή μιας βέλτιστης ακολουθίας· ο χρήστης πρέπει να μπορεί να κρίνει αν η πρόταση τον εξυπηρετεί, γεγονός που επιβάλλει την εμφάνιση των χρόνων αναμονής, του φορέα κάθε σκέλους και της διάκρισης ανάμεσα σε χρόνο ταξιδιού και συνολικό χρόνο.

Η τέταρτη είναι η αναπαραγωγικότητα και η ευκολία εγκατάστασης. Το σύστημα εκτελείται με μοναδική προϋπόθεση την ύπαρξη διερμηνέα Python, χωρίς εξυπηρετητή βάσης δεδομένων, χωρίς στάδιο μεταγλώττισης της διεπαφής και χωρίς κλειδιά πρόσβασης σε εμπορικές υπηρεσίες.

Η πέμπτη, τέλος, είναι η συντηρησιμότητα απέναντι σε μεταβαλλόμενα δεδομένα. Καθώς τα χρονοδιαγράμματα αλλάζουν και οι πηγές δεν προειδοποιούν, η ανανέωση του συνόλου

δεδομένων οφείλει να μη συνεπάγεται καμία μεταβολή στον υπολογιστικό πυρήνα: αρκεί η επανεκτέλεση της αλυσίδας φόρτωσης και προϋπολογισμού.

5.3 Case studies

Η μετάβαση από τις απαιτήσεις στη σχεδίαση διευκολύνθηκε από τη διατύπωση τεσσάρων αντιπροσωπευτικών περιπτώσεων χρήσης, οι οποίες λειτούργησαν ως κριτήρια αποδοχής καθ' όλη τη διάρκεια της ανάπτυξης. Οι περιπτώσεις αυτές δεν επιλέχθηκαν ως ακραίες αλλά ως τυπικές, με την έννοια ότι καθεμία ενεργοποιεί διαφορετικό μηχανισμό του συστήματος· η εκτέλεσή τους επί των πραγματικών δεδομένων παρουσιάζεται αναλυτικά στο έκτο κεφάλαιο.

Η πρώτη περίπτωση είναι η απευθείας σύνδεση μεταξύ δύο πόλεων που εξυπηρετούνται από το ίδιο ΚΤΕΛ. Πρόκειται για το ερώτημα που ήδη απαντούν οι υφιστάμενοι ιστότοποι και το οποίο το σύστημα οφείλει να απαντά τουλάχιστον εξίσου καλά, με το πρόσθετο όφελος ότι ο χρήστης δεν χρειάζεται να γνωρίζει ποιον φορέα να ρωτήσει. Λειτουργεί, επομένως, ως έλεγχος βάσης: κάθε απόκλιση από τα δημοσιευμένα δρομολόγια εδώ υποδεικνύει σφάλμα στα δεδομένα ή στη μοντελοποίηση.

Η δεύτερη είναι η διαδρομή με μία ή περισσότερες ανταποκρίσεις μεταξύ περιοχών που δεν συνδέονται απευθείας, όπως μια μετακίνηση από τη Μακεδονία προς την Πελοπόννησο. Εδώ ενεργοποιείται ο πυρήνας της συνεισφοράς: η σύνθεση σκελών διαφορετικών φορέων μέσω ενός κόμβου μετεπιβίβασης τον οποίο ο χρήστης δεν είχε προσδιορίσει. Είναι ακριβώς η περίπτωση που καμία υφιστάμενη πηγή δεν καλύπτει.

Η τρίτη είναι η διαδρομή που δεν ολοκληρώνεται αυθημερόν και απαιτεί διανυκτέρευση σε ενδιάμεσο κόμβο. Η περίπτωση αυτή ελέγχει τη σωστή λειτουργία της αναπαράστασης του χρόνου σε απόλυτα λεπτά και της διαχείρισης του εβδομαδιαίου προτύπου εκτέλεσης, καθώς το δεύτερο σκέλος ενδέχεται να μην εκτελείται καθόλου την ημέρα άφιξης στον ενδιάμεσο κόμβο.

Η τέταρτη είναι το ερώτημα διαθεσιμότητας χωρίς δεσμευμένη ημέρα, το οποίο απαντά στην ερώτηση «ποιες ημέρες μπορώ να πάω» και όχι «πώς πηγαίνω αύριο». Πρόκειται για ερώτημα σχεδιασμού και όχι εκτέλεσης, ιδιαίτερα κρίσιμο σε αραιό δίκτυο όπου πολλές συνδέσεις εκτελούνται δύο ή τρεις ημέρες την εβδομάδα, και το οποίο καμία υφιστάμενη διεπαφή δεν υποστηρίζει.

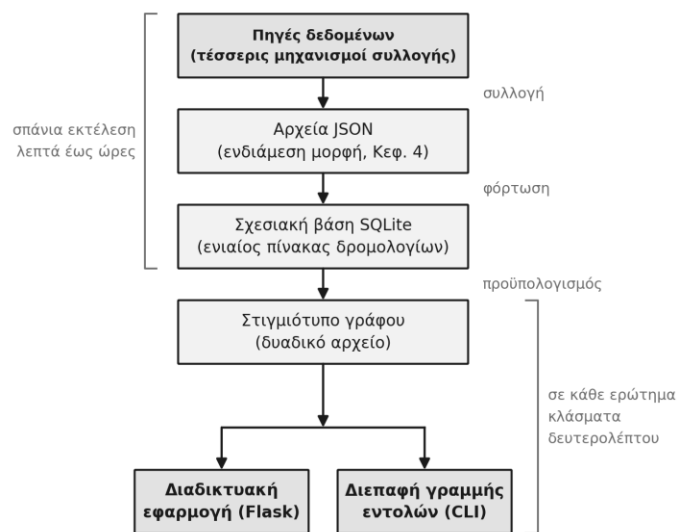
5.4 Αρχιτεκτονική συστήματος

5.4.1 Επισκόπηση και διάγραμμα ροής δεδομένων

Ο MITOY δεν είναι μία εφαρμογή αλλά μια αλυσίδα διακριτών σταδίων, τα οποία επικοινωνούν αποκλειστικά μέσω αρχείων και εκτελούνται με εντελώς διαφορετική συχνότητα. Στο ένα άκρο βρίσκονται οι τέσσερις μηχανισμοί συλλογής, οι οποίοι εκτελούνται σπάνια —όποτε κρίνεται ότι τα χρονοδιαγράμματα έχουν μεταβληθεί— και παράγουν τα αρχεία JSON του τετάρτου κεφαλαίου. Ακολουθεί η φόρτωση στη σχεσιακή βάση, η οποία μετατρέπει τα αρχεία αυτά σε ενιαίο πίνακα δρομολογίων. Στη συνέχεια, ένα χωριστό στάδιο προϋπολογισμού διαβάσει τη βάση,

εφαρμόζει την αντιστοίχιση των κανονικών στάσεων και παράγει τον έτοιμο γράφο ως δυαδικό στιγμιότυπο. Στο άλλο άκρο βρίσκονται οι δύο καταναλωτές του γράφου, η διαδικτυακή εφαρμογή και η διεπαφή γραμμής εντολών, οι οποίοι εκτελούνται συχνά και δεν γνωρίζουν τίποτε ούτε για τις πηγές ούτε για τη βάση.

Η ροή, δηλαδή, είναι αυστηρά μονόδρομη: πηγές, αρχεία JSON, σχεσιακή βάση, στιγμιότυπο γράφου, απάντηση σε ερώτημα. Κανένα στάδιο δεν γράφει πίσω στο προηγούμενο και κάθε ενδιάμεσο προϊόν είναι αυτοτελές αρχείο που μπορεί να επιθεωρηθεί, να αντιγραφεί ή να αντικατασταθεί χειροκίνητα. Η ιδιότητα αυτή αποδείχθηκε καθοριστική κατά την ανάπτυξη, διότι επιτρέπει την απομόνωση κάθε σφάλματος στο στάδιο που το παρήγαγε: μια εσφαλμένη ώρα άφιξης εντοπίζεται είτε στο αρχείο JSON, οπότε ευθύνεται η συλλογή, είτε μόνο στον γράφο, οπότε ευθύνεται ο μετασχηματισμός.



Σχήμα 5.1: Διάγραμμα ροής δεδομένων του MITOY. Η ροή είναι αυστηρά μονόδρομη και κάθε ενδιάμεσο προϊόν αποτελεί αυτοτελές αρχείο.

Η ασυμμετρία των συχνотήτων εκτέλεσης δικαιολογεί από μόνη της τη διάσπαση: τα τρία πρώτα στάδια εκτελούνται σπάνια και κοστίζουν από λεπτά έως ώρες, ενώ τα δύο τελευταία εκτελούνται σε κάθε ερώτημα και οφείλουν να κοστίζουν κλάσματα του δευτερολέπτου.

5.4.2 Διαχωρισμός επιπέδων

Η αλυσίδα αντιστοιχεί σε τέσσερα εννοιολογικά επίπεδα με σαφώς οριοθετημένες ευθύνες. Το επίπεδο συλλογής γνωρίζει τις ιδιαιτερότητες κάθε πηγής —τα σημεία πρόσβασης, την κρυπτογράφηση, τα πεδία κατάστασης— και τίποτε άλλο· η μοναδική του υποχρέωση απέναντι στο υπόλοιπο σύστημα είναι η παραγωγή εγγραφών κοινού σχήματος. Το επίπεδο αποθήκευσης γνωρίζει το σχήμα της βάσης και τη διαδικασία ταυτοδύναμης φόρτωσης, αγνοώντας πλήρως την προέλευση των δεδομένων. Το επίπεδο υπολογισμού γνωρίζει μόνο τον γράφο και τον αλγόριθμο,

χωρίς καμία αναφορά σε αρχεία, σε SQL ή σε HTTP. Το επίπεδο παρουσίασης, τέλος, γνωρίζει μόνο τη μορφή των αποτελεσμάτων.

Ο διαχωρισμός αυτός δεν είναι σχολαστικισμός· έχει τρεις μετρήσιμες συνέπειες. Πρώτον, ο υπολογιστικός πυρήνας είναι δοκιμάσιμος αυτοτελώς, με δεδομένα κατασκευασμένα στο χέρι και χωρίς καμία υποδομή. Δεύτερον, η προσθήκη μιας πέμπτης πηγής απαιτεί μόνο έναν νέο μηχανισμό συλλογής που παράγει το ίδιο σχήμα εγγραφών, χωρίς καμία μεταβολή στα υπόλοιπα τρία επίπεδα. Τρίτον, ο πυρήνας εξυπηρετεί ταυτόχρονα δύο εντελώς διαφορετικές διεπαφές — μια διαδικτυακή εφαρμογή και ένα πρόγραμμα γραμμής εντολών— χωρίς καμία γραμμή κώδικα ειδικά για καθεμία, γεγονός που αποτελεί και τον πρακτικότερο έλεγχο της ορθότητας του διαχωρισμού.

Υπάρχει, ωστόσο, ένα σημείο όπου ο διαχωρισμός δεν είναι απόλυτος και οφείλει να δηλωθεί. Ο κανόνας με τον οποίο ένα ακατέργαστο όνομα στάσης μεταφράζεται στην κανονική του μορφή εμφανίζεται τόσο στο στάδιο του προϋπολογισμού όσο και στη διαδικτυακή εφαρμογή, η οποία τον χρειάζεται για να αντιστοιχίσει τις συντεταγμένες των στάσεων στα ονόματα του γράφου. Η επανάληψη συνιστά σύζευξη: κάθε μεταβολή του κανόνα πρέπει να εφαρμοστεί και στα δύο σημεία, αλλιώς οι δείκτες του χάρτη παύουν να ταυτίζονται με τους κόμβους του.

5.4.3 Προϋπολογισμός γράφου και μηχανισμός cache

Η κατασκευή του γράφου από τη βάση δεν είναι δαπανηρή σε απόλυτους όρους, απαιτεί όμως την ανάγνωση και τον μετασχηματισμό όλων των εγγραφών: μετατροπή των ωρών σε λεπτά, αποσειριοποίηση των ημερών εκτέλεσης, αντιστοίχιση κάθε άκρου στην κανονική του οντότητα και ταξινόμηση των αναχωρήσεων κάθε στάσης. Η επανάληψή της σε κάθε εκκίνηση του εξυπηρετητή ή σε κάθε κλήση από τη γραμμή εντολών θα ήταν καθαρή σπατάλη, αφού το αποτέλεσμα μεταβάλλεται μόνο όταν αλλάξουν τα δεδομένα.

Για τον λόγο αυτό ο προϋπολογισμός απομονώθηκε σε χωριστό πρόγραμμα, το οποίο εκτελείται ρητά μετά από κάθε ανανέωση της βάσης και αποθηκεύει το αποτέλεσμα σε ένα αρχείο. Το αποθηκευμένο στιγμιότυπο περιλαμβάνει τέσσερις δομές —τις λίστες γειτνίασης του γράφου, το σύνολο όλων των στάσεων, τον κατάλογο των κανονικών οντοτήτων και το λεξικό αντιστοίχισης— καθώς και τη χρονική σήμανση της κατασκευής του, ώστε να είναι πάντοτε γνωστό σε ποια έκδοση των δεδομένων αντιστοιχεί η τρέχουσα λειτουργία. Χωριστή λειτουργία ελέγχου εκτυπώνει τα συγκεντρωτικά μεγέθη του στιγμιότυπου χωρίς να το ανακατασκευάσει, διευκολύνοντας την επαλήθευση ότι το σύστημα εκτελείται πράγματι επί των αναμενόμενων δεδομένων.

Η ανάκτηση του έτοιμου γράφου κοστίζει τάξεις μεγέθους λιγότερο από την ανακατασκευή του και εκτελείται μία μόνο φορά, κατά την εκκίνηση της εφαρμογής· εν συνεχεία ο γράφος παραμένει στη μνήμη και κάθε αίτημα εξυπηρετείται χωρίς καμία πρόσβαση στον δίσκο. Το τίμημα είναι η ρητή εξάρτηση από τη σειρά εκτέλεσης: αν το στιγμιότυπο λείπει ή είναι παλαιότερο της βάσης, το σύστημα δεν το ανακατασκευάζει σιωπηρά αλλά τερματίζει με μήνυμα που υποδεικνύει τα βήματα που πρέπει να προηγηθούν.

