



UNIVERSITY OF THESSALY

MASTER THESIS

---

# CROWDSOURCING TECHNOLOGIES AND APPLICATIONS

---

*Author:*

DIMITRIOS K. ARCHONTIS

*Supervisor:*

PANAYIOTIS BOZANIS

*Associate Professor*

*This Master Thesis, submitted in fulfilment of the requirements  
for a Master of Science Degree*

*in the*

COMPUTER SCIENCE AND ENGINEERING, TELECOMMUNICATIONS AND NETWORKS

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING



VOLOS, OCTOBER 2013





ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

---

## ΠΛΗΘΟΠΟΡΙΣΜΟΣ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

---

Συντάκτης:

ΔΗΜΗΤΡΙΟΣ Κ. ΑΡΧΟΝΤΗΣ

Επιβλέπων:

ΠΑΝΑΓΙΩΤΗΣ ΜΠΟΖΑΝΗΣ

Αναπληρωτής Καθηγητής

*Μεταπτυχιακή Διατριβή, υποβληθείσα ως μέρος των απαιτήσεων  
για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης*

*στην*

ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ, ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΔΙΚΤΥΩΝ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ



ΒΟΛΟΣ, ΟΚΤΩΒΡΙΟΣ 2013



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Πολυτεχνική Σχολή

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Μ.Δ.Ε. “Επιστήμη και Τεχνολογία Υπολογιστών, Τηλεπικοινωνιών και Δικτύων”

# Πληθοπορισμός Τεχνολογίες και Εφαρμογές

## ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΔΗΜΗΤΡΙΟΣ Κ. ΑΡΧΟΝΤΗΣ

Επιβλέπων: **ΜΠΟΖΑΝΗΣ ΠΑΝΑΓΙΩΤΗΣ**  
*Αναπληρωτής Καθηγητής Π.Θ.*

(Υπογραφή)

---

**ΜΠΟΖΑΝΗΣ ΠΑΝΑΓΙΩΤΗΣ**  
*Αν. Καθηγητής Π.Θ.*

(Υπογραφή)

---

**ΤΣΟΜΠΑΝΟΠΟΥΛΟΥ ΠΑΝΑΓΙΩΤΑ**  
*Επ. Καθηγήτρια Π.Θ.*

(Υπογραφή)

---

**ΚΑΤΣΑΡΟΣ ΔΗΜΗΤΡΙΟΣ**  
*Λέκτορας Π.Θ.*

(Υπογραφή)

---

ΔΗΜΗΤΡΙΟΣ Κ. ΑΡΧΟΝΤΗΣ

©2013 – ALL RIGHTS RESERVED.

COPYRIGHT © ΔΗΜΗΤΡΙΟΣ Κ. ΑΡΧΟΝΤΗΣ, 2013

ΜΕ ΕΠΙΦΥΛΑΞΗ ΠΑΝΤΟΣ ΔΙΚΑΙΩΜΑΤΟΣ. ALL RIGHTS RESERVED

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

Η ΣΕΛΙΔΑ ΑΥΤΗ ΕΙΝΑΙ ΣΚΟΠΙΜΑ ΚΕΝΗ



# Ευχαριστίες

*Με την παρούσα εργασία περατώνονται οι σπουδές μου στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας.*

*Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Μποζάνη Παναγιώτη για την εμπιστοσύνη και το ενδιαφέρον που έδειξε τόσο κατά την ανάθεση της εργασίας, όσο και κατά τη διάρκεια διεξαγωγής των μαθημάτων. Θα ήθελα επίσης να τον ευχαριστήσω για τη συμπαράσταση και την καθοδήγησή του καθ' όλη τη διάρκεια εκπόνησης της εργασίας μου.*

*Στη συνέχεια θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα κ. Φεύγα Αθανάσιο ο οποίος με τα πλούσια πνευματικά προσόντα και το ήθος του συνέβαλε ουσιαστικά στην ολοκλήρωση αυτής της εργασίας.*

*Πάνω απ' όλα, είμαι ευγνώμων στους γονείς μου, για την ολόψυχη αγάπη και υποστήριξή τους όλα αυτά τα χρόνια. Αφιερώνω αυτή την εργασία στην μητέρα μου και στον πατέρα μου.*

ΔΗΜΗΤΡΙΟΣ Κ. ΑΡΧΟΝΤΗΣ



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

# Περίληψη

Πολυτεχνική Σχολή

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Μ.Δ.Ε. “Επιστήμη και Τεχνολογία Υπολογιστών, Τηλεπικοινωνιών και Δικτύων”

Μεταπτυχιακή Διατριβή

**Πληθοπορισμός**

**Τεχνολογίες και Εφαρμογές**

ΔΗΜΗΤΡΙΟΣ Κ. ΑΡΧΟΝΤΗΣ

Την σημερινή εποχή η πλειονότητα του πληθυσμού έχει στην κατοχή της ένα ή περισσότερα κινητά τηλέφωνα. Οι κινητές συσκευές διαθέτουν εξελιγμένους αισθητήρες όπως είναι το GPS, η κάμερα, το μικρόφωνο ή η ψηφιακή πυξίδα. Ο συνδυασμός της ανάπτυξης εφαρμογών και των τεχνολογικών υποδομών, παρέχει την δυνατότητα στα κινητά τηλέφωνα να λειτουργούν ως τοπικοί σταθμοί βάσης. Οποιοσδήποτε χρήστης, κάνοντας χρήση των εφαρμογών που είναι εγκατεστημένες στις συσκευές, μπορεί να μοιραστεί δεδομένα με άλλους χρήστες. Ο πληθοπορισμός είναι ο νέος ορισμός που δίνεται, για να προσδιοριστεί η ανοιχτή πρόσκληση σε ένα πλήθος ατόμων, ώστε να επιλύσουν προβλήματα και να προτείνουν λύσεις προς βελτίωση των συνθηκών που επικρατούν στην κοινωνία. Το πλήθος έχει ενεργή συμμετοχή στις εφαρμογές, που η κάθε μια προσφέρει ανωνυμία και ασφάλεια στον κάθε χρήστη. Στην παρούσα εργασία, θα παρουσιάσουμε τεχνολογίες και εφαρμογές που αναπτύχθηκαν, για την συλλογή δεδομένων από τους χρήστες.

**Λέξεις κλειδιά:** *πληθοπορισμός, κινητά τηλέφωνα, τοπικός σταθμός βάσης, κυκλοφοριακό σύστημα, Παγκόσμιο Σύστημα Θεσιθεσίας*

UNIVERSITY OF THESSALY

# *Abstract*

Faculty of Engineering

Department of Electrical and Computer Engineering

M.Sc. "Computer Science and Engineering, Telecommunications and Networks"

Master Thesis

**Crowdsourcing**

**Technologies and Applications**

DIMITRIOS K. ARCHONTIS

The present year's the majority of the population owns one or several mobile phones. The mobile devices includes advanced blunder detection sensors such as GPS, camera, microphone or digital compass. The combination of application development and technology infrastructure give the opportunity to mobile phones to operate as a local base stations. Any user making use of the applications installed on the device can share data with other users. Crowdsourcing is the new definition given to determine the open invitation to a number of individuals to solve problems and suggest solutions to improve conditions in society. The crowd participates actively, where each one application offers anonymity and security to each user. In this thesis, we will present technologies and applications that were developed to collect data from users.

**Keywords:** *crowdsourcing, smartphones, Location-Based Services, Participatory Sensing, Traffic Conditions, Global Positioning System*

# Περιεχόμενα

Ευχαριστίες	vi
Περίληψη	viii
Abstract	ix
Κατάλογος Σχημάτων	xiii
Συντομογραφίες	xiv
<b>I ΘΕΩΡΗΤΙΚΕΣ ΕΝΝΟΙΕΣ</b>	<b>1</b>
<b>1 Εισαγωγή</b>	<b>2</b>
1.1 Γενική Επισκόπηση	2
1.2 Αντικείμενο της παρούσας διατριβής	4
1.3 Οργάνωση του τόμου	6
<b>2 Ορισμός και μοντέλο του Crowdsourcing</b>	<b>7</b>
2.1 Εισαγωγή	7
2.2 Παράγοντες εξασφάλισης επιτυχούς crowdsourcing μοντέλου	8
2.3 Χρησιμότητα των smartphones στο crowdsourcing μοντέλο	11
<b>3 Spatial Crowdsourcing</b>	<b>13</b>
3.1 Εισαγωγή	13
3.2 Κατηγοριοποίηση του Spatial Crowdsourcing	13
3.3 Περιγραφή Ανάθεσης των Εργασιών	15
3.4 Μέθοδοι Ανάθεσης των Εργασιών	17
3.4.1 Στρατηγική Greedy	17
3.4.2 Στρατηγική Least Location Entropy Priority (LLEP)	19
3.5 Επίλογος	20
<b>4 Η εξέλιξη και οι δυνατότητες των κινητών τηλεφώνων</b>	<b>21</b>
4.1 Εισαγωγή	21
4.2 Εξέλιξη των Κινητών Τηλεφώνων	21

4.3	Πεδία και παροχή εφαρμογών	23
4.4	Γενική Αρχιτεκτονική των Εφαρμογών	26
4.4.1	Συμμετέχοντες	26
4.4.2	Κύριες λειτουργίες των υποδομών των εφαρμογών	28
4.5	Location-Based Υπηρεσίες	29
4.5.1	Αρχιτεκτονική των Location-Based Υπηρεσιών	30
4.5.2	Προσδιορισμός της τοποθεσίας των χρηστών	31
4.6	Επίλογος	31
<b>5</b>	<b>Ασφάλεια προσωπικών δεδομένων</b>	<b>33</b>
5.1	Εισαγωγή	33
5.2	Δομή της Αρχιτεκτονικής για Επίτευξη Ανωθυμίας	34
5.3	Πληροφορίες Θέσεως μέσω της έννοιας $k$ -anonymity	35
5.3.1	Αλγόριθμος Adaptive-Interval	35
5.3.2	Μέθοδος Casper	36
5.4	Η μετρική των mix zones	37
5.5	Επίλογος	39
<b>II</b>	<b>ΕΦΑΡΜΟΓΕΣ</b>	<b>40</b>
<b>6</b>	<b>VTrack Project</b>	<b>41</b>
6.1	Εισαγωγή	41
6.2	Αρχιτεκτονική του Συστήματος VTrack	41
6.3	Βασικοί Στόχοι του VTrack Project	43
6.4	Αλγόριθμοι Υλοποίησης	44
6.4.1	Hidden Markov Model και Μονοπάτι Viterbi	44
6.4.2	Χαρτογράφηση Διαδρομών	45
6.5	Εκτίμηση της Χρονικής Διάρκειας της Διαδρομής	47
6.6	Αξιολόγηση του VTrack project	48
6.6.1	Λογικές Τιμές Δεδομένων και Πληροφοριών	48
6.6.2	Χαρτογράφηση των Διαδρομών	49
6.6.3	Ανίχνευση Ενεργών Σημείων	51
6.7	Επίλογος	53
<b>7</b>	<b>Mobile Millennium Project</b>	<b>54</b>
7.1	Εισαγωγή	54
7.2	Το σύστημα του Mobile Millennium Project	54
7.2.1	Δομή της Αρχιτεκτονικής	55
7.2.2	Λειτουργία του Συστήματος	55
7.2.3	Virtual Trip Line (VTL)	57
7.3	Αλγόριθμος Εκτίμησης της Κυκλοφορίας	57
7.4	Επίλογος	60
<b>8</b>	<b>Crowdsourcing εφαρμογές των τεσσάρων κύριων τομέων</b>	<b>61</b>
8.1	Εισαγωγή	61
8.2	Εφαρμογές Παρακολούθησης της Υγείας	61
8.3	Εφαρμογές Παρακολούθησης του Περιβάλλοντος	62