5.5 Ανάπτυξη του backend

5.5.1 Κατασκευή του γράφου δρομολογίων από τη βάση

Ο γράφος υλοποιείται ως λεξικό που αντιστοιχίζει κάθε στάση αναχώρησης στον κατάλογο των δρομολογίων που ξεκινούν από αυτήν, δηλαδή ως λίστες γειτνίασης, σύμφωνα με την επιλογή που τεκμηριώθηκε στην ενότητα 2.1.3. Κάθε δρομολόγιο αποθηκεύεται ως εγγραφή που φέρει τη στάση προέλευσης και προορισμού στην κανονική τους μορφή, την ώρα αναχώρησης τόσο σε λεπτά όσο και στην αρχική μορφή εμφάνισης, τη διάρκεια, τον προϋπολογισμένο χρόνο άφιξης, το σύνολο των ημερών εκτέλεσης και το αρχείο προέλευσης, το οποίο επιτρέπει την ανάκτηση του φορέα κατά την παρουσίαση.

Το κρίσιμο βήμα της κατασκευής είναι η αντιστοίχιση των ονομάτων. Για κάθε άκρο κάθε δρομολογίου σχηματίζεται η τριάδα πηγής, φορέα και ακατέργαστου ονόματος και αναζητείται στο λεξικό κανονικοποίησης του αρχείου κανονικών στάσεων. Όταν η αναζήτηση αποτύχει, δεν εφαρμόζεται καμία λεξιλογική προσέγγιση: η στάση λαμβάνει σύνθετο αναγνωριστικό που συνενώνει το αρχικό της όνομα με τον φορέα, παραμένοντας πλήρως λειτουργική εντός του δικτύου του φορέα της χωρίς κίνδυνο εσφαλμένης συγχώνευσης με ομώνυμη άλλου. Η συμπεριφορά αυτή, που αναλύθηκε στην ενότητα 4.6.5, είναι το σημείο όπου η αρχή της ασφαλούς αστοχίας μεταφράζεται σε λίγες γραμμές κώδικα.

Ολοκληρώνοντας, οι λίστες κάθε στάσης ταξινομούνται κατά ώρα αναχώρησης. Η ταξινόμηση αυτή δεν είναι διακοσμητική αλλά προϋπόθεση της ορθότητας του επόμενου βήματος: ο αλγόριθμος στηρίζεται στο ότι η πρώτη έγκυρη αναχώρηση που θα συναντήσει κατά τη σάρωση είναι πράγματι η πρωιμότερη, δηλαδή ακριβώς στην υλοποίηση της συνάρτησης ελαχίστου της ενότητας 2.3.1 από την οποία απορρέει η ιδιότητα FIFO.

5.5.2 Αναπαράσταση του χρόνου σε απόλυτα λεπτά και υποστήριξη πολλαπλών ημερών

Ο χρόνος εκφράζεται καθολικά σε λεπτά, με δύο διακριτές έννοιες που δεν συγχέονται ποτέ. Η ώρα της ημέρας είναι ακέραιος στο διάστημα από μηδέν έως χίλια τετρακόσια σαράντα και αποθηκεύεται στις εγγραφές των δρομολογίων. Ο απόλυτος χρόνος είναι ο δείκτης της ημέρας επί χίλια τετρακόσια σαράντα, προσαυξημένος κατά την ώρα της ημέρας, και χρησιμοποιείται αποκλειστικά από την αναζήτηση.

Η μετατροπή αυτή, θεωρητικά ασήμαντη, αποδεικνύεται πρακτικά καθοριστική. Επαναφέρει τη γραμμική διάταξη του χρόνου, ώστε η ουρά προτεραιότητας να συγκρίνει απλούς ακεραίους και όχι ζεύγη ημέρας και ώρας, και εξαλείφει ολόκληρη κατηγορία σφαλμάτων γύρω από τα μεσάνυχτα. Ένα δρομολόγιο που αναχωρεί στις είκοσι τρεις και μηδέν και διαρκεί τέσσερις ώρες παράγει χρόνο άφιξης μεγαλύτερο του ορίου της ημέρας· η ημέρα άφιξης προκύπτει τότε αυτομάτως με ακέραια διαίρεση, χωρίς καμία ειδική περίπτωση στον κώδικα.

Η ίδια αναπαράσταση καθιστά τις διανυκτερεύσεις φυσικό επακόλουθο και όχι εξαίρεση. Η διαφορά των δεικτών ημέρας ανάμεσα στην άφιξη ενός σκέλους και στην αναχώρηση του επόμενου δίνει απευθείας το πλήθος των διανυκτερεύσεων, ενώ η μετατροπή ενός δείκτη ημέρας

σε ημέρα της εβδομάδας —απαραίτητη για τον έλεγχο του εβδομαδιαίου προτύπου εκτέλεσης— είναι μια απλή πρόσθεση με υπόλοιπο διαίρεσης δια επτά ως προς την ημέρα εκκίνησης.

Αξίζει να σημειωθεί ότι η ημέρα εκκίνησης δηλώνεται ως ημέρα της εβδομάδας και όχι ως ημερολογιακή ημερομηνία, επιλογή που απορρέει ευθέως από τη φύση των δεδομένων: το σύνολο του τετάρτου κεφαλαίου αποτυπώνει εβδομαδιαίο πρότυπο και όχι ημερολόγιο με εξαιρέσεις. Η αναπαράσταση σε απόλυτα λεπτά παραμένει, ωστόσο, συμβατή με ημερολογιακό ορίζοντα, ώστε η μελλοντική προσθήκη εξαιρέσεων για αργίες να μην απαιτεί μεταβολή του πυρήνα.

5.5.3 Υλοποίηση του αλγορίθμου αναζήτησης διαδρομών

Η αναζήτηση υλοποιεί τον χρονοεξαρτώμενο Dijkstra της ενότητας 2.9.3. Η ουρά προτεραιότητας είναι δυαδικός σωρός της πρότυπης βιβλιοθήκης και κάθε καταχώρισή της περιλαμβάνει τον απόλυτο χρόνο άφιξης, έναν αύξοντα μετρητή, το πλήθος των σκελών, την τρέχουσα στάση και το μονοπάτι που οδήγησε σε αυτήν. Ο μετρητής εξυπηρετεί καθαρά τεχνικό σκοπό αλλά είναι απαραίτητος: χωρίς αυτόν, δύο καταχωρίσεις με ίδιο χρόνο θα οδηγούσαν σε σύγκριση των υπόλοιπων πεδίων και σε σφάλμα εκτέλεσης, αφού τα μονοπάτια δεν είναι διατάξιμα.

Η κατάσταση της αναζήτησης δεν είναι η στάση αλλά το ζεύγος στάσης και πλήθους σκελών. Η επιλογή αυτή διατηρεί διαδρομές που φτάνουν αργότερα αλλά με λιγότερες μετεπιβιβάσεις, οι οποίες σε ένα αυστηρά μονοκριτηριακό σχήμα θα απορρίπτονταν, και υλοποιεί έτσι στην πράξη τη διάσταση του πλήθους σκελών που περιγράφηκε θεωρητικά στην ενότητα 2.6. Ο έλεγχος κυριαρχίας εφαρμόζεται εντός κάθε κλάσης χωριστά: μια καταχώριση απορρίπτεται ως παρωχημένη μόνο αν έχει ήδη οριστικοποιηθεί άφιξη όχι μεταγενέστερη για την ίδια στάση με το ίδιο πλήθος σκελών. Πρόκειται για την τεχνική της τεμπέλικης διαγραφής της ενότητας 2.2.5, η οποία αποφεύγει την πράξη μείωσης κλειδιού με τίμημα ορισμένες περιττές καταχωρίσεις στον σωρό.

Σε κάθε εξαγωγή, ο αλγόριθμος σαρώνει τα δρομολόγια που αναχωρούν από την τρέχουσα στάση και για καθένα αναζητά την πρωιμότερη έγκυρη ημέρα επιβίβασης. Το πρώτο σκέλος περιορίζεται στην ημέρα εκκίνησης, ενώ κάθε επόμενο εξετάζει τις ημέρες από την τρέχουσα έως τον καθορισμένο ορίζοντα, ελέγχοντας ταυτόχρονα ότι ο υποψήφιος απόλυτος χρόνος αναχώρησης δεν προηγείται του πρωιμότερου επιτρεπτού και ότι η αντίστοιχη ημέρα της εβδομάδας ανήκει στο σύνολο εκτέλεσης του δρομολογίου.

Η αναζήτηση δεν τερματίζεται στην πρώτη άφιξη στον προορισμό: οι διαδρομές που καταλήγουν σε αυτόν συσσωρεύονται σε χωριστό κατάλογο και η εξερεύνηση συνεχίζεται έως ότου συγκεντρωθεί ένα πολλαπλάσιο των αποτελεσμάτων που θα παρουσιαστούν, ώστε το στάδιο της κατάταξης να έχει πραγματικές εναλλακτικές να συγκρίνει.

5.5.4 Διαχείριση ανταποκρίσεων και διανυκτερεύσεων

Ο περιορισμός ανταπόκρισης εφαρμόζεται με ενιαία τιμή δεκαπέντε λεπτών, σύμφωνα με την παραδοχή της ενότητας 2.9.6, και μόνο όταν πράγματι μεσολαβεί αλλαγή οχήματος: για το πρώτο σκέλος ο επιβάτης θεωρείται ήδη παρών στην αφετηρία και δεν επιβαρύνεται με κανένα περιθώριο. Η διάκριση γίνεται με τον απλούστερο δυνατό έλεγχο, το αν το μονοπάτι είναι ακόμη κενό, και

έχει αισθητή επίπτωση στις πρωινές αναχωρήσεις, όπου δεκαπέντε λεπτά συχνά καθορίζουν αν προλαβαίνεται το πρώτο δρομολόγιο της ημέρας.

Δεύτερος περιορισμός είναι η μέγιστη ανεκτή αναμονή, με προεπιλογή τις τέσσερις ώρες. Ο περιορισμός αυτός δεν πηγάζει από τη θεωρία αλλά από την εμπειρία του χρήστη: μια ανταπόκριση με εξάωρη αναμονή σε έναν υπεραστικό σταθμό είναι τεχνικά έγκυρη και πρακτικά απαράδεκτη. Καθοριστικό είναι ότι ο περιορισμός εφαρμόζεται αποκλειστικά σε ανταποκρίσεις της ίδιας ημέρας. Όταν η επιβίβαση μετατίθεται σε επόμενη ημέρα, η αναμονή παύει να είναι αναμονή σε σταθμό και γίνεται διανυκτέρευση, δηλαδή διάστημα κατά το οποίο ο επιβάτης δεν περιμένει αλλά διαμένει· η εφαρμογή του ορίου εκεί θα απέκλειε αδικαιολόγητα ολόκληρη την κατηγορία των πολυήμερων διαδρομών.

Οι διανυκτερεύσεις, όπως προβλέπει η ενότητα 2.9.3, δεν αντιμετωπίζονται ως ειδική περίπτωση αλλά προκύπτουν αυτομάτως όποτε η πρωιμότερη έγκυρη επιβίβαση εντοπίζεται σε επόμενη ημέρα, είτε επειδή έχουν εξαντληθεί οι αναχωρήσεις της τρέχουσας είτε επειδή το δρομολόγιο δεν εκτελείται τη συγκεκριμένη ημέρα. Το πλήθος τους καταγράφεται ανά σκέλος και ο μέγιστος επιτρεπτός αριθμός τους ταυτίζεται με τον ορίζοντα των πρόσθετων ημερών, με προεπιλογή τη μία.

Ξεχωριστή μεταχείριση επιφυλάσσεται, τέλος, στα σκέλη άγνωστης διάρκειας, τα οποία, παρά τη συμπλήρωση της ενότητας 4.7, θα μπορούσαν να εμφανιστούν σε μελλοντικά δεδομένα: γίνονται δεκτά ως τελευταία, με την άφιξη να δηλώνεται άγνωστη, αλλά ουδέποτε ως ενδιάμεσα.

5.5.5 Μετα-βελτιστοποίηση διαδρομής: ελαχιστοποίηση χρόνων αναμονής

Το στάδιο αυτό υλοποιεί την οπισθοδρομική βελτιστοποίηση της ενότητας 2.9.4, όπου προσδιορίστηκε και η θέση της ως προς τη βιβλιογραφία. Το πρόβλημα που επιλύει είναι εγγενές στην πρωιμότερη άφιξη και ιδιαίτερα οξύ σε αραιά δίκτυα: ο αλγόριθμος επιλέγει πάντοτε το πρωιμότερο δρομολόγιο κάθε σκέλους, ακόμη και όταν αυτό οδηγεί σε πολύωρη αναμονή στον επόμενο κόμβο χωρίς κανένα απολύτως όφελος στην τελική άφιξη. Ο επιβάτης που αναχωρεί στις έξι το πρωί για να προλάβει μια ανταπόκριση των δέκα, ενώ υπάρχει δρομολόγιο των οκτώ και μισή που τον φέρνει εγκαίρως στον ίδιο κόμβο, δεν κερδίζει τίποτε από την πρωινή έγερση.

Η υλοποίηση διατρέχει τα σκέλη της διαδρομής από το προτελευταίο προς το πρώτο. Για κάθε σκέλος υπολογίζεται η προθεσμία άφιξης ως ο απόλυτος χρόνος αναχώρησης του επόμενου μείον τον ελάχιστο χρόνο ανταπόκρισης, και αναζητείται μεταξύ των δρομολογίων που συνδέουν τις ίδιες δύο στάσεις την ίδια ημέρα εκείνο με τη μέγιστη ώρα αναχώρησης του οποίου η άφιξη δεν υπερβαίνει την προθεσμία. Για το πρώτο σκέλος τίθεται επιπλέον κάτω φράγμα η ώρα εκκίνησης που δήλωσε ο χρήστης. Ολοκληρώνοντας, οι χρόνοι αναμονής και τα πλήθη διανυκτερεύσεων επανυπολογίζονται από τους νέους απόλυτους χρόνους, ώστε τα εμφανιζόμενα μεγέθη να αντιστοιχούν στη μετασχηματισμένη και όχι στην αρχική διαδρομή.

Δύο ιδιότητες αξίζουν έμφαση. Πρώτον, ο χρόνος τελικής άφιξης παραμένει απαράλλακτος: το τελευταίο σκέλος δεν θίγεται ποτέ και κάθε αντικατάσταση διατηρεί εξ ορισμού την εφικτότητα της επόμενης ανταπόκρισης, οπότε η λύση εξακολουθεί να είναι βέλτιστη ως προς το αρχικό κριτήριο ενώ η συνολική της διάρκεια συμπίπτει. Δεύτερον, το σφίξιμο αναστέλλεται πάνω από διανυκτέρευση, κατά συνειδητή επιλογή και όχι λόγω τεχνικού περιορισμού: εκεί η πρωιμότερη

άφιξη στον ενδιαμέσο κόμβο είναι προτιμότερη, καθώς αφήνει περιθώριο για τη διευθέτηση της διαμοή. Αντ' αυτού, το σύστημα υπολογίζει και παρουσιάζει ρητά όλες τις μεταγενέστερες αναχωρήσεις του ίδιου τμήματος —περιορισμένες, όταν το σκέλος είναι ενδιαμέσο, σε όσες προλαβαίνουν ακόμη την επόμενη ανταπόκριση— μεταθέτοντας τη σχετική απόφαση στον χρήστη.