8.4	Εφαρμογές Κοινωνικής Δικτύωσης . . . . .	63
8.5	Εφαρμογές Παρακολούθησης της Οδικής Κυκλοφορίας . . . . .	64
<b>III</b>	<b>ΕΠΙΛΟΓΟΣ</b>	<b>65</b>
<b>9</b>	<b>Επίλογος και Μελλοντικές Έρευνες</b>	<b>66</b>
9.1	Επίλογος . . . . .	66
9.2	Μελλοντικές έρευνες . . . . .	66
<b>IV</b>	<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>	<b>68</b>
	<b>Βιβλιογραφία</b>	<b>69</b>

# Κατάλογος σχημάτων

1.1	Γενικό σύστημα σχεδιασμού των εφαρμογών . . . . .	5
2.1	Παράγοντες διασφάλισης επιτυχούς crowdsourcing μοντέλου . . . . .	9
2.2	Επίδραση περιφερειακών παραγόντων στους καθοριστικούς παράγοντες . . . .	11
2.3	Διαδικασία του crowdsourcing . . . . .	11
3.1	Κατηγοριοποίηση του spatial crowdsourcing . . . . .	14
3.2	Πλαίσιο ανάθεσης εργασιών του spatial crowdsourcing . . . . .	15
3.3	Παράδειγμα μετατροπής προβλήματος μέγιστης ανάθεσης εργασιών σε πρό- βλημα εύρεσης μέγιστης ροής του δικτύου . . . . .	18
4.1	Αισθητήρες που είναι εγκατεστημένα στο κινητό iPhone 4 της Apple . . . . .	22
4.2	Διαχωρισμός εφαρμογών ανά κατηγορίες . . . . .	24
4.3	Δομή της αρχιτεκτονικής μιας τυπικής crowdsourcing εφαρμογής . . . . .	26
4.4	Απεικόνιση συμμετεχόντων και βασικών λειτουργιών των υποδομών μιας crowdsourcing εφαρμογής . . . . .	27
4.5	Σύγκλιση τεχνολογιών για την υποστήριξη Location-Based Υπηρεσιών . . . . .	30
4.6	Δομή της αρχιτεκτονικής των location-based υπηρεσιών . . . . .	30
5.1	Δομή της αρχιτεκτονικής των συστημάτων για επίτευξη ανωνυμίας . . . . .	34
5.2	Adaptive–Interval αλγόριθμος . . . . .	36
5.3	Ανωνυμία με την μέθοδο Casper . . . . .	37
5.4	Παράδειγμα εφαρμογής της mix zones . . . . .	38
6.1	Η εφαρμογή VTrack απεικονίζει σημεία κυκλοφοριακής συμφόρησης . . . . .	42
6.2	Αρχιτεκτονική του VTrack project . . . . .	42
6.3	Λειτουργία του server του VTrack . . . . .	43
6.4	Παράδειγμα Hidden Markov Model . . . . .	45
6.5	Η διαδικασία χαρτογράφησης διαδρομών . . . . .	45
6.6	Αθροιστικές Συναρτήσεις Κατανομής . . . . .	51
6.7	Ποσοστά επιτυχούς ανίχνευσης Hotspot . . . . .	52
6.8	Ποσοστά λανθασμένων αποτελεσμάτων ανίχνευσης Hotspot . . . . .	52
7.1	Αρχιτεκτονική του mobile millennium συστήματος . . . . .	55
7.2	Διαδικασία που ακολουθείται για την εκτίμηση της κυκλοφορίας . . . . .	56
7.3	Απεικόνιση της εξόδου του προσομοιωτή της εκτίμησης της κυκλοφορίας . . . .	57
7.4	Απεικόνιση των Virtual Trip Lines ως μέρος του mobile millennium συστήματος	58
7.5	Διάγραμμα ροής της λειτουργίας του EM αλγορίθμου . . . . .	59



# Συντομογραφίες

<b>3G</b>	<b>3rd Generation</b>
<b>AP</b>	<b>Access Point</b>
<b>AS</b>	<b>Anonymizer Server</b>
<b>ASR</b>	<b>Anonymizing Spatial Region</b>
<b>CDFs</b>	<b>Cumulative Distribution Functions</b>
<b>GIS</b>	<b>Geographical Information System</b>
<b>GPRS</b>	<b>General Packet Radio Service</b>
<b>GPS</b>	<b>Global Positioning System</b>
<b>GSM</b>	<b>Global System for Mobile communications</b>
<b>HMM</b>	<b>Hidden Markov Model</b>
<b>LAN</b>	<b>Local Area Network</b>
<b>LBS</b>	<b>Location Based Service</b>
<b>QoS</b>	<b>Quality of Service</b>
<b>SMS</b>	<b>Short Message Service</b>
<b>TCP</b>	<b>Transmission Control Protocol</b>
<b>VTL</b>	<b>Virtual Trip Line</b>
<b>WiFi</b>	<b>Wireless Fidelity</b>

*αφιερωμένη στους γονείς μου...*

**Μέρος Ι**

**ΘΕΩΡΗΤΙΚΕΣ ΕΝΝΟΙΕΣ**

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Γενική Επισκόπηση

Την τελευταία δεκαετία οι υπηρεσίες που προσφέρονται μέσω του διαδικτύου, παρέχουν τη δυνατότητα συλλογής και αποθήκευσης πληθώρα όγκου δεδομένων και πληροφοριών. Η βελτίωση των υπάρχοντων συσκευών για την επεξεργασία, την ανάλυση και την εξαγωγή συμπερασμάτων από τα συλλεγόμενα δεδομένα, αποτελεί επιτακτική ανάγκη. Η εξέλιξη της τεχνολογίας έχει επιτρέψει την κατασκευή μικρότερων, γρηγορότερων και πιο αξιόπιστων συσκευών. Το αποτέλεσμα που προκύπτει είναι η μετάβαση, από τους παραδοσιακούς σταθμούς εργασίας τύπου desktop ή τα παραδοσιακά κινητά τηλέφωνα, σε συσκευές τύπου tablets και smartphones.

Ο ταχύς ρυθμός της τεχνολογικής ανάπτυξης της ψηφιακής τηλεπικοινωνίας, σε συνδυασμό με την ανάπτυξη της διαδικτυακής τεχνολογίας και την εξέλιξη των smartphones, δίνει την δυνατότητα να παρέχονται δεδομένα και πληροφορίες που σχετίζονται με τη θέση της τοποθεσίας των φυσικών προσώπων. Τα σημερινά κινητά εκτός του ότι εξελίσσονται ως ο κεντρικός υπολογιστής και η συσκευή επικοινωνίας μεταξύ των ανθρώπων, έχουν και ένα πλούσιο σύνολο αισθητήρων. Η ύπαρξη του επιταχυνσιόμετρου, της ψηφιακής πυξίδας, του γυροσκοπίου, του παγκοσμίου συστήματος θεσιθεσίας, του μικροφώνου και της κάμερας, σε συνδυασμό με την βελτίωση της επεξεργαστικής ισχύς και τον αυξημένο χώρο αποθήκευσης δεδομένων, τα καθιστά ως μια συσκευή πολύτιμη. *Αξίζει να σημειωθεί ότι πάνω από 5 δισεκατομμύρια άνθρωποι παγκοσμίως έχουν στην κατοχή τους κινητά τηλέφωνα.* Δεδομένα άνευ προηγουμένου ποσότητας και ποιότητας μπορούν να καταγραφούν και να υποστούν επεξεργασία, από μια πιθανή βάση δισεκατομμυρίων χρηστών κινητών τηλεφώνων, παγκοσμίως [27, 34].

Η παρουσία όλων των προαναφερθέντων αισθητήρων στα κινητά τηλέφωνα επιτρέπει ένα ευρύ φάσμα καινοτόμων νέων εφαρμογών. Συνολικά, οι αισθητήρες αυτοί είναι που επιτρέπουν την ανάπτυξη νέων εφαρμογών, σε μια ευρεία κλίμακα διαδεδομένων τομέων, όπως είναι η υγειονομική περίθαλψη, τα κοινωνικά δίκτυα, η παρακολούθηση του περιβάλλοντος

[29], καθώς και των μεταφορών [3, 6, 15, 30, 37]. Επιπρόσθετα, τα κινητά τηλέφωνα λειτουργούν και σαν τοπικοί σταθμοί βάσης, βασιζόμενα στους αισθητήρες εύρεσης της γεωγραφικής τοποθεσίας [8, 34].

Κάθε υπηρεσία που αφορά δεδομένα τοποθεσίας καθορίζει την θέση του χρήστη για να παρέχει πληροφορίες. Αυτές οι πληροφορίες θέσεως μπορούν να χρησιμοποιηθούν μια φορά και για έναν μόνο σκοπό, ή μπορεί να αποθηκευτούν, ή να συνδυαστούν με άλλες πληροφορίες για να παράγουν ένα ιστορικό των δραστηριοτήτων του χρήστη, ή για άλλους σκοπούς. Οι Location-Based εφαρμογές (LBS), παρέχουν ένα ευρύ φάσμα υπηρεσιών που δύναται να εκτελεστούν από πολλές διαφορετικές συσκευές. Για παράδειγμα, εργαλεία πλοήγησης που παρέχουν οδηγίες μετάβασης και πληροφορίες σχετικά με την κυκλοφορία στους οδικούς άξονες σε πραγματικό χρόνο [3, 37], όπως επίσης και εφαρμογές κοινωνικής δικτύωσης που ενημερώνουν τους χρήστες πότε οι γνωστοί τους βρίσκονται σε κοντινή, με αυτούς, απόσταση. (όπως το foursquare [13, 38])[34].

*Η δημοτικότητα των LBS είναι ταχέως αναπτυσσόμενη, λόγω της αυξανόμενης χρήσης των location-enabled συσκευών, τύπου smartphones και tablets.*

Οι LBS μπορούν να συλλέγουν και να επεξεργάζονται τεράστιο όγκο πληροφοριών και δεδομένων που αφορούν τους χρήστες, μέσω ενός ευρέως φάσματος εφαρμογών. Οι μηχανές αναζήτησης, είναι δυνατόν να συνδυάσουν πληροφορίες γεωγραφικής τοποθεσίας, με όρους αναζήτησης που έχουν εισαχθεί ή με αποτελέσματα που έχουν επιλεγεί. Τα εργαλεία πλοήγησης, θα μπορούσαν καθορίζοντας την ταχύτητα, να κάνουν εκτίμηση της κυκλοφορίας [34]. Τα δεδομένα όμως των τοπικών σταθμών βάσης συχνά μεταβάλλονται. Αυτή η μεταβολή εισάγει νέες προκλήσεις στην κατηγοριοποίηση και την κατάταξη της εκάστοτε θέσης, που βρίσκεται ο χρήστης. Ο τεράστιος όγκος πληροφοριών που συλλέγεται, δύναται να υποστεί επεξεργασία για εξαγωγή λύσεων και ιδεών, από το πλήθος. Μια εφαρμογή, που η λειτουργία της πραγματοποιείται μέσω του πλήθους, επιστρέφει αποτελέσματα σε σημαντικά χαμηλότερο κόστος.