5.5.6 Κριτήρια κατάταξης και απαλοιφή διπλοτύπων

Οι υποψήφιες διαδρομές που συγκεντρώθηκαν κατά την αναζήτηση δεν είναι μεταξύ τους συγκρίσιμες με ένα μόνο μέγεθος, οπότε η τελική επιλογή στηρίζεται στη λεξικογραφική διάταξη της ενότητας 2.6.3. Η ιεραρχία των κριτηρίων είναι, κατά σειρά, το πλήθος των διανυκτερεύσεων, το πλήθος των σκελών και ο απόλυτος χρόνος άφιξης. Η σειρά αυτή αποτυπώνει την πραγματική ιεραρχία των προτιμήσεων του επιβάτη: μια διαδρομή που ολοκληρώνεται αυθημερόν υπερτερεί κάθε άλλης ανεξαρτήτως ώρας άφιξης, ενώ μεταξύ διαδρομών ίδιας κατηγορίας προτιμώνται οι λιγότερες αλλαγές και, τελευταία, η πρωιμότερη άφιξη. Οι διαδρομές των οποίων το τελευταίο σκέλος έχει άγνωστη διάρκεια κατατάσσονται συστηματικά τελευταίες, μέσω μιας συμβατικά μεγάλης τιμής άφιξης.

Πρέπει να διευκρινιστεί η ακριβής εμβέλεια της διάταξης αυτής. Η κατάταξη δεν αποτελεί απόδειξη βελτιστότητας ως προς τα τρία κριτήρια συνολικά, αλλά διάταξη επί του συνόλου των υποψηφίων που η αναζήτηση πρόλαβε να συγκεντρώσει· η εγγύηση της ενότητας 2.9.5 αφορά την πρωιμότερη άφιξη εντός κάθε κλάσης πλήθους σκελών και όχι το σύνολο των εναλλακτικών που παρουσιάζονται. Οι υπόλοιπες προτάσεις είναι έγκυρες και χρήσιμες, χωρίς να διεκδικούν βελτιστότητα.

Ακολουθεί η απαλοιφή διπλοτύπων. Το φαινόμενο δεν είναι σπάνιο: η ίδια ακριβώς φυσική διαδρομή μπορεί να εντοπιστεί από περισσότερα του ενός μονοπάτια της αναζήτησης, ιδίως μετά το στάδιο του σφιδήματος, το οποίο ενδέχεται να συγκλίνει διαφορετικές αρχικές επιλογές στην ίδια τελική ακολουθία δρομολογίων. Ως ταυτότητα κάθε διαδρομής ορίζεται η ακολουθία των τριάδων αφετηρίας, προορισμού και απόλυτου χρόνου αναχώρησης κάθε σκέλους, και διατηρείται μόνο η πρώτη εμφάνιση κάθε ταυτότητας. Επειδή η απαλοιφή εκτελείται μετά την ταξινόμηση, η διατηρούμενη εκδοχή είναι πάντοτε η καλύτερα κατατεταγμένη. Από τον καθαρό αυτόν κατάλογο επιστρέφονται έως πέντε διαδρομές.

5.5.7 Ασαφής αναζήτηση και κανονικοποίηση ελληνικού κειμένου

Η αντιστοίχιση της εισόδου του χρήστη σε στάση του γράφου είναι λιγότερο τετριμμένη από όσο φαίνεται, διότι η ελληνική γραφή εισάγει δύο πηγές απόκλισης που δεν έχουν αντίστοιχο στα λατινικά αλφάβητα: τον τονισμό και τη διάκριση του τελικού σίγμα. Ο χρήστης πληκτρολογεί «Ιωαννίνα» ή «ιωάννινα», ενώ τα δεδομένα περιέχουν «ΙΩΑΝΝΙΝΑ»· η άμεση σύγκριση συμβολοσειρών θα απέτυχε σε κάθε μία από αυτές τις περιπτώσεις.

Η κανονικοποίηση εκτελείται σε τρία βήματα. Πρώτα το κείμενο αποσυντίθεται κατά Unicode στη μορφή NFD, η οποία διαχωρίζει κάθε τονισμένο γράμμα σε βασικό χαρακτήρα και διακριτικό σημείο. Έπειτα απαλείφονται όλοι οι χαρακτήρες που ανήκουν στην κατηγορία των μη διαστηματικών σημείων, δηλαδή ακριβώς οι τόνοι και τα διαλυτικά. Τέλος το αποτέλεσμα

μετατρέπεται σε κεφαλαία και περικόπτονται τα περιττά κενά· η μετατροπή σε κεφαλαία εξαλείφει και τη διάκριση του τελικού σίγμα, το οποίο αντιστοιχεί στο ίδιο κεφαλαίο γράμμα.

Επί των κανονικοποιημένων μορφών εφαρμόζεται αναζήτηση δύο επιπέδων. Καταγράφονται χωριστά οι ακριβείς ταυτίσεις και οι μερικές, δηλαδή οι περιπτώσεις όπου η μία συμβολοσειρά περιέχεται στην άλλη προς οποιαδήποτε κατεύθυνση, ώστε να καλύπτεται τόσο η πληκτρολόγηση τμήματος του ονόματος όσο και η αναζήτηση με πληρέστερο όνομα από το καταχωρισμένο. Τα αποτελέσματα κατατάσσονται με προτεραιότητα στις στάσεις που διαθέτουν εξερχόμενα δρομολόγια, καθώς μόνο αυτές μπορούν να χρησιμεύσουν ως αφετηρία. Ιδιαίτερη πρόβλεψη υπάρχει για την περίπτωση όπου υφίσταται ακριβής ταύτιση αλλά καμία από τις ταυτισμένες στάσεις δεν έχει αναχωρήσεις: τότε επιστρέφονται και οι μερικές ταυτίσεις, ώστε ο χρήστης να μη λάβει άδείο αποτέλεσμα ενώ υπάρχει προφανής παραλλαγή του ονόματος που εξυπηρετείται.

5.5.8 Σχεδίαση του REST API

Η διαδικτυακή υπηρεσία υλοποιήθηκε σε Flask και εκθέτει τρία σημεία πρόσβασης, σύμφωνα με τις αρχές που περιγράφηκαν στην ενότητα 3.5. Το πρώτο επιστρέφει, ως απάντηση σε αίτημα ανάκτησης, τον πλήρη κατάλογο των στάσεων με συντεταγμένες, συνοδευόμενη καθεμία από ένδειξη του αν διαθέτει εξερχόμενα δρομολόγια· ο κατάλογος τροφοδοτεί τον χάρτη και υπολογίζεται μία φορά κατά την εκκίνηση. Το δεύτερο δέχεται συμβολοσειρά αναζήτησης ως παράμετρο διεύθυνσης και επιστρέφει έως δεκαπέντε ονόματα στάσεων για την αυτόματη συμπλήρωση. Το τρίτο δέχεται, με αίτημα υποβολής και σώμα σε μορφή JSON, τις παραμέτρους της αναζήτησης και επιστρέφει τις υπολογισμένες διαδρομές.

Κρίσιμη σχεδιαστική λεπτομέρεια του τρίτου σημείου είναι η διπλή του συμπεριφορά. Όταν δηλώνεται ημέρα, εκτελείται μία αναζήτηση και επιστρέφεται η βέλτιστη διαδρομή με πλήρη ανάλυση των σκελών της. Όταν το πεδίο της ημέρας παραλείπεται, εκτελούνται επτά αναζητήσεις, μία ανά ημέρα της εβδομάδας, και επιστρέφεται δομή που περιλαμβάνει για καθεμία το πλήθος των διαθέσιμων διαδρομών, μια μονόγραμμη σύνοψη της καλύτερης και την πλήρη της ανάλυση, υλοποιώντας έτσι την απαίτηση της εβδομαδιαίας σύνοψης.

Η μετατροπή του εσωτερικού μονοπατιού σε απόκριση JSON συγκεντρώνεται σε μία συνάρτηση, η οποία υπολογίζει τα συγκεντρωτικά μεγέθη —συνολική διάρκεια, καθарός χρόνος ταξιδιού, άθροισμα αναμονών, πλήθος διανυκτερεύσεων— και μορφοποιεί κάθε σκέλος με τις ημέρες αναχώρησης και άφιξης, την ετικέτα του φορέα και, όπου υπάρχει διανυκτέρευση, τον κατάλογο των εναλλακτικών αναχωρήσεων. Ο εξυπηρετητής παράγει, δηλαδή, αποκλειστικά δεδομένα και ουδεμία σήμανση παρουσίας, ώστε η ίδια υπηρεσία να μπορεί να καταναλωθεί από οποιονδήποτε άλλο πελάτη χωρίς τροποποίηση.

5.5.9 Διεπαφή γραμμής εντολών (CLI)

Παράλληλα με τη διαδικτυακή εφαρμογή, ο υπολογιστικός πυρήνας εκτίθεται και μέσω διεπαφής γραμμής εντολών, η οποία δέχεται την αφετηρία και τον προορισμό ως θεσιακά ορίσματα και το σύνολο των παραμέτρων της αναζήτησης ως προαιρετικές σημαίες. Η συμπεριφορά της ως προς την ημέρα είναι ακριβώς εκείνη της διαδικτυακής υπηρεσίας: με δηλωμένη ημέρα εκτυπώνονται οι αναλυτικές διαδρομές, ενώ χωρίς αυτήν παράγεται η εβδομαδιαία σύνοψη με μία γραμμή ανά

ημέρα. Χωριστή λειτουργία επιτρέπει την αναζήτηση στάσεων με βάση τμήμα του ονόματος, σημειώνοντας ποιες από τις ταυτισμένες διαθέτουν εξερχόμενα δρομολόγια, και είναι ιδιαίτερα χρήσιμη για τη διερεύνηση του συνόλου δεδομένων.

Μία σχεδιαστική διαφορά ως προς τη διαδικτυακή εφαρμογή αξίζει να επισημανθεί: η γραμμή εντολών εκτυπώνει το σύνολο των διαδρομών που επέστρεψε η αναζήτηση, έως πέντε, ενώ η διαδικτυακή εφαρμογή προβάλλει μόνο την πρώτη της κατάταξης. Η επιλογή αντανακλά το διαφορετικό κοινό των δύο διεπαφών, καθώς ο χρήστης του χάρτη αναζητά μία σαφή πρόταση, ενώ για τον προγραμματιστή ή ένα σενάριο αυτοματοποίησης η πλήρης εικόνα του χώρου των λύσεων έχει αξία.

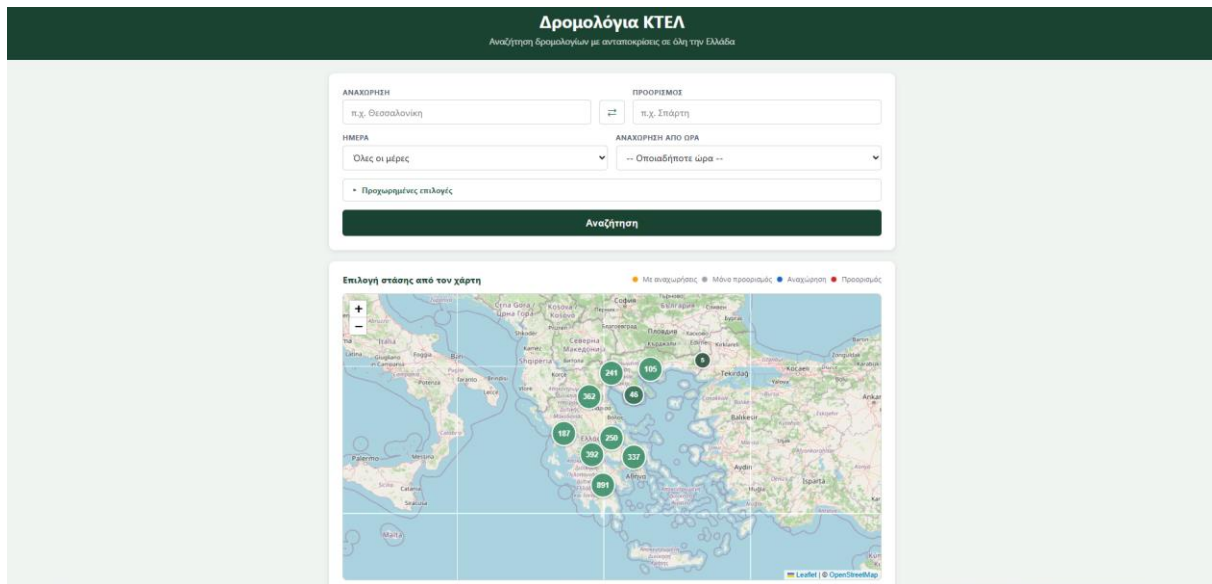
Η διεπαφή αυτή δεν αποτελεί απλή εναλλακτική διευκόλυνση. Υπήρξε το κύριο εργαλείο ανάπτυξης και ελέγχου του αλγορίθμου, καθώς επιτρέπει την ταχεία επανάληψη ερωτημάτων χωρίς επανεκκίνηση εξυπηρετητή, ενώ η μορφοποιημένη έξοδος της αποτυπώνει κάθε σκέλος με τον φορέα του, τις αναμονές και τις διανυκτερεύσεις σε αναγνώσιμη μορφή. Αξιοσημείωτη είναι, τέλος, η διαγνωστική της συμπεριφορά όταν δεν εντοπίζεται διαδρομή: αντί για ένα απλό αρνητικό μήνυμα, ελέγχεται και αναφέρεται αν η αφετηρία στερείται εξερχόμενων δρομολογίων, αν ο προορισμός δεν εμφανίζεται πουθενά ως άκρο άφιξης και πόσα δρομολόγια αναχωρούν από την αφετηρία τη συγκεκριμένη ημέρα, υποδεικνύοντας παράλληλα ποιες παραμέτρους θα είχε νόημα να χαλαρώσει ο χρήστης.