Ο **πληθοπορισμός** είναι ο νέος ορισμός που δίνεται, για να προσδιοριστεί η ανοιχτή πρόσκληση σε ένα πλήθος ατόμων, ώστε να επιλύσουν προβλήματα, ύστερα από συλλογή δεδομένων από τους τοπικούς σταθμούς βάσης, και με τις κατάλληλες εφαρμογές να προτείνουν λύσεις προς βελτίωση των συνθηκών που επικρατούν στην κοινωνία. Το πλήθος που έχει υπό την κατοχή του smartphone ή tablet, εφόσον λάβει τις κατάλληλες εφαρμογές, μπορεί ομαδικά να εντοπίσει λύσεις, να εξάγει ιδέες, ή συμπεράσματα από τα συλλεγόμενα δεδομένα. Ο **πληθοπορισμός** είναι διαφορετικός από μια συνηθισμένη εξωτερική ανάθεση εργασιών, δεδομένου ότι μια εργασία ή ένα πρόβλημα ανατίθεται εξωτερικά σε ένα απροσδιόριστο πλήθος και όχι σε μια συγκεκριμένη ομάδα [19, 38].

*Με την γενικευμένη παρουσία των κινητών συσκευών, ο πληθοπορισμός αναδύεται σαν*

την νέα πλατφόρμα, που επιτρέπει οι εργασίες να ανατίθενται και να εκτελούνται από το πλήθος.

Οι εφαρμογές που αναπτύσσονται, έχουν ως στόχο τη δημιουργία ενός ευρέως φάσματος καινοτόμων *sensing* εφαρμογών, που μπορούν να κατηγοριοποιηθούν είτε ως **ανθρωπο-κεντρικές**, είτε ως **περιβαλλοντο-κεντρικές**. Οι **ανθρωπο-κεντρικές** εφαρμογές, επικεντρώνονται κυρίως στην κατανόηση της συμπεριφοράς των χρηστών. Δηλαδή χρησιμοποιούν τους αισθητήρες, που βρίσκονται στα τηλέφωνα, για να συλλέξουν δεδομένα που αφορούν τον χρήστη. Αντίθετα, οι **περιβαλλοντο-κεντρικές** εφαρμογές, συλλέγουν περιβαλλοντικές παραμέτρους. Τα τηλέφωνα παρέχουν πληροφορίες μέσω των αισθητήρων τους και πιθανών συμπληρωματικών περιφερειακών αισθητήρων, για την περιοχή που βρίσκονται οι χρήστες [10]. Περισσότερες λεπτομέρειες σχετικά με κατηγοριοποίηση των εφαρμογών, παρέχονται στην ενότητα 4.3.

Καθώς αυξάνονται οι εφαρμογές συλλογής δεδομένων, παρόμοια είναι και η ανάγκη αύξησης της προστασίας των προσωπικών δεδομένων των χρηστών. Η προστασία των προσωπικών δεδομένων είναι σημαντική τόσο σε ατομικό όσο και σε κοινωνικό επίπεδο. Παρέχεται η αυτονομία στο πλήθος, να αντιμετωπίζει ευαίσθητα θέματα χωρίς φόβο δημόσιας έκθεσης. Η διασφάλιση του απορρήτου επιτρέπει μια υγιή κοινωνία να πειραματιστεί και να αναπτυχθεί. Επιπλέον προσφέρεται ισορροπία μεταξύ των ατομικών ελευθεριών και των αρμοδιοτήτων της κυβέρνησης. Ως εκ τούτου, η προστασία των προσωπικών δεδομένων, αποτελεί θεμελιώδες στοιχείο για την οικοδόμηση μιας εύρωστης δημοκρατίας. Όμως η προστασία της ιδιωτικής ζωής, η αυτονομία, και ο έλεγχος των προσωπικών πληροφοριών μπορεί να διατρέχουν κίνδυνο, καθώς όλο και περισσότερα προσωπικά μας στοιχεία βρίσκονται στις εφαρμογές των LBS. [10, 34].

*Οι νόμοι περί προστασίας προσωπικών δεδομένων είναι εγκλωβισμένοι στο παρελθόν και αδυνατούν να εξασφαλίσουν τις απαραίτητες νομικές προστασίες για τον τρόπο που συλλέγονται οι ευαίσθητες πληροφορίες.*

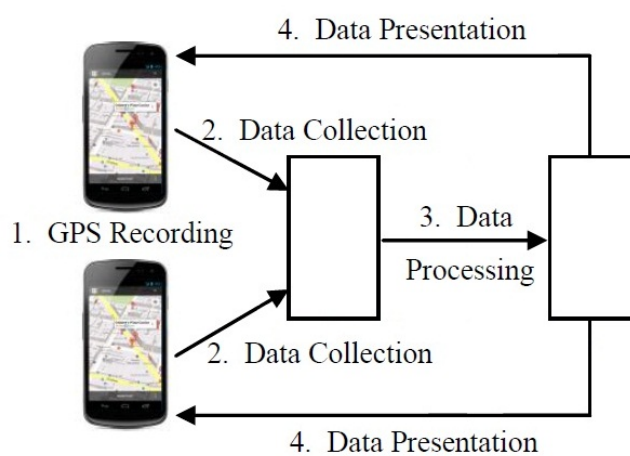
Χαρακτηριστικό παράδειγμα είναι ότι κατά τη διάρκεια των τελευταίων 40 ετών, οι δικαστικές αποφάσεις αφήναν αδιευκρίνιστο, κατά πόσον το σύνταγμα παρείχε το δικαίωμα απόκτησης ενός δικαστικώς-εγκεκριμένου εντάλματος έρευνας από τα όργανα επιβολής του νόμου, για την πρόσβαση στους διάφορους τύπους πληροφοριών και δεδομένων των LBSs [34].