5.6 Ανάπτυξη του frontend

5.6.1 Σχεδίαση της διεπαφής χρήστη

Η διεπαφή οργανώνεται σε μία μόνο σελίδα με τρεις κάθετα διατεταγμένες περιοχές: τη φόρμα αναζήτησης, την περιοχή αποτελεσμάτων που παραμένει κρυφή έως την πρώτη αναζήτηση, και τον διαδραστικό χάρτη. Η επιλογή της μονοσέλιδης δομής απορρέει από τη φύση της εργασίας του χρήστη, η οποία είναι επαναληπτική: σπανίως υποβάλλεται ένα μόνο ερώτημα, και η διατήρηση της φόρμας ορατής μαζί με τα αποτελέσματα επιτρέπει την άμεση τροποποίηση μιας παραμέτρου και την εκ νέου υποβολή χωρίς πλοήγηση.

Η φόρμα ακολουθεί τη λογική της προοδευτικής αποκάλυψης. Στο ορατό της τμήμα βρίσκονται μόνο τα τέσσερα στοιχεία που απαιτούνται πάντοτε —αφετηρία, προορισμός, ημέρα και πρωιμότερη ώρα αναχώρησης— ενώ οι τέσσερις παράμετροι του αλγορίθμου αποσύρονται σε πτυσσόμενη ενότητα προχωρημένων επιλογών. Μεταξύ των δύο πεδίων στάσης παρεμβάλλεται πλήκτρο εναλλαγής, το οποίο αντιστρέφει ταυτόχρονα τα περιεχόμενα των πεδίων και τη χρωματική σήμανση των αντίστοιχων δεικτών στον χάρτη, καλύπτοντας τη συχνή ανάγκη αναζήτησης της επιστροφής. Το Σχήμα 5.2 αποδίδει τη μορφή της αρχικής οθόνης.



Σχήμα 5.2: Η αρχική οθόνη της διαδικτυακής εφαρμογής. Διακρίνονται η φόρμα αναζήτησης με τα τέσσερα πάντοτε ορατά πεδία, το πλήκτρο εναλλαγής αφετηρίας και προορισμού, η πτυσσόμενη ενότητα των προχωρημένων επιλογών και, κάτω από αυτήν, ο διαδραστικός χάρτης με τα συσσωματώματα των στάσεων. Η περιοχή αποτελεσμάτων παραμένει κρυφή έως την πρώτη αναζήτηση.

Η χρωματική παλέτα και οι υπόλοιπες σταθερές της εμφάνισης συγκεντρώνονται σε μεταβλητές CSS ορισμένες σε ένα σημείο, ώστε κάθε μεταβολή του ύφους να είναι εντοπισμένη, ενώ ολόκληρη η διεπαφή αποδίδεται στα ελληνικά με ρητή δήλωση της γλώσσας του εγγράφου. Η κατάσταση της εφαρμογής περιορίζεται σε ελάχιστα μεγέθη —τις δύο επιλεγμένες στάσεις, τις παραμέτρους της φόρμας και το τελευταίο αποτέλεσμα— και διατηρείται σε απλές μεταβλητές, χωρίς μηχανισμό διαχείρισης κατάστασης, επιβεβαιώνοντας στην πράξη την επιλογή της ενότητας 3.6 να αποφευχθεί κάθε πλαίσιο διεπαφής.

5.6.2 Αναζήτηση στάσεων με αυτόματη συμπλήρωση

Η επιλογή στάσης από κατάλογο τριακοσίων πενήντα τριών κανονικών οντοτήτων και δύο χιλιάδων τριακοσίων δέκα συνολικών σημείων θα ήταν αδύνατη με πτυσσόμενο κατάλογο· επιβάλλεται αναζήτηση με υποδείξεις κατά την πληκτρολόγηση. Η υλοποίηση στέλνει αίτημα στο δεύτερο σημείο πρόσβασης της υπηρεσίας μόλις το κείμενο υπερβεί τους δύο χαρακτήρες και εμφανίζει τις προτάσεις σε κατάλογο κάτω από το πεδίο. Το Σχήμα 5.3 παρουσιάζει τον κατάλογο σε λειτουργία.

Σχήμα 5.3: Αυτόματη συμπλήρωση στο πεδίο της αφετηρίας. Οι υποδείξεις εμφανίζονται μετά τον δεύτερο χαρακτήρα και καλύπτουν τόσο την κανονική οντότητα όσο και τις επιμέρους καταχωρίσεις που περιέχουν τον ίδιο όρο.

Ο άμεσος κίνδυνος μιας τέτοιας υλοποίησης είναι η αποστολή ενός αιτήματος ανά πληκτρολόγηση. Αντιμετωπίστηκε με χρονική απόσβεση των γεγονότων: κάθε νέα πληκτρολόγηση ακυρώνει τον προηγούμενο χρονιστή και το αίτημα εκπέμπεται μόλις μεσολαβήσουν διακόσια είκοσι χιλιοστά του δευτερολέπτου χωρίς νέα εισαγωγή. Η τιμή επιλέχθηκε ώστε να μην είναι αντιληπτή ως καθυστέρηση, μειώνοντας ωστόσο τα αιτήματα κατά μία τάξη μεγέθους.

Ο κατάλογος των υποδείξεων υποστηρίζει πλήρη πλοήγηση από το πληκτρολόγιο, με τα βέλη για τη μετακίνηση της επιλογής, το πλήκτρο επιβεβαίωσης για την αποδοχή της και το πλήκτρο διαφυγής για την απόρριψη του καταλόγου, ενώ η επιλογή με το ποντίκι δεσμεύεται στο γεγονός της πίεσης και όχι της απελευθέρωσης, ώστε να μην προλάβει να μεσολαβήσει η απώλεια εστίασης του πεδίου. Η αποδοχή μιας υπόδειξης δεν συμπληρώνει απλώς το πεδίο αλλά μεταφέρει και τον χάρτη στην αντίστοιχη στάση, συνδέοντας τις δύο μεθόδους επιλογής. Το Σχήμα 5.4 δείχνει το αποτέλεσμα για το πεδίο του προορισμού.

Σχήμα 5.4: Αυτόματη συμπλήρωση στο πεδίο του προορισμού, με την αφετηρία ήδη επιλεγμένη. Ο νομός που συνοδεύει κάθε υπόδειξη προέρχεται από τη γεωκωδικοποίηση της ενότητας 4.6 και επιτρέπει στον χρήστη να διακρίνει ομόνυμες ή παραπλήσιες στάσεις πριν από την επιλογή.

5.6.3 Οπτικοποίηση διαδρομών και διαδραστικός χάρτης

Ο χάρτης υλοποιείται με τη βιβλιοθήκη Leaflet επί υποβάθρου OpenStreetMap και εμφανίζει το σύνολο των στάσεων με συντεταγμένες. Κάθε στάση αποδίδεται ως κυκλικός δείκτης, με το χρώμα και το μέγεθος να διακρίνουν εκείνες που διαθέτουν εξερχόμενα δρομολόγια από εκείνες που εμφανίζονται μόνο ως προορισμοί — διάκριση ουσιώδης, αφού οι δεύτερες δεν μπορούν να χρησιμεύσουν ως αφετηρία. Λόγω του πλήθους των σημείων αξιοποιείται η επέκταση ομαδοποίησης δεικτών, η οποία συγχωνεύει γειτονικά σημεία σε συσσωματώματα ανάλογα με το επίπεδο μεγέθυνσης και διατηρεί την απόκριση του χάρτη ομαλή. Η επιλογή μιας στάσης ανοίγει αναδυόμενο παράθυρο με το όνομά της και δύο πλήκτρα, τα οποία την ορίζουν ως αφετηρία ή ως προορισμό· το πρώτο εμφανίζεται μόνο όταν η στάση διαθέτει αναχωρήσεις.

Τα αποτελέσματα παρουσιάζονται ως κάρτες, μία ανά διαδρομή. Η κεφαλίδα κάθε κάρτας φέρει τον χαρακτηρισμό της —άμεση σύνδεση ή πλήθος ανταποκρίσεων, με ρητή σήμανση των διανυκτερεύσεων— και ακολουθεί η σύνοψη με τους συγκεντρωτικούς χρόνους. Τα σκέλη αποδίδονται διαδοχικά, με τα ενδιάμεσα διαστήματα να εμφανίζονται είτε ως αναμονή σε συγκεκριμένη στάση είτε, όταν μεσολαβεί αλλαγή ημέρας, ως διανυκτέρευση με οπτικά διακριτή σήμανση και συνοδευόμενη από τον κατάλογο των μεταγενέστερων αναχωρήσεων που υπολόγισε ο εξυπηρετητής. Στην εβδομαδιαία σύνοψη τα αποτελέσματα παρουσιάζονται ως πίνακας επτά γραμμών, καθεμία από τις οποίες αναπτύσσεται κατ' απαίτηση σε πλήρη κάρτα διαδρομής.

5.6.4 Παραμετροποίηση αναζήτησης από τον χρήστη

Οι τέσσερις παράμετροι του αλγορίθμου εκτίθενται στον χρήστη μέσω της ενότητας προχωρημένων επιλογών, με τις προεπιλεγμένες τιμές να ταυτίζονται με τις σταθερές του πυρήνα: τέσσερα σκέλη, τέσσερις ώρες μέγιστης αναμονής, δεκαπέντε λεπτά ελάχιστης ανταπόκρισης και μία διανυκτέρευση. Καθεμία αντιστοιχεί σε αριθμητικό πεδίο με ρητά όρια, ώστε να αποκλείονται τιμές που θα οδηγούσαν σε παράλογη ή δυσανάλογα δαπανηρή αναζήτηση.

Η έκθεση αυτή δεν είναι απλή τεχνική διαφάνεια αλλά λειτουργική ανάγκη, διότι οι προεπιλογές αποτελούν συμβιβασμούς που δεν εξυπηρετούν κάθε χρήστη. Ένας επιβάτης με χρονική ευχέρεια μπορεί να αυξήσει τη μέγιστη αναμονή και να αποκαλύψει συνδέσεις που η προεπιλογή απέκλειε· ένας επιβάτης με αποσκευές μπορεί να ανεβάσει τον ελάχιστο χρόνο ανταπόκρισης· και η αναζήτηση μιας διαδρομής προς απομακρυσμένη περιοχή μπορεί να απαιτεί την αύξηση των επιτρεπτών διανυκτερεύσεων. Το σύστημα καθιστά έτσι ρητό ότι μια αρνητική απάντηση δεν σημαίνει πάντοτε ανυπαρξία διαδρομής αλλά ενδεχομένως υπερβολικά αυστηρούς περιορισμούς, και προσφέρει στον χρήστη τα μέσα να το διαπιστώσει ο ίδιος.

Η ώρα αναχώρησης αντιμετωπίζεται με ανάλογη λογική: προσφέρεται ως επιλογή ώρας με ρητή δυνατότητα καθαρισμού, ώστε η προεπιλεγμένη συμπεριφορά να είναι η αναζήτηση χωρίς κανέναν χρονικό περιορισμό εντός της ημέρας.

5.6.5 Αποκρίσιμος σχεδιασμός

Η διεπαφή σχεδιάστηκε ώστε να παραμένει πλήρως λειτουργική σε οθόνες κινητών συσκευών, απαίτηση που δεν είναι διακοσμητική για μια εφαρμογή σχεδιασμού μετακινήσεων: το ερώτημα «πώς φτάνω από εδώ εκεί» τίθεται συχνά καθ' οδόν και όχι μπροστά σε σταθερό υπολογιστή.

Η προσαρμογή στηρίζεται στους μηχανισμούς διάταξης Flexbox και Grid, οι οποίοι επιτρέπουν στις ομάδες πεδίων να αναδιαταχθούν χωρίς πολλαπλές παραλλαγές σήμανσης, καθώς και σε σχετικές μονάδες μέτρησης για τις διαστάσεις και τα μεγέθη γραμματοσειράς. Ένα ερώτημα μέσω των εξακόσια σαράντα εικονοστοιχεία μεταβάλλει τη διάταξη των γραμμών της φόρμας από οριζόντια σε κατακόρυφη, συμπύκνωση το πλέγμα των προχωρημένων επιλογών σε δύο στήλες και αναδιατάσσει την κεφαλίδα του χάρτη. Η δήλωση της περιοχής προβολής εξασφαλίζει, τέλος, ότι η σελίδα αποδίδεται στο πραγματικό πλάτος της συσκευής και όχι σε πλάτος επιτραπέζιας οθόνης σε σμίκρυνση.

Δύο στοιχεία απαίτησαν ιδιαίτερη μεταχείριση: ο χάρτης, του οποίου το ύψος πρέπει να ορίζεται ρητά για να αποδοθεί και εκφράζεται γι' αυτό σε σχετικές μονάδες, και ο εγγενώς πλατύς πίνακας της εβδομαδιαίας σύνοψης, ο οποίος τοποθετείται σε περιέκτη με οριζόντια κύλιση ώστε η σελίδα να μη μετατοπίζεται ποτέ πλευρικά. Δεν επιχειρήθηκε, αντιθέτως, καμία απόκρυψη λειτουργικότητας στις μικρές οθόνες: όλες οι παράμετροι και όλα τα στοιχεία των αποτελεσμάτων παραμένουν προσβάσιμα, με μόνη διαφορά τη διάταξή τους.

Επίλογος κεφαλαίου

Το κεφάλαιο διέτρεξε την πορεία από τις απαιτήσεις έως το λειτουργικό σύστημα: κατέγραψε τι οφείλει να κάνει και με ποιες ποιοτικές ιδιότητες, διατύπωσε τις τέσσερις περιπτώσεις χρήσης που λειτούργησαν ως κριτήρια αποδοχής, παρουσίασε την τετραεπίπεδη αρχιτεκτονική και τον μηχανισμό προϋπολογισμού του γράφου, και ανέλυσε την υλοποίηση του πυρήνα — από την αναπαράσταση του χρόνου σε απόλυτα λεπτά και τη διαχείριση ανταποκρίσεων και διανυκτερεύσεων έως το στάδιο οπισθοδρομικής βελτιστοποίησης, τη λεξικογραφική κατάταξη και τις δύο ισοδύναμες διεπαφές. Κοινός άξονας των επιλογών υπήρξε η αντιστοιχία ανάμεσα στην πολυπλοκότητα του μέσου και στην πολυπλοκότητα του προβλήματος: ρητά στάδια που επικοινωνούν μέσω αρχείων και ένας πυρήνας που αγνοεί τον τρόπο με τον οποίο θα κληθεί. Ο MITOY, δηλαδή, υπάρχει· το επόμενο κεφάλαιο τον θέτει σε λειτουργία επί των πραγματικών δεδομένων, το παρουσιάζει μέσα από τα σενάρια που διατυπώθηκαν εδώ ως περιπτώσεις χρήσης και αποτιμά αν κάνει πράγματι όσα υπόσχεται.

Κεφάλαιο 6

Παρουσίαση του ΜΙΤΟΥ

Το προηγούμενο κεφάλαιο περιέγραψε πώς κατασκευάστηκε ο ΜΙΤΟΣ· το παρόν τον θέτει σε λειτουργία και εξετάζει αν κάνει όσα υπόσχεται. Παρουσιάζονται πρώτα οι προϋποθέσεις εκτέλεσης και έπειτα τα τέσσερα σενάρια της ενότητας 5.3, εκτελεσμένα επί του στιγμιότυπου δεδομένων της 23ης Ιουνίου 2026 με τα 106.380 δρομολόγια και τις 2.310 στάσεις του. Ακολουθεί η ποσοτική αποτίμηση της απόδοσης και της ορθότητας και, τέλος, η σύγκριση με τις υφιστάμενες λύσεις. Όλα τα μεγέθη που παρατίθενται προέκυψαν από εκτελέσεις του ίδιου του ΜΙΤΟΥ και όχι από εκτίμηση.

6.1 Εγκατάσταση και εκτέλεση

Μοναδική προϋπόθεση εγκατάστασης είναι η ύπαρξη διερμηνέα Python 3.10 ή νεότερου και ενός μικρού συνόλου βιβλιοθηκών. Δεν απαιτείται εξυπηρετητής βάσης δεδομένων, ούτε στάδιο μεταγλώττισης της διεπαφής, ούτε κλειδί πρόσβασης σε εμπορική υπηρεσία, σύμφωνα με τη μη λειτουργική απαίτηση της ενότητας 5.2.

Η αλυσίδα εκτελείται σε τρία βήματα: το `json_to_db.py` φορτώνει τα αρχεία JSON στη βάση, το `build_graph.py` κατασκευάζει τον γράφο και τον αποθηκεύει ως δυαδικό στιγμιότυπο, και εν συνεχεία ο ΜΙΤΟΣ είναι διαθέσιμος μέσω της γραμμής εντολών, με το `route_finder.py`, ή της διαδικτυακής εφαρμογής, με το `UI/app.py`. Η λειτουργία `build_graph.py --check` επιβεβαιώνει ποιο στιγμιότυπο βρίσκεται σε χρήση — έλεγχος απαραίτητος, καθώς κάθε αριθμός του κεφαλαίου αναφέρεται σε συγκεκριμένη έκδοση των δεδομένων.

6.2 Σενάριο 1: άμεση σύνδεση

Το πρώτο σενάριο αφορά το ερώτημα που ήδη απαντούν οι υφιστάμενοι ιστότοποι και λειτουργεί ως έλεγχος βάσης. Για τη διαδρομή Θεσσαλονίκη προς Κατερίνη τη Δευτέρα από τις 08:00, το σύστημα επιστρέφει πέντε απευθείας συνδέσεις, με πρώτη την αναχώρηση των 08:10 και άφιξη στις 08:55.

Ενδιαφέρον παρουσιάζει η δεύτερη πρόταση: αναχώρηση στις 08:15 με άφιξη επίσης στις 08:55, διάρκειας σαράντα αντί σαράντα πέντε λεπτών, η οποία όμως δεν εκτελείται από το ΚΤΕΛ Πιερίας αλλά από το ΚΤΕΛ Αχαΐας και δημοσιεύεται σε διαφορετική πλατφόρμα, καθώς πρόκειται για διερχόμενο δρομολόγιο μεγάλης απόστασης που εξυπηρετεί την Κατερίνη ως ενδιάμεσο σταθμό. Ήδη, δηλαδή, στην απλούστερη περίπτωση η ενοποίηση των πηγών αποδίδει την ταχύτερη σύνδεση, την οποία κανένας μεμονωμένος ιστότοπος δεν εμφανίζει.

6.3 Σενάριο 2: διαδρομή με μία ή περισσότερες ανταποκρίσεις

Το δεύτερο σενάριο ενεργοποιεί τον πυρήνα της συνεισφοράς. Για τη διαδρομή Θεσσαλονίκη προς Σπάρτη τη Δευτέρα από τις 06:00, η βέλτιστη λύση συντίθεται από δύο σκέλη: αναχώρηση στις 07:15 με άφιξη στην Αθήνα στις 12:51 από το ΚΤΕΛ Θεσσαλονίκης, αναμονή τριάντα εννέα λεπτών, και αναχώρηση στις 13:30 για τη Σπάρτη με άφιξη στις 16:30 από το ΚΤΕΛ Λακωνίας. Ο συνολικός χρόνος ανέρχεται σε εννέα ώρες και δεκαπέντε λεπτά, εκ των οποίων μόλις τα τριάντα εννέα λεπτά είναι αναμονή. Το Σχήμα 6.1 παρουσιάζει τη λύση όπως αποδίδεται στη διεπαφή.

Θεσσαλονίκη → ΚΤΕΛ ΣΠΑΡΤΗΣ

Ημέρα	Διαθεσιμότητα	Καλύτερη διαδρομή	
Δευτέρα	Βρέθηκε διαδρομή	Δε 07:15 → Δε 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Τρίτη	Βρέθηκε διαδρομή	Τρ 07:15 → Τρ 16:30 (9ω 15λ, 1 ανταπόκριση)	Απόκρυψη ▴

Διαδρομή #1
1 ανταπόκριση

Τρ 07:15 → Τρ 16:30 (9ω 15λ, 1 ανταπόκριση)

Συνολικός χρόνος: 9ω 15λ · Ταξίδι: 8ω 36λ · Αναμονές: 39λ

Θεσσαλονίκη → Αθήνα
 Τρίτη 07:15 → Τρίτη 12:51 | 5ω 36λ
 ΚΤΕΛ Θεσσαλονίκης (e-ticketing.gr)

Αναμονή στον «Αθήνα»: 39λ

Αθήνα → ΚΤΕΛ ΣΠΑΡΤΗΣ
 Τρίτη 13:30 → Τρίτη 16:30 | 3ω 00λ
 ΚΤΕΛ Λακωνίας (ktelbus.gr)

Άφιξη: Τρίτη 16:30 στον «ΚΤΕΛ ΣΠΑΡΤΗΣ»

Τετάρτη	Βρέθηκε διαδρομή	Τε 07:15 → Τε 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Πέμπτη	Βρέθηκε διαδρομή	Πε 07:15 → Πε 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Παρασκευή	Βρέθηκε διαδρομή	Πα 07:15 → Πα 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Σάββατο	Βρέθηκε διαδρομή	Σα 07:15 → Σα 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Κυριακή	Βρέθηκε διαδρομή	Κυ 07:15 → Κυ 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾

Σχήμα 6.1: Η διαδρομή Θεσσαλονίκη προς ΚΤΕΛ Σπάρτης όπως αποδίδεται στη διεπαφή, με τη γραμμή της Τρίτης αναπτυγμένη σε πλήρη κάρτα· η λύση είναι πανομοιότυπη με εκείνη της Δευτέρας που περιγράφεται στο κείμενο. Διακρίνονται τα δύο σκέλη, ο φορέας και η πλατφόρμα

προέλευσης καθενός, η ενδιάμεση αναμονή των τριάντα εννέα λεπτών στην Αθήνα και οι συγκεντρωτικοί χρόνοι της διαδρομής.

Η απάντηση δεν μπορεί να παραχθεί από καμία υφιστάμενη πηγή: τα δύο σκέλη ανήκουν σε διαφορετικούς φορείς, δημοσιεύονται σε διαφορετικές πλατφόρμες, και ο κόμβος μετεπιβίβασης προέκυψε από τον υπολογισμό και όχι από δήλωση του χρήστη. Το σύστημα εντοπίζει επιπλέον εναλλακτική μέσω Πάτρας, με μετεπιβίβαση από το ΚΤΕΛ Αχαΐας στο ΚΤΕΛ Αρκαδίας, καθώς και λύσεις τριών σκελών μέσω Κατερίνης.

6.4 Σενάριο 3: διαδρομή με διανυκτέρευση

Για τη διαδρομή από τη Σάρτη Χαλκιδικής προς την Καφιόνα Λακωνίας την Παρασκευή από τις 06:00, καμία σύνδεση δεν ολοκληρώνεται αυθημερόν. Η λύση περιλαμβάνει αναχώρηση στις 07:00 με άφιξη στη Θεσσαλονίκη στις 10:20, μετεπιβίβαση στο δρομολόγιο των 12:00 προς την Αθήνα με άφιξη στις 17:36, διανυκτέρευση εκεί, και αναχώρηση το Σάββατο στις 06:30 με τελική άφιξη στις 12:54.

Δύο στοιχεία αξίζουν σχολιασμού. Πρώτον, η διανυκτέρευση σημαίνεται ρητά και διακρίνεται από την απλή αναμονή, ώστε ο χρήστης να αντιλαμβάνεται ότι πρόκειται για διαμονή και όχι για δεκατρίωρη παραμονή σε σταθμό. Δεύτερον, η αναμονή μίας ώρας και σαράντα λεπτών στη Θεσσαλονίκη διατηρείται αντί να συμπιεστεί: το σκέλος προηγείται διανυκτέρευσης και, κατά την επιλογή της ενότητας 5.5.5, η οπισθοδρομική βελτιστοποίηση αναστέλλεται εκεί ώστε ο επιβάτης να φτάνει νωρίτερα στον κόμβο διαμονής.

6.5 Σενάριο 4: εβδομαδιαία σύνοψη διαθεσιμότητας

Το τέταρτο σενάριο απαντά στο ερώτημα «ποιες ημέρες μπορώ να πάω». Για τη διαδρομή από τις Λιβανάτες προς τα Ιωάννινα, η σύνοψη αποκαλύπτει έντονη ασυμμετρία μέσα στην εβδομάδα: από Δευτέρα έως Παρασκευή η μετακίνηση ολοκληρώνεται αυθημερόν με δύο ανταποκρίσεις, σε οκτώ ώρες και σαράντα τρία λεπτά τις δύο πρώτες ημέρες και σε δέκα ώρες και πενήντα πέντε λεπτά τις υπόλοιπες, ενώ το Σάββατο και την Κυριακή καμία λύση δεν ολοκληρώνεται εντός της ημέρας και η ταχύτερη απαιτεί τρεις ανταποκρίσεις και μία διανυκτέρευση. Η μορφή με την οποία αποδίδεται η σύνοψη φαίνεται στο Σχήμα 6.2.

Θεσσαλονίκη → ΚΤΕΛ ΣΠΑΡΤΗΣ			
Ημέρα	Διαθεσιμότητα	Καλύτερη διαδρομή	
Δευτέρα	Βρέθηκε διαδρομή	Δε 07:15 → Δε 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Τρίτη	Βρέθηκε διαδρομή	Τρ 07:15 → Τρ 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Τετάρτη	Βρέθηκε διαδρομή	Τε 07:15 → Τε 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Πέμπτη	Βρέθηκε διαδρομή	Πε 07:15 → Πε 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Παρασκευή	Βρέθηκε διαδρομή	Πα 07:15 → Πα 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Σάββατο	Βρέθηκε διαδρομή	Σα 07:15 → Σα 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾
Κυριακή	Βρέθηκε διαδρομή	Κυ 07:15 → Κυ 16:30 (9ω 15λ, 1 ανταπόκριση)	Αναλυτικά ▾

Σχήμα 6.2: Η εβδομαδιαία σύνοψη διαθεσιμότητας σε συνεπτυγμένη μορφή, εδώ για τη διαδρομή της ενότητας 6.3: μία γραμμή ανά ημέρα με την ένδειξη διαθεσιμότητας και τη βέλτιστη διαδρομή της ημέρας, η οποία αναπτύσσεται κατ' απαίτηση σε πλήρη κάρτα.

Η πληροφορία αυτή είναι εκείνη ακριβώς που χρειάζεται ο επιβάτης πριν επιλέξει ημερομηνία, και καμία υφιστάμενη διεπαφή δεν την παρέχει: οι μηχανές κράτησης απαντούν μόνο για συγκεκριμένη ημερομηνία, οπότε η ίδια διαπίστωση θα απαιτούσε επτά χωριστές αναζητήσεις σε καθέναν από τους φορείς.

6.6 Χρήση μέσω γραμμής εντολών

Η διεπαφή γραμμής εντολών δέχεται την αφετηρία και τον προορισμό ως θεσιακά ορίσματα και τις παραμέτρους της αναζήτησης ως προαιρετικές σημαίες, ενώ η παράλειψη της ημέρας ενεργοποιεί αυτομάτως την εβδομαδιαία σύνοψη.

Δύο συμπεριφορές αναδεικνύουν την πρακτική της αξία. Όταν η αναζήτηση αποτύχει, το σύστημα διαγιγνώσκει την αιτία: για ερώτημα με αφετηρία την Ορεστιάδα αναφέρει ότι η στάση δεν διαθέτει καμία αναχώρηση στη βάση, υποδεικνύοντας ότι το πρόβλημα εντοπίζεται στην κάλυψη των δεδομένων και όχι στις παραμέτρους. Όταν πάλι το κείμενο του χρήστη ταιριάζει σε περισσότερες κανονικές οντότητες, η επιλογή που έγινε δηλώνεται με ρητή προειδοποίηση, ώστε η ασάφεια να μην περνά απαρατήρητη.

6.7 Αξιολόγηση απόδοσης

Οι μετρήσεις εκτελέστηκαν σε φορητό υπολογιστή με επεξεργαστή AMD Ryzen 7 5800H και 16 GB μνήμης, υπό Windows 11 και Python 3.14. Η φόρτωση του έτοιμου γράφου από το στιγμιότυπο απαιτεί 0,27 δευτερόλεπτα, έναντι 1,10 δευτερολέπτων για την πλήρη ανακατασκευή του από τη βάση.

Η καθαυτό αναζήτηση μετρήθηκε σε εξακόσια τυχαία ζεύγη αφετηρίας και προορισμού. Ο μέσος χρόνος απόκρισης ανήλθε σε 274 χιλιοστά του δευτερολέπτου με ημέρα εκκίνησης Δευτέρα και 257 με Κυριακή, με ενενηκοστό πέμπτο εκατοστημόριο 540 και 582 χιλιοστά και μέγιστη παρατηρηθείσα τιμή 795. Ο στόχος της ενότητας 5.2, δηλαδή απόκριση σαφώς κάτω του δευτερολέπτου, ικανοποιείται σε ολόκληρο το δείγμα χωρίς καμία βελτιστοποίηση χαμηλού επιπέδου, επιβεβαιώνοντας την εκτίμηση της ενότητας 3.1 ότι η ταχύτητα της Python δεν αποτελεί περιοριστικό παράγοντα στην κλίμακα του προβλήματος. Η εβδομαδιαία σύνοψη, που συνίσταται σε επτά διαδοχικές αναζητήσεις, ολοκληρώνεται σε περίπου 1,5 δευτερόλεπτο, ενώ η αναζήτηση στάσης για την αυτόματη συμπλήρωση απαιτεί περί τα εννέα χιλιοστά και είναι ανεπαίσθητη κατά την πληκτρολόγηση.