## 1.2 Αντικείμενο της παρούσας διατριβής

Η πλοήγηση μέσα στο πλήθος, όπου χιλιάδες άνθρωποι κινούνται προς διαφορετικές κατευθύνσεις, μπορεί να έχει τεράστιες δυσκολίες. Εάν δεν διατίθεται μια πανοραμική άποψη

ολόκληρης της περιοχής, τότε υπάρχει περίπτωση να υπάρχει περιορισμένη ορατότητα και έλλειψη χαρτογραφικών πληροφοριών, που σαν αποτέλεσμα θα ήταν να μην καταστεί δυνατόν να εντοπιστεί η βέλτιστη διαδρομή. Τα τηλέφωνα με την αυξανόμενη υπολογιστική ισχύ και την συνδεσιμότητα τους, σε συνδυασμό με την ευρέως διαδεδομένη χρήση τους, τα καθιστά ως ένα από τα σημαντικότερα εργαλεία συλλογής δεδομένων. Υπάρχει η δυνατότητα να κάνουν χρήση των συλλεγόμενων συνολικών δεδομένων του αισθητήρα θέσιθειας (GPS) και αφού τα επεξεργαστούν, να διανείμουν πίσω στους χρήστες την κατάσταση της κυκλοφορίας σε πραγματικό χρόνο.

Το γενικό σύστημα σχεδιασμού όλων των εφαρμογών παρουσιάζεται στο ΣΧΗΜΑ: 1.1. Το σύστημα περιλαμβάνει τέσσερα κύρια στάδια: **1. Data Recording**, **2. Data Collection**, **3. Data Processing**, και **4. Data Presentation** [33].



ΣΧΗΜΑ 1.1: Γενικό σύστημα σχεδιασμού των εφαρμογών [33].

Το στάδιο **1. GPS Recording** λαμβάνει χώρα στα κινητά τηλέφωνα των χρηστών. Πραγματοποιείται καταγραφή της τρέχουσας τοποθεσίας και του χρόνου λήψης, κάνοντας χρήση του ενσωματωμένου αισθητήρα GPS.

Στο στάδιο **2. Data Collection** πραγματοποιείται συλλογή των δεδομένων. Τα δεδομένα στην συνέχεια, αποστέλλονται στον εξυπηρετητή όπου και αποθηκεύονται.

Στο στάδιο **3. Data Processing** τα δεδομένα τα επεξεργάζεται ο εξυπηρετητής σύμφωνα με τον αλγόριθμο που υλοποιεί η εκάστοτε εφαρμογή.

Στο στάδιο **4. Data Presentation** τα κινητά τηλέφωνα δημοσιεύουν, πίσω στους χρήστες, τα δεδομένα και τα αποτελέσματα που έχουν υποστεί επεξεργασία.

*Στην παρούσα διατριβή γίνεται αναφορά τόσο στο θεωρητικό μέρος, όσο και στην παρουσίαση τεχνολογιών και εφαρμογών που αναπτύχθηκαν για την συλλογή δεδομένων από τα smartphones. Κυρίως επικεντρωνόμαστε σε εφαρμογές, που σαν στόχο έχουν την βελτίωση των συνθηκών της κυκλοφορίας στις οδικές αρτηρίες.*

### 1.3 Οργάνωση του τόμου

Η παρούσα εργασία είναι οργανωμένη ως εξής:

Στο επόμενο **κεφάλαιο** αναλύεται ο όρος **crowdsourcing** δίνοντας ορισμούς και περιγράφοντας επιτυχή μοντέλα. Στο **κεφάλαιο 3**, πραγματευόμαστε τον όρο **spatial crowdsourcing**, ο οποίος αποτελεί κατηγορία του crowdsourcing, και γίνεται χρήση των δεδομένων τοποθεσίας και των χωροτακτικών εργασιών. Επίσης παρουσιάζουμε τα πρωτόκολλα που χρησιμοποιούν οι εφαρμογές. Στο **κεφάλαιο 4** γίνεται αναφορά στην τεχνολογική εξέλιξη των κινητών τηλεφώνων, και των δυνατοτήτων που προσφέρουν στο πλήθος. Στο **κεφάλαιο 5** παρουσιάζεται το ευαίσθητο θέμα της ασφάλειας των *προσωπικών δεδομένων*, και πώς υπάρχει η δυνατότητα να διαφυλάξουμε τα προσωπικά δεδομένα από τους LBS εξυπηρετητές. Στα **κεφάλαια 6 και 7** παρουσιάζονται και αναλύονται αλγοριθμικές εφαρμογές, και αφορούν την βελτίωση των συνθηκών της κυκλοφορίας στις οδικές αρτηρίες. Στο **κεφάλαιο 8** γίνεται παρουσίαση crowdsourcing εφαρμογών. Τέλος στο **κεφάλαιο 9** γίνεται απολογισμός της εργασίας. Επίσης προτείνουμε μελλοντικές έρευνες που μπορούν να γίνουν, ώστε να βελτιωθούν ακόμα περισσότερα τα υπάρχοντα συστήματα ή την δημιουργία νέων, μιας και βρισκόμαστε ακόμα σε πρώιμο επίπεδο της τεχνολογίας που αναφέρουμε.



## Κεφάλαιο 2

# Ορισμός και μοντέλο του Crowdsourcing

### 2.1 Εισαγωγή

Ο όρος **crowdsourcing** προέρχεται από την ένωση των λέξεων crowd και outsourcing. Το πλήθος, ενεργώντας συλλογικά, δύναται να επεξεργαστεί δεδομένα που έχουν συλλεχθεί από κατάλληλες εφαρμογές, και να εξάγει ιδέες και συμπεράσματα.