Πίνακας 6.1: Χρόνοι απόκρισης της αναζήτησης σε δείγμα 600 τυχαίων ζευγών αφετηρίας και προορισμού. Μέγιστη παρατηρηθείσα τιμή σε ολόκληρο το δείγμα: 795 ms.

Ημέρα εκκίνησης	Μέσος χρόνος (ms)	95ο εκατοστημόριο (ms)
Δευτέρα	274	540
Κυριακή	257	582

6.8 Αξιολόγηση ορθότητας έναντι πραγματικών δρομολογίων

Η ορθότητα ελέγχθηκε με αυτοματοποιημένη επαλήθευση αναλλοίωτων επί τετρακοσίων τυχαίων ερωτημάτων, τα οποία παρήγαγαν 1.144 διαδρομές και 3.533 σκέλη. Για κάθε σκέλος επιβεβαιώθηκε ότι αντιστοιχεί σε υπαρκτή εγγραφή του συνόλου δεδομένων, ότι η ημέρα αναχώρησης ανήκει στις ημέρες εκτέλεσής του, ότι η άφιξη ισούται με την αναχώρηση συν τη δημοσιευμένη διάρκεια και ότι κάθε μετεπιβίβαση τηρεί το ελάχιστο περιθώριο των δεκαπέντε λεπτών. Καμία παραβίαση δεν εντοπίστηκε σε κανέναν από τους ελέγχους αυτούς.

Ένα εύρημα, ωστόσο, οφείλει να δηλωθεί: σε ένδεκα σκέλη, ήτοι 0,3% του δείγματος, η αναμονή ίδιας ημέρας υπερβαίνει το όριο των τεσσάρων ωρών. Πριν από τη μετα-βελτιστοποίηση οι παραβιάσεις ήταν μηδενικές, γεγονός που εντοπίζει την αιτία: το σφίξιμο ενός σκέλους συμπιέζει την αναμονή που έπεται, διευρύνοντας όμως εκείνη που προηγείται, και όταν το προηγούμενο σκέλος δεν διαθέτει μεταγενέστερη εναλλακτική η διεύρυνση δεν απορροφάται. Το όριο επιβάλλεται κατά την αναζήτηση αλλά δεν επανελέγχεται μετά από αυτήν — πραγματικό, αν και σπάνιο, ελάττωμα που αίρεται με έναν τελικό έλεγχο συμβατότητας.

Αποτιμήθηκε χωριστά η μετα-βελτιστοποίηση, κεντρικό στοιχείο της υλοποίησης. Σε δείγμα 1.123 επιστρεφόμενων διαδρομών δύο ή περισσότερων σκελών, βελτίωσε το 15,6% από αυτές, μειώνοντας τη μέση αναμονή ίδιας ημέρας από 139 σε 125 λεπτά· όπου ενεργοποιήθηκε, το μέσο κέρδος ανήλθε σε 93 λεπτά και το μέγιστο σε τέσσερις ώρες και σαράντα τρία λεπτά. Επιβεβαιώθηκε ταυτόχρονα ότι σε καμία περίπτωση δεν μεταβλήθηκε ο χρόνος τελικής άφιξης, όπως απαιτεί η ενότητα 2.9.4.

Ελέγχθηκε, τέλος, η ποιότητα των διαρκειών που συμπληρώθηκαν μέσω OSRM, ζήτημα που η ενότητα 4.7 παρέπεμψε ρητά εδώ. Για 125 ζεύγη στάσεων υπάρχει τόσο δημοσιευμένη διάρκεια

όσο και εκτίμηση· η μέση απόκλιση είναι +13 λεπτά και η διάμεση +11, με το 52% των ζευγών εντός ± 15 λεπτών. Η εκτίμηση τείνει, δηλαδή, να υπερεκτιμά τη διάρκεια — κατεύθυνση ασφαλής, αφού οδηγεί σε απώλεια οριακών ανταποκρίσεων και όχι σε πρόταση που ο επιβάτης δεν προλαβαίνει.

Πίνακας 6.2: Αποτελέσματα της αυτοματοποιημένης επαλήθευσης αναλλοίωτων επί 400 τυχαίων ερωτημάτων, τα οποία παρήγαγαν 1.144 διαδρομές και 3.533 σκέλη.

Έλεγχος	Παραβιάσεις
Αντιστοιχία σκέλους σε υπαρκτή εγγραφή	0
Ημέρα αναχώρησης εντός των ημερών εκτέλεσης	0
Άφιξη = αναχώρηση + δημοσιευμένη διάρκεια	0
Ελάχιστο περιθώριο μετεπιβίβασης (15 λεπτά)	0
Όριο αναμονής ίδιας ημέρας (4 ώρες)	11 (0,3%)

Πίνακας 6.3: Αποτίμηση της μετα-βελτιστοποίησης σε δείγμα 1.123 διαδρομών δύο ή περισσότερων σκελών.

Μέγεθος	Τιμή
Διαδρομές που βελτιώθηκαν	15,6%
Μέση αναμονή ίδιας ημέρας — πριν	139 λεπτά
Μέση αναμονή ίδιας ημέρας — μετά	125 λεπτά
Μέσο κέρδος όπου ενεργοποιήθηκε	93 λεπτά
Μέγιστο κέρδος	4 ώρες 43 λεπτά
Μεταβολή χρόνου τελικής άφιξης	καμία

6.9 Σύγκριση με υφιστάμενες λύσεις

Η ουσιαστικότερη σύγκριση δεν αφορά την ταχύτητα αλλά το εύρος των ερωτημάτων που μπορούν καν να απαντηθούν. Σε πεντακόσια τυχαία ερωτήματα, το 92,2% των λύσεων που εντοπίστηκαν συνθέτει δρομολόγια περισσότερων του ενός ΚΤΕΛ, με τις περισσότερες να εμπλέκουν δύο ή τρεις φορείς. Εννέα, δηλαδή, στις δέκα απαντήσεις βρίσκονται εξ ορισμού εκτός της εμβέλειας κάθε μεμονωμένου ιστότοπου ΚΤΕΛ, ο οποίος γνωρίζει αποκλειστικά τα δικά του δρομολόγια.

Η ίδια μέτρηση, επαναλαμβανόμενη για καθεμία από τις επτά ημέρες εκκίνησης, αναδεικνύει και το πόσο αραιό είναι το δίκτυο: το ποσοστό των ζευγών για τα οποία υπάρχει οποιαδήποτε λύση κυμαίνεται από 87% τη Δευτέρα έως 50% το Σάββατο, διαφορά δυσανάλογη της μεταβολής των

δρομολογίων, αφού η Δευτέρα συγκεντρώνει το 78,5% των αναχωρήσεων και το Σάββατο το 48,9%. Σε αραιό δίκτυο, δηλαδή, μια μέτρια μείωση της προσφοράς αποδιαρθρώνει δυσανάλογα τις αλυσίδες ανταποκρίσεων.

Έναντι των υφιστάμενων επιλογών η διαφορά είναι, επομένως, κατηγορική και όχι βαθμού. Οι ιστότοποι των φορέων απαντούν σωστά αλλά μόνο εντός του δικού τους δικτύου και μόνο για δεδομένη ημερομηνία, ενώ οι διεθνείς σχεδιαστές ταξιδιού, μη διαθέτοντας τα ελληνικά υπεραστικά χρονοδιαγράμματα σε μηχανικά αναγνώσιμη μορφή, καλύπτουν το δίκτυο αποσπασματικά ή καθόλου. Ο MITOS δεν υπερτερεί, συνεπώς, σε κάτι που οι υπάρχουσες λύσεις κάνουν χειρότερα· απαντά σε ερώτημα που εκείνες δεν θέτουν.

Πίνακας 6.4: Διαθεσιμότητα λύσεων ανά ημέρα εκκίνησης — τα δύο άκρα του παρατηρηθέντος εύρους.

Ημέρα εκκίνησης	Ζεύγη με λύση	Μερίδιο αναχωρήσεων
Δευτέρα	87%	78,5%
Σάββατο	50%	48,9%

Επίλογος κεφαλαίου

Το κεφάλαιο έθεσε τον MITO σε λειτουργία επί των πραγματικών δεδομένων και τον αποτίμησε. Τα τέσσερα σενάρια επιβεβαίωσαν ότι καλύπτονται όλες οι περιπτώσεις χρήσης της ενότητας 5.3, από την απευθείας σύνδεση έως τη διαδρομή με διανυκτέρευση και την εβδομαδιαία σύνοψη. Η απόδοση παρέμεινε εντός του ορίου διαδραστικότητας, ενώ ο έλεγχος ορθότητας δεν ανέδειξε καμία παραβίαση των θεμελιωδών περιορισμών του μοντέλου, παρά μόνο ένα σπάνιο και εντοπισμένο ελάττωμα της μετα-βελτιστοποίησης. Ποσοτικοποιήθηκαν, τέλος, το όφελος της οπισθοδρομικής βελτιστοποίησης και η ακρίβεια των εκτιμώμενων διαρκειών. Το επόμενο κεφάλαιο συνοψίζει τα συμπεράσματα, αναγνωρίζει τους περιορισμούς που αναδείχθηκαν εδώ και διατυπώνει προτάσεις για τη συνέχεια.

Κεφάλαιο 7

Συμπεράσματα και Μελλοντικές επεκτάσεις

Η εργασία ξεκίνησε από μια διαπίστωση καθημερινής εμπειρίας — ότι δεν υπάρχει τρόπος να απαντηθεί το ερώτημα «πώς φτάνω από το Α στο Β με υπεραστικό λεωφορείο» όταν η μετακίνηση διασχίζει περισσότερες από μία περιοχές ευθύνης ΚΤΕΛ — και κατέληξε στον ΜΙΤΟ, ένα λειτουργικό σύστημα που το απαντά για το σύνολο της ηπειρωτικής χώρας. Το παρόν κεφάλαιο συνοψίζει τη διαδρομή αυτή, αποτιμά κατά πόσο εκπληρώθηκαν οι στόχοι της ενότητας 1.5, δηλώνει ρητά τα όρια εντός των οποίων ισχύουν τα αποτελέσματα, και διατυπώνει τις κατευθύνσεις προς τις οποίες η προσπάθεια μπορεί να συνεχιστεί — τεχνικές αλλά και θεσμικές.

7.1 Σύνοψη της εργασίας

Το πρώτο κεφάλαιο οριοθέτησε το πρόβλημα: ένα υπεραστικό δίκτυο που προέκυψε ως άθροισμα αυτοτελών νομαρχιακών δικτύων, χωρίς ενιαίο μητρώο στάσεων, χωρίς συντονισμό ανταποκρίσεων και χωρίς καμία μηχανικά αναγνώσιμη δημοσίευση των χρονοδιαγραμμάτων του. Ο πυρήνας της δυσκολίας εντοπίστηκε σε μια κυκλική εξάρτηση: για να ρωτήσει κανείς τον σωστό φορέα πρέπει να γνωρίζει ήδη τον ενδιαμέσο κόμβο, ο οποίος όμως είναι ακριβώς το ζητούμενο.

Το δεύτερο κεφάλαιο συγκρότησε το θεωρητικό υπόβαθρο, από τους ορισμούς της θεωρίας γράφων και τον αλγόριθμο Dijkstra έως τα χρονοεξαρτώμενα δίκτυα, την ιδιότητα FIFO και τα δύο καθιερωμένα μοντέλα αναπαράστασης δικτύων χρονοδιαγράμματος, και κατέληξε στην τυπική διατύπωση του προτεινόμενου αλγορίθμου, στη θεμελίωση της ορθότητάς του και στη ρητή δήλωση των παραδοχών του. Το τρίτο κεφάλαιο τεκμηρίωσε τις τεχνολογικές επιλογές ως απαντήσεις σε συγκεκριμένα προβλήματα και όχι ως προτιμήσεις.

Το τέταρτο κεφάλαιο περιέγραψε το πλέον χρονοβόρο στάδιο: την αντίστροφη μηχανίκευση τεσσάρων μη τεκμηριωμένων πλατφορμών, συμπεριλαμβανομένης της αποκρυπτογράφησης μίας εξ αυτών, την ανακατασκευή του εβδομαδιαίου προτύπου εκτέλεσης μέσω δειγματοληψίας επτά ημερών, και την ενοποίηση των οντοτήτων στάσεων με συνδυασμό γεωκωδικοποίησης, χωρικής ομαδοποίησης και ημιαυτόματης επικύρωσης. Αποτέλεσμα υπήρξε ένα ενιαίο σύνολο 106.380 δρομολογίων από 35 ΚΤΕΛ, με 2.310 κόμβους και 353 κανονικές οντότητες στάσεων.

Το πέμπτο κεφάλαιο μετέτρεψε το σύνολο αυτό στον ΜΙΤΟ, ένα σύστημα τεσσάρων επιπέδων, με τον χρόνο εκφρασμένο σε απόλυτα λεπτά, ρητή διαχείριση ανταποκρίσεων και διανυκτερεύσεων,

στάδιο οπισθοδρομικής βελτιστοποίησης, λεξικογραφική κατάταξη των υποψηφίων και δύο ισοδύναμες διεπαφές, μία διαδικτυακή με διαδραστικό χάρτη και μία γραμμής εντολών. Το έκτο κεφάλαιο, τέλος, έθεσε τον ΜΙΤΟ σε λειτουργία επί των πραγματικών δεδομένων: επιβεβαίωσε ότι καλύπτονται και οι τέσσερις περιπτώσεις χρήσης, από την απευθείας σύνδεση έως τη διαδρομή με διανυκτέρευση και την εβδομαδιαία σύνοψη διαθεσιμότητας, και αποτίμησε ποσοτικά την απόδοση, την ορθότητα και το εύρος των ερωτημάτων που καθιστά για πρώτη φορά απαντήσιμα.

7.2 Συμπεράσματα

7.2.1 Εκπλήρωση των στόχων

Ο πρώτος στόχος, η συστηματική ανάκτηση των δεδομένων, εκπληρώθηκε πλήρως: καταγράφηκαν και ταξινομήθηκαν οι πηγές κατά τεχνολογική υποδομή και αναπτύχθηκε μηχανισμός συλλογής για καθεμία από τις τέσσερις οικογένειες, με τη συλλογή να ολοκληρώνεται σε ώρες και με μόλις δύο επίμονες αστοχίες σε δεκάδες χιλιάδες κλήσεις. Ο δεύτερος, η ενοποίηση σε ενιαίο μοντέλο, επιτεύχθηκε με την αντιστοίχιση 1.157 επιμέρους καταχωρίσεων σε 353 κανονικές οντότητες και τη φόρτωσή τους σε σχεσιακή βάση με εγγυημένη ταυτοδυναμία: ουσιώδες στοιχείο υπήρξε ότι η αντιστοίχιση κλειδώνεται στην τριάδα πηγής, φορέα και ονόματος, ώστε τα ομώνυμα τοπωνύμια να μη συγχέονται ποτέ.

Ο τρίτος και ο τέταρτος στόχος, η τυπική διατύπωση και η ανάπτυξη του αλγορίθμου, καλύφθηκαν στην ενότητα 2.9 και υλοποιήθηκαν στην ενότητα 5.5, με τεκμηριωμένη ορθότητα εντός κάθε κλάσης πλήθους σκελών και με το στάδιο μετα-βελτιστοποίησης να αποτελεί το διακριτό στοιχείο της υλοποίησης. Ο πέμπτος στόχος υλοποιήθηκε ως διαδικτυακή εφαρμογή με διαδραστικό χάρτη και ως ισοδύναμη διεπαφή γραμμής εντολών.

Ο έκτος στόχος, η αξιολόγηση, τεκμηριώνεται από τις μετρήσεις του έκτου κεφαλαίου: μέσος χρόνος απόκρισης 274 χιλιοστών του δευτερολέπτου με μέγιστη παρατηρηθείσα τιμή 795, καμία παραβίαση των θεμελιωδών περιορισμών σε 3.533 ελεγχθέντα σκέλη, ποσοτικοποιημένο όφελος της οπισθοδρομικής βελτιστοποίησης και χωριστή αποτίμηση της ακρίβειας των εκτιμώμενων διαρκειών. Η σύγκριση με τις υφιστάμενες λύσεις έδειξε, εξάλλου, ότι το 92,2% των διαδρομών που εντοπίζονται συνθέτει δρομολόγια περισσότερων του ενός φορέα, δηλαδή βρίσκεται εξ ορισμού εκτός της εμβέλειας κάθε μεμονωμένου ιστοτόπου ΚΤΕΛ. Και οι έξι στόχοι, επομένως, θεωρούνται εκπληρωμένοι, με τη διευκρίνιση ότι ο έλεγχος ορθότητας ανέδειξε ένα σπάνιο και εντοπισμένο ελάττωμα, το οποίο δηλώνεται στην ενότητα 7.3.2.

7.2.2 Τεχνικά και ερευνητικά συμπεράσματα

Το πρώτο και σημαντικότερο συμπέρασμα είναι ότι το εμπόδιο δεν ήταν αλγοριθμικό. Ο υπολογιστικός πυρήνας γράφτηκε και επαληθεύτηκε σε κλάσμα του χρόνου που απαιτήσε η συγκρότηση των δεδομένων, και αποδίδει σε διαδραστικό χρόνο χωρίς καμία βελτιστοποίηση χαμηλού επιπέδου· η πολυπλοκότητα των σύγχρονων αλγορίθμων του πεδίου, όπως τεκμηριώθηκε στην ενότητα 2.7.4, δεν δικαιολογείται στην κλίμακα του προβλήματος. Αντιθέτως, η ταύτιση των οντοτήτων στάσεων αποδείχθηκε το πραγματικό κρίσιμο σημείο, με την έννοια ότι κάθε αστοχία της αποκόπτει ή κατασκευάζει συνδέσεις σε ολόκληρο το δίκτυο.

Δεύτερον, επιβεβαιώθηκε ότι η αραιότητα μεταβάλλει ποιοτικά το πρόβλημα. Σε πυκνά αστικά δίκτυα ο χρόνος αναμονής είναι αμελητέος και το ερευνητικό ενδιαφέρον μετατοπίζεται στην ταχύτητα απόκρισης· εδώ η αναμονή κυριαρχεί, και η μετα-βελτιστοποίηση βελτίωσε το 15,6% των διαδρομών με μέσο κέρδος 93 λεπτά χωρίς καμία μεταβολή της ώρας άφιξης. Η ίδια αραιότητα εξηγεί και το εύρημα της ενότητας 6.9: η μείωση των αναχωρήσεων από 78,5% τη Δευτέρα σε 48,9% το Σάββατο αντιστοιχεί σε πτώση της καλυπτόμενης ζήτησης από 87% σε 50%, δηλαδή σε δυσανάλογη αποδιάρθρωση των αλυσίδων ανταπόκρισης.

Τρίτον, η ασυμμετρία του κόστους των σφαλμάτων αποδείχθηκε γόνιμη σχεδιαστική αρχή. Από την αντιστοίχιση των στάσεων έως τη μεταχείριση των σκελών άγνωστης διάρκειας, η συστηματική προτίμηση της χαμένης έναντι της ανύπαρκτης σύνδεσης παρήγαγε ένα σύστημα του οποίου οι αστοχίες είναι παραλείψεις και όχι ψευδείς υποσχέσεις.

Τέταρτον, η αυστηρά μονόδρομη αρχιτεκτονική διακριτών σταδίων που επικοινωνούν μέσω αυτοτελών αρχείων αποδείχθηκε πολυτιμότερη από όσο υποδήλωνε η απλότητά της. Επέτρεψε την απομόνωση κάθε σφάλματος στο στάδιο που το παρήγαγε και, κυρίως, την επανειλημμένη αναθεώρηση των κανόνων ενοποίησης χωρίς νέα προσφυγή στις πηγές — ιδιότητα καθοριστική όταν το κόστος κάθε συλλογής μετριέται σε ώρες και εξαρτάται από υπηρεσίες τρίτων.

7.2.3 Η κατάσταση των ανοικτών συγκοινωνιακών δεδομένων στην Ελλάδα

Το τέταρτο κεφάλαιο επιτρέπει μια εμπειρικά θεμελιωμένη διαπίστωση: τα δεδομένα υπάρχουν. Είναι ψηφιακά, δομημένα, ενημερώνονται καθημερινά για εμπορικούς σκοπούς και εκτίθενται ήδη μέσω υπηρεσιών που επιστρέφουν, σε τρεις από τις τέσσερις περιπτώσεις, καθαρά μηχανικά επεξεργασίμες αποκρίσεις. Δεν λείπει η πληροφορία· λείπει η δημοσίευσή της σε κοινή, τεκμηριωμένη και σταθερή μορφή.

Η απόσταση μεταξύ των δύο καταστάσεων είναι μικρότερη απ' όσο υποδηλώνει η δυσκολία της παρούσας εργασίας. Το μεγαλύτερο μέρος του κόστους δεν αναλώθηκε στην άντληση καθεαυτήν αλλά στην ανασύσταση πληροφορίας που οι φορείς ήδη κατέχουν και δεν εκθέτουν: το εβδομαδιαίο πρότυπο εκτέλεσης, το οποίο ανακατασκευάστηκε με δειγματοληψία επτά ημερών επειδή καμία πηγή δεν το δηλώνει· τις διάρκειες διαδρομής, που εκτιμήθηκαν μέσω OSRM για δύο πηγές οι οποίες δημοσιεύουν μόνο ώρα αναχώρησης· και την ταυτότητα των στάσεων, που επιλύθηκε με γεωκωδικοποίηση και χειροκίνητη επικύρωση επειδή δεν υφίσταται εθνικό μητρώο.

Η αντιπαράβολή με τις αστικές συγκοινωνίες των δύο μεγάλων κέντρων, οι οποίες δημοσιεύουν GTFS και είναι ορατές στους διεθνείς σχεδιαστές ταξιδιού, καθιστά σαφές ότι το εμπόδιο δεν είναι ούτε τεχνολογικό ούτε οικονομικό αλλά θεσμικό: απουσιάζει η υποχρέωση και το κοινό πρότυπο. Το γεγονός ότι η ενοποίηση κατέστη εφικτή με τα μέσα μιας διπλωματικής εργασίας αποτελεί, από την άποψη αυτή, το ισχυρότερο επιχείρημα υπέρ της θεσμικής λύσης.

7.3 Περιορισμοί

7.3.1 Περιορισμοί που πηγάζουν από τις πηγές δεδομένων

Η κάλυψη περιορίζεται στην ηπειρωτική χώρα και στα 35 ΚΤΕΛ που εξυπηρετούνται από τις τέσσερις εντοπισμένες υποδομές. Τα νησιωτικά ΚΤΕΛ απουσιάζουν πλήρως, ενώ δύο φορείς της πλατφόρμας ΙΜΕ δεν επέστρεψαν κανένα δρομολόγιο και εξαιρέθηκαν. Στάσεις χωρίς εξερχόμενες αναχωρήσεις —όπως η περίπτωση της Ορεστιάδας στην ενότητα 6.6— δεν συνιστούν αδυναμία του αλγορίθμου αλλά κενό των δεδομένων.

Η κάλυψη είναι, επιπλέον, άνιση εντός του δικτύου, καθώς η κατανομή των εγγραφών ανά πηγή είναι έντονα ασύμμετρη: η πλατφόρμα ΙΜΕ συνεισφέρει το μεγαλύτερο μέρος του συνόλου, ενώ οι δύο μικρότερες πηγές αθροίζουν λιγότερες από χίλιες τριακόσιες εγγραφές, γεγονός που αντανακλά όχι μόνο το μέγεθος των αντίστοιχων δικτύων αλλά και τη διαφορετική λεπτομέρεια με την οποία κάθε πλατφόρμα εκθέτει τις ενδιάμεσες στάσεις των γραμμών της.

Ποιοτικοί περιορισμοί υπάρχουν και εντός της κάλυψης. Τα δεδομένα αποτυπώνουν μεμονωμένες συνδέσεις ζευγών στάσεων και όχι τη σύνθεση κάθε δρομολογίου ως ακολουθίας ενδιάμεσων στάσεων, γεγονός που εμποδίζει την άμεση εξαγωγή σε GTFS. Οι διάρκειες που συμπληρώθηκαν μέσω OSRM φέρουν συστηματική απόκλιση περίπου δεκατριών λεπτών προς τα άνω, κατεύθυνση μεν ασφαλής αλλά αιτία απώλειας οριακών ανταποκρίσεων. Το κόστος του εισιτηρίου δεν συλλέγεται καθόλου. Τέλος, ολόκληρη η υποδομή συλλογής παραμένει εύθραυστη: μια ανασχεδίαση ιστοτόπου αρκεί για να ακυρώσει τον αντίστοιχο μηχανισμό, χωρίς καμία προειδοποίηση.

7.3.2 Παραδοχές του μοντέλου

Ισχύουν, πρώτον, οι παραδοχές που δηλώθηκαν στην ενότητα 2.9.6: ενιαίος χρόνος ανταπόκρισης δεκαπέντε λεπτών ανεξαρτήτως σταθμού, αποκλειστική στήριξη στο προγραμματισμένο χρονοδιάγραμμα, και αποκοπή του χώρου αναζήτησης μέσω του άνω φράγματος σκελών και του ορίζοντα ημερών. Οι δύο πρώτες είναι αναγκαστικές, ελλείψει δεδομένων ανά αποβάθρα και πραγματικού χρόνου· η τρίτη είναι συνειδητή επιλογή.

Δεύτερον, η εγγύηση βελτιστότητας αφορά την πρωιμότερη άφιξη εντός κάθε κλάσης πλήθους σκελών και όχι το σύνολο των εναλλακτικών που παρουσιάζονται. Η λεξικογραφική διάταξη της ενότητας 5.5.6 αποτελεί πρακτική προσέγγιση και όχι υπολογισμό πλήρους συνόλου Pareto.

Τρίτον, οφείλει να επαναληφθεί το ελάττωμα που ανέδειξε η ενότητα 6.8: σε 0,3% των σκελών η αναμονή ίδιας ημέρας υπερβαίνει το δηλωμένο όριο, επειδή το σφίξιμο ενός σκέλους διευρύνει την προηγούμενη αναμονή χωρίς να επανελεγχεται. Πρόκειται για πραγματική, αν και σπάνια, απόκλιση μεταξύ προδιαγραφής και υλοποίησης, η οποία αίρεται με έναν τελικό έλεγχο συμβατότητας μετά τη μετα-βελτιστοποίηση.

7.3.3 Εγκυρότητα και χρονική ισχύς των δεδομένων

Το σύνολο δεδομένων είναι στιγμιότυπο του Ιουνίου 2026 και αποτυπώνει το εβδομαδιαίο πρότυπο μιας συγκεκριμένης περιόδου. Δεν καλύπτει τη θερινή και χειμερινή διαφοροποίηση των

προγραμμάτων, ούτε τις εξαιρέσεις των αργιών και της σχολικής περιόδου, οι οποίες σε πολλές γραμμές μεταβάλλουν δραστικά τη συχνότητα.

Ο περιορισμός επιτείνεται από την ίδια τη συμπεριφορά των πηγών, οι οποίες, όπως επισημάνθηκε στην ενότητα 1.3, δεν δηλώνουν περίοδο εγκυρότητας ούτε τηρούν ιστορικό εκδόσεων: μια αλλαγή αντικαθιστά σιωπηρά την προηγούμενη κατάσταση. Κάθε αριθμητικό αποτέλεσμα των κεφαλαίων 4 και 6 αναφέρεται, συνεπώς, στο συγκεκριμένο στιγμιότυπο και είναι αναπαραγώγιμο μόνο ως προς αυτό — γι' αυτό άλλωστε τα ενδιάμεσα αρχεία JSON διατηρούνται χρονοσημασμένα και η λειτουργία ελέγχου του γράφου δηλώνει πάντοτε ποια έκδοση βρίσκεται σε χρήση. Η ανανέωση, πάντως, δεν απαιτεί καμία μεταβολή του πυρήνα: αρκεί η επανεκτέλεση της αλυσίδας.