Το *crowdsourcing* ανατρέπει εκ βαθέων τον τρόπο με τον οποίο οι επιχειρήσεις, οι κυβερνήσεις ακόμα και οι ανθρωπιστικές οργανώσεις αλληλεπιδρούν με το διαδίκτυο και την χρήση του κινητού τηλεφώνου [2]. Ο όρος χρησιμοποιήθηκε για πρώτη φορά από τον Jeff Howe τον Ιούνιο του 2006 στην έκδοση του *Wired Magazine* με τίτλο “*The Rise of Crowdsourcing*”. Ο Jeff Howe καθόρισε, πως η έννοια του crowdsourcing εξαρτάται από το γεγονός, πως εφόσον πρόκειται για ανοικτή πρόσκληση προς ένα άγνωστο πλήθος ατόμων, συγκεντρώνει τα άτομα που είναι ικανά να αναλάβουν καθήκοντα, ώστε να λύσουν πολύπλοκα προβλήματα, και να συνεισφέρουν με νέες ιδέες [38].

*“Το crowdsourcing αντιπροσωπεύει την πράξη μιας εταιρείας ή ιδρύματος, κάνοντας λήψη μιας λειτουργίας, αφού εκτελεστεί μια φορά από απασχολούμενους, πραγματοποιώντας εξωτερική ανάθεση σε ένα αόριστο ή μη προκαθορισμένο (γενικά μεγάλο) δίκτυο ανθρώπων, υπό τη μορφή μιας ανοικτής πρόσκλησης.”*

---

JEFF HOWE  
ΙΟΥΝΙΟΣ, 2006

Σύμφωνα με το “*Merriam-Webster Dictionary*”, το *crowdsourcing* είναι η πρακτική απόκτησης αναγκαίων υπηρεσιών ή ιδεών μέσω μιας μεγάλης ομάδας ανθρώπων, και συγκεκριμένα μιας *online* οντότητας. Συνδυάζει τις προσπάθειες του πλήθους των αυτοπροσδιοριζομένων εθελοντών ή χρηστών, όπου ο κάθε ένας με δική του πρωτοβουλία, προσθέτει ένα μικρό τμήμα και ο συνδυασμός όλων των τμημάτων αποφέρει ένα μεγαλύτερο αποτέλεσμα [38].

Ο *Daren C. Brabham* ήταν ο πρώτος που όρισε τον όρο “crowdsourcing” στην επιστημονική βιβλιογραφία τον Φεβρουάριο του 2008.

“Το crowdsourcing είναι ένα διαδικτυακό μοντέλο καταναεμημένης επίλυσης προβλημάτων και λειτουργίας.”

DAREN C. BRABHAM  
ΦΕΒΡΟΥΑΡΙΟΣ, 2008

Οι *Enrique Estellés-Arolas* και *Fernando González Ladrón-de-Guevara* ανέπτυξαν έναν νέο ορισμό.

“Το crowdsourcing είναι μια μορφή συλλογικής διαδικτυακής δραστηριότητας, στην οποία ένα άτομο, ένα ίδρυμα, ένας μη κερδοσκοπικός οργανισμός ή μια εταιρεία προτείνει σε μια ομάδα ατόμων με ποικίλλες γνώσεις, ετερογένεια και αριθμό, να αναλάβει εθελοντικά μια εργασία, μέσω μιας ανοιχτής πρόσκλησης. Η ανάληψη της εργασίας η οποία ποικίλλει σε πολυπλοκότητα, το πλήθος πρέπει να συμμετάσχει με προσωπική εργασία, χρήματα, γνώση ή εμπειρία και περιλαμβάνει πάντοτε όφελος για όλους τους συμμετέχοντες. Οι συμμετέχοντες λαμβάνουν την ικανοποίηση κάποιας ανάγκης τους, είτε αυτή είναι οικονομική, είτε κοινωνική αναγνώριση, είτε προσωπική ικανοποίηση, είτε ανάπτυξη ατομικών ικανοτήτων σε κάποιο τομέα. Ο εκκινητής της πρωτοβουλίας (πληθοποριστής ή αλλιώς *crowdsourcer*) αποκτά και χρησιμοποιεί προς όφελος του, αυτά που έχει συνεισφέρει το πλήθος στο εγχείρημα.”