7.4 Μελλοντικές επεκτάσεις

7.4.1 Πολυτροπική δρομολόγηση

Η προφανέστερη επέκταση είναι η ενσωμάτωση των υπόλοιπων μέσων. Ο σιδηρόδρομος, η ακτοπλοΐα και οι αεροπορικές συνδέσεις μοιράζονται την ίδια δομή δεδομένων —διακριτές αναχωρήσεις μεταξύ ζευγών σημείων με ημέρες εκτέλεσης— και εντάσσονται στον υπάρχοντα γράφο χωρίς μεταβολή του αλγορίθμου, με μόνη προσθήκη ένα πεδίο τύπου μέσου και διαφοροποιημένους χρόνους ανταπόκρισης, σαφώς μεγαλύτερους στα λιμάνια και στα αεροδρόμια. Κρισιμότερη είναι η σύνδεση με τις αστικές συγκοινωνίες, καθώς μια διαδρομή δεν τελειώνει στον υπεραστικό σταθμό αλλά στον τελικό προορισμό εντός της πόλης· εδώ η ύπαρξη δεδομένων GTFS για τα δύο μεγάλα κέντρα καθιστά την επέκταση άμεσα εφικτή, ενώ η μετάβαση πεζή μεταξύ γειτονικών στάσεων θα απαιτούσε ρητές ακμές μεταφοράς.

7.4.2 Διασυνοριακή επέκταση και ενοποίηση με ευρωπαϊκά δίκτυα

Οι υπεραστικές συνδέσεις προς τη Βουλγαρία, τη Βόρεια Μακεδονία, την Αλβανία και την Τουρκία, καθώς και τα πανευρωπαϊκά δίκτυα λεωφορείων, δεν αποτυπώνονται στο παρόν σύνολο δεδομένων. Η ενσωμάτωσή τους έχει το πλεονέκτημα ότι οι αντίστοιχοι φορείς δημοσιεύουν συχνά δεδομένα σε τυποποιημένη μορφή, οπότε το βάρος μετατοπίζεται από την ανάκτηση στην αντιστοίχιση των συνοριακών κόμβων. Θα απαιτούσε, ωστόσο, ρητή διαχείριση των ζωνών ώρας και της διάκρισης μεταξύ προγραμματισμένου χρόνου και χρόνου διέλευσης των συνόρων, ο οποίος είναι εγγενώς μεταβλητός και θα λειτουργούσε στην πράξη ως πρόσθετος, μη ντετερμινιστικός χρόνος ανταπόκρισης στους συνοριακούς κόμβους.

7.4.3 Δρομολόγηση υπό αβεβαιότητα

Το παρόν μοντέλο θεωρεί τους χρόνους ντετερμινιστικούς. Μια στοχαστική επέκταση θα αντιμετώπιζε τη διάρκεια κάθε σκέλους ως τυχαία μεταβλητή και θα ζητούσε όχι την πρωιμότερη άφιξη αλλά τη διαδρομή με τη μέγιστη πιθανότητα εμπρόθεσμης άφιξης, ή εκείνη που μεγιστοποιεί την πιθανότητα επίτευξης κάθε ανταπόκρισης. Η διάκριση έχει άμεσο πρακτικό νόημα σε αραιό δίκτυο: μια χαμένη ανταπόκριση με επόμενη αναχώρηση την επομένη έχει ασύγκριτα μεγαλύτερο

κόστος από μια καθυστέρηση εικοσαλέπτου, και μια διαδρομή με περιθώριο σαράντα λεπτών μπορεί να είναι προτιμότερη από μια ταχύτερη με περιθώριο δεκαπέντε.

7.4.4 Ενσωμάτωση δεδομένων πραγματικού χρόνου και δυναμική επαναδρομολόγηση

Τρεις από τις τέσσερις πλατφόρμες περιλαμβάνουν ήδη υποδομή τηλεματικής και άρα διαθέτουν εσωτερικά τη θέση των οχημάτων. Η έκθεση των δεδομένων αυτών θα επέτρεπε την τροφοδότηση του συστήματος με πληροφορία καθυστερήσεων και ματαιώσεων και, κατ' επέκταση, τη δυναμική επαναδρομολόγηση: την ειδοποίηση του επιβάτη ότι η ανταπόκρισή του κινδυνεύει και τον άμεσο υπολογισμό εναλλακτικής από τον τρέχοντα κόμβο. Ο υφιστάμενος πυρήνας υποστηρίζει ήδη τέτοιο ερώτημα, αφού η αναζήτηση από ενδιάμεση στάση με δεδομένη ώρα δεν διαφέρει σε τίποτε από μια νέα αναζήτηση· το ζητούμενο δεν είναι, δηλαδή, αλγοριθμικό αλλά και εδώ ζήτημα διάθεσης δεδομένων που ήδη υπάρχουν.

7.4.5 Εξατομικευμένη και επεξηγήσιμη δρομολόγηση

Η λεξικογραφική διάταξη ενσωματώνει μια συγκεκριμένη ιεραρχία προτιμήσεων, η οποία δεν είναι κοινή σε όλους. Μια εξατομικευμένη εκδοχή θα επέτρεπε στον χρήστη να δηλώσει τα δικά του βάρη — ανοχή σε αναμονή, αποστροφή προς μετεπιβιώσεις, προτίμηση πρωινής ή απογευματινής αναχώρησης — ή θα τα συνήγαγε από το ιστορικό των επιλογών του. Συμπληρωματικά, η επεξηγήσιμη δρομολόγηση θα αιτιολογούσε τόσο τις προτάσεις όσο και τις αρνήσεις, δηλώνοντας ποιος ακριβώς περιορισμός απέκλεισε μια εναλλακτική· η διαγνωστική συμπεριφορά της γραμμής εντολών, που περιγράφηκε στην ενότητα 5.5.9, αποτελεί ήδη ένα πρώτο βήμα προς την κατεύθυνση αυτή.

7.4.6 Από τον σχεδιασμό στον ανασχεδιασμό του δικτύου: ανάλυση προσβασιμότητας ως εργαλείο πολιτικής

Το σύνολο δεδομένων επιτρέπει ερωτήματα που υπερβαίνουν τον μεμονωμένο επιβάτη. Ο υπολογισμός, για κάθε οικισμό, του πλήθους των προορισμών που είναι προσβάσιμοι εντός δεδομένου χρόνου παράγει έναν χάρτη προσβασιμότητας ολόκληρης της χώρας, ενώ η επανάληψη ανά ημέρα αναδεικνύει τις περιοχές που απομονώνονται τα σαββατοκύριακα. Η ίδια ανάλυση εντοπίζει τις ελλείπουσες ακμές: ζεύγη περιοχών με σημαντική ζήτηση των οποίων η σύνδεση απαιτεί δυσανάλογο χρόνο ή διανυκτέρευση, και όπου συνεπώς η προσθήκη ενός δρομολογίου ή απλώς ο συγχρονισμός δύο υφιστάμενων θα είχε δυσανάλογο όφελος. Το εργαλείο σχεδιασμού ταξιδιού μετατρέπεται έτσι σε εργαλείο τεκμηρίωσης συγκοινωνιακής πολιτικής.

7.4.7 Προς ένα εθνικό πρότυπο ανοικτών συγκοινωνιακών δεδομένων

Η ουσιαστικότερη συνέχεια, ωστόσο, δεν είναι τεχνική. Όσο η πληροφορία ανακτάται με αντίστροφη μηχανίκευση, κάθε τέτοιο σύστημα θα παραμένει εύθραυστο και θα αναπαράγει την ίδια εργασία. Το άμεσο βήμα είναι η εξαγωγή του παρόντος συνόλου σε μορφή GTFS, όπως προαναγγέλθηκε στην ενότητα 3.8, ώστε να καταστεί άμεσα καταναλώσιμο από υφιστάμενα

εργαλεία· η κύρια εκκρεμότητα εκεί είναι η ανασύσταση των ενδιάμεσων στάσεων κάθε γραμμής. Το ουσιώδες βήμα, όμως, ανήκει στους φορείς και στη ρυθμιστική αρχή: η υιοθέτηση κοινού προτύπου δημοσίευσης, με σαφή αδειοδότηση, δηλωμένη περίοδο εγκυρότητας και σταθερά αναγνωριστικά στάσεων, όπως άλλωστε προδιαγράφει και το ευρωπαϊκό πλαίσιο της ενότητας 1.1. Η παρούσα εργασία δείχνει τι γίνεται εφικτό όταν τα δεδομένα ενοποιηθούν· ένα εθνικό πρότυπο θα καθιστούσε το αποτέλεσμα διαρκές αντί για στιγμιαίο.

Επίλογος κεφαλαίου

Η εργασία επιχείρησε να απαντήσει σε ένα ερώτημα που φαινομενικά έχει ήδη λυθεί παντού αλλού και το οποίο, εντούτοις, παρέμενε αναπάντητο για το ελληνικό υπεραστικό δίκτυο. Η απάντηση απαίτησε πολύ περισσότερη εργασία στα δεδομένα παρά στους αλγορίθμους, και ακριβώς αυτή η δυσαναλογία αποτελεί το κεντρικό της εύρημα: το δίκτυο συνδέει τη χώρα σαφώς περισσότερο απ' όσο φαίνεται, και η αοριστία γύρω από το τι είναι εφικτό δεν είναι ιδιότητα του δικτύου αλλά της πληροφόρησης γι' αυτό. Ο ΜΙΤΟΣ δεν φιλοδοξεί να υποκαταστήσει μια θεσμική λύση· αποδεικνύει, με συγκεκριμένο και μετρήσιμο τρόπο, ότι η απουσία της δεν είναι αναπόφευκτη.

Βιβλιογραφία

- [1] Ευρωπαϊκή Επιτροπή. Κατ' εξουσιοδότηση Κανονισμός (ΕΕ) 2017/1926 της Επιτροπής, της 31ης Μαΐου 2017, για τη συμπλήρωση της οδηγίας 2010/40/ΕΕ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου όσον αφορά την παροχή σε επίπεδο ΕΕ υπηρεσιών πληροφόρησης για τις πολυτροπικές μετακινήσεις. Επίσημη Εφημερίδα της Ευρωπαϊκής Ένωσης, L 272, 21.10.2017.
- [2] Νόμος 2963/2001 «Οργάνωση και λειτουργία των δημόσιων επιβατικών μεταφορών με λεωφορεία, τεχνικός έλεγχος οχημάτων και ασφάλεια χειρσαίων μεταφορών και άλλες διατάξεις». ΦΕΚ Α' 268/23.11.2001.
- [3] Υπουργική απόφαση Β/οικ.69088/4971/15/12.11.2015 «Καθορισμός των κομίστρων των αστικών και υπεραστικών ΚΤΕΛ και ΚΤΕΛ Α.Ε. της χώρας». ΦΕΚ Β' 2498/19.11.2015, όπως τροποποιήθηκε τελευταία με την υπ' αρ. 37279 απόφαση του Αναπληρωτή Υπουργού Υποδομών και Μεταφορών, ΦΕΚ Β' 1258/06.03.2026.
- [4] Οργανισμός Αστικών Συγκοινωνιών Αθηνών (Συγκοινωνίες Αθηνών). Δρομολόγια Οδικών Συγκοινωνιών (Λεωφορεία, Τρόλεϊ) — ανοικτά δεδομένα σε μορφή GTFS. <https://data.gov.gr/dataset/dromologia-odikon-sygkoinonion-leoforeia-trolei>
- [5] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.
- [6] M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM*, 34(3):596–615, 1987.
- [7] A. Orda and R. Rom. Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length. *Journal of the ACM*, 37(3):607–625, 1990.
- [8] E. Pyrga, F. Schulz, D. Wagner and C. Zaroliagis. Efficient models for timetable information in public transportation systems. *ACM Journal of Experimental Algorithmics*, 12, Article 2.4, 1–39, 2008.
- [9] J. Dibbelt, T. Pajor, B. Strasser and D. Wagner. Connection Scan Algorithm. *ACM Journal of Experimental Algorithmics*, 23, Article 1.7, 1–56, 2018.
- [10] D. Delling, T. Pajor and R. F. Werneck. Round-Based Public Transit Routing. *Transportation Science*, 49(3):591–604, 2015.
- [11] H. Bast, E. Carlsson, A. Eigenwillig, R. Geisberger, C. Harrelson, V. Raychev and F. Viger. Fast Routing in Very Large Public Transportation Networks using Transfer Patterns. In *Proceedings of the 18th Annual European Symposium on Algorithms (ESA 2010)*, LNCS 6346, pages 290–301. Springer, 2010.

- [12] A. K. Elmagarmid, P. G. Ipeirotis and V. S. Verykios. Duplicate Record Detection: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):1–16, 2007.
- [13] H. Bast, D. Delling, A. Goldberg, M. Müller-Hannemann, T. Pajor, P. Sanders, D. Wagner and R. F. Werneck. Route Planning in Transportation Networks. In *Algorithm Engineering*, LNCS 9220, pages 19–80. Springer, 2016.
- [14] M. Müller-Hannemann and M. Schnee. Finding All Attractive Train Connections by Multi-Criteria Pareto Search. In *Algorithmic Methods for Railway Optimization*, LNCS 4359, pages 246–263. Springer, 2007.
- [15] Y. Disser, M. Müller-Hannemann and M. Schnee. Multi-Criteria Shortest Paths in Time-Dependent Train Networks. In *Proceedings of the 7th International Workshop on Experimental Algorithms (WEA 2008)*, LNCS 5038, pages 347–361. Springer, 2008.
- [16] General Transit Feed Specification (GTFS) Schedule Reference. <https://gtfs.org/documentation/schedule/reference/>
- [17] CEN. CEN/TS 16614, Public transport — Network and Timetable Exchange (NeTEx). European Committee for Standardization.
- [18] Requests: HTTP for Humans — Documentation. <https://requests.readthedocs.io>
- [19] Beautiful Soup Documentation. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- [20] Python Software Foundation. asyncio — Asynchronous I/O. <https://docs.python.org/3/library/asyncio.html>
- [21] SQLite Documentation. <https://www.sqlite.org/docs.html>
- [22] Python Software Foundation. pickle — Python object serialization. <https://docs.python.org/3/library/pickle.html>
- [23] Flask Documentation. <https://flask.palletsprojects.com>
- [24] R. T. Fielding. Architectural Styles and the Design of Network-based Software Architectures. Διδακτορική διατριβή, University of California, Irvine, 2000.
- [25] OpenStreetMap Foundation. Copyright and License (Open Database License). <https://www.openstreetmap.org/copyright>
- [26] OpenStreetMap Foundation. Nominatim Usage Policy. <https://operations.osmfoundation.org/policies/nominatim/>
- [27] D. Luxen and C. Vetter. Real-time routing with OpenStreetMap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '11)*, pages 513–516. ACM, 2011.
- [28] Leaflet — an open-source JavaScript library for mobile-friendly interactive maps. <https://leafletjs.com>
- [29] PyCryptodome Documentation. <https://pycryptodome.readthedocs.io>