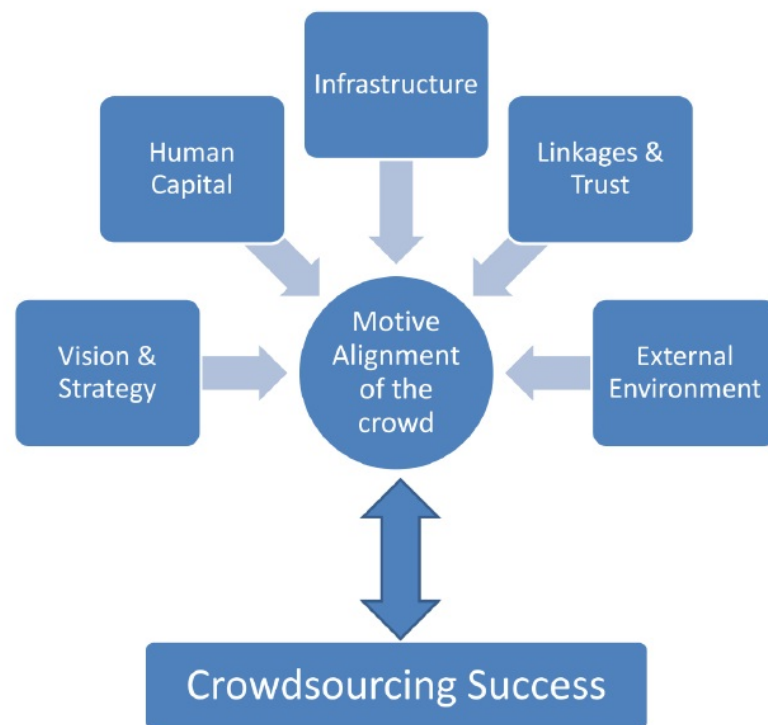
ENRIQUE ESTELLÉS-AROLAS & FERNANDO GONZÁLEZ LADRÓN-DE-GUEVARA

Με την τυπική έννοια του όρου, τα προβλήματα παρουσιάζονται σε ένα άγνωστο πλήθος χρηστών, μέσω μιας ανοιχτής πρόσκλησης, για την παροχή λύσεων. Οι χρήστες – οι οποίοι αναφέρονται και ως πλήθος – σχηματίζουν οντότητες στο διαδίκτυο, υποβάλλοντας λύσεις. Επίσης δύναται να τις αξιολογούν, επιλέγοντας τις καλύτερες. Αυτές οι λύσεις επιλέγονται από την αρχική οντότητα που έθεσε το πρόβλημα, ενώ τα φυσικά πρόσωπα που τις πρότειναν, ανταμείβονται ορισμένες φορές. Σε ορισμένες περιπτώσεις, αυτή η εργασία ανταμείβεται είτε χρηματικά, είτε με βραβεία, είτε με αναγνώριση. Σε κάποιες άλλες περιπτώσεις, η ανταμοιβή είναι απλά φήμη, ή διανοητική ικανοποίηση [38].

Την σημερινή εποχή, το crowdsourcing έχει μεταφερθεί κυρίως στο διαδίκτυο. Η πλατφόρμα του διαδικτύου παρέχεται ως ένα ιδιαίτερα καλό μέσο για crowdsourcing επικοινωνία, λόγω της ταχύτητας και της ευκαιρίας για ασύγχρονη επικοινωνία. Εξαιτίας της ανωνυμίας που παρέχεται, το πλήθος τείνει να είναι περισσότερο ανοικτό σε web-based και mobile-based projects [9, 38].

## 2.2 Παράγοντες εξασφάλισης επιτυχούς crowdsourcing μοντέλου

Στο ΣΧΗΜΑ: 2.1 παρουσιάζονται οι παράγοντες που καθιστούν μια crowdsourcing εφαρμογή επιτυχημένη. Ο κύριος παράγοντας είναι ο *συγχρονισμός των κινήτρων του πλήθους*. Οι περιφερειακοί παράγοντες είναι *το όραμα & η στρατηγική*, *το πλήθος*, *οι υποδομές, οι διασυνδέσεις & η εμπιστοσύνη*, και *το εξωτερικό περιβάλλον*. Το μοντέλο που προκύπτει έχει σαν στόχο να παρουσιάσει μια νέα προσέγγιση του crowdsourcing [2].



ΣΧΗΜΑ 2.1: Παράγοντες διασφάλισης επιτυχούς crowdsourcing μοντέλου [2].

Όλες οι crowdsourcing εφαρμογές εισέρχονται στην αγορά με ένα καλώς ορισμένο σύνολο στόχων και ιδανικών. Το όραμα της εφαρμογής είναι πολύ σημαντικό για το πλήθος. Είναι επιτακτική ανάγκη ο κόσμος να αντιλαμβάνεται την εφαρμογή ως πολύτιμη και καλοπροαίρετη. Εάν είναι δυνατόν, θα πρέπει επίσης να παρέχονται και κίνητρα συμμετοχής. Η ορθή διαχείριση του οράματος και της στρατηγικής εξασφαλίζει, κατά κύριο λόγο, επαρκή συμμετοχή του πλήθους.

Επιπλέον, μια σωστά τεκμηριωμένη παρουσίαση του οράματος, μπορεί να οδηγήσει στην υποστήριξη της εφαρμογής από τις κυβερνήσεις, τις επιχειρήσεις ή άλλων άμεσα ενδιαφερομένων. Η κρατική υποστήριξη προσθέτει έναν επαρκή παράγοντα εμπιστοσύνης. Επιπρόσθετα, η υποστήριξη εγγυάται μια ευρύτερη συμμετοχή, και αυξάνει την προβολή της εφαρμογής εξασφαλίζοντας με τον τρόπο αυτόν τη συμμετοχή του πλήθους.

*Η λογική συνοχή του οράματος και της στρατηγικής, σε συνδυασμό με τις προσδοκίες του πλήθους, διασφαλίζει ότι ο κόσμος πρόθυμα θα συμμετάσχει στην εφαρμογή.*

Ένας άλλος καθοριστικός παράγοντας της επιτυχίας του crowdsourcing είναι οι δεξιότητες και οι ικανότητες που διαθέτει το πλήθος. Ο *Erran Carmel* περιγράφει τα χαρακτηριστικά, τις δεξιότητες και τις ικανότητες που διαθέτει συλλογικά το πλήθος ως *Human Capital*. Το “ανθρώπινο κεφάλαιο” περιλαμβάνει γλωσσικές δεξιότητες, διαχειριστικές ικανότητες, εθνικό προσανατολισμό, παραδόσεις, καθώς και το επίπεδο εκπαίδευσης των ανθρώπων [2, 12].

Συνεπώς, προκειμένου να καταστεί ουσιαστική η συμμετοχή του πλήθους σε crowdsourcing εφαρμογές, υπάρχει ανάγκη να αναπτυχθούν κατάλληλες δεξιότητες. Η εξέλιξη του κινητού τηλεφώνου παρέχει την δυνατότητα υποστήριξης crowdsourcing εφαρμογών. Επίσης, οι επιχειρηματίες πρέπει να διαθέτουν τις απαραίτητες δεξιότητες, την εμπειρία, και το όραμα, για να προωθήσουν μια crowdsourcing εφαρμογή επιτυχημένα στο πλήθος.

Επομένως η επιτυχία του crowdsourcing εξαρτάται από την προσέλκυση ανθρώπων να συμμετάσχουν στις εφαρμογές. Όσο περισσότερα άτομα συμμετέχουν, τόσο πιο επιτυχημένη κρίνεται. Το πλήθος θα πρέπει να είναι σε θέση να χρησιμοποιεί τις εφαρμογές, χωρίς προηγούμενη εκπαίδευση και με ελάχιστες παρεμβάσεις από τους προγραμματιστές.

Οι περισσότερες από τις crowdsourcing εφαρμογές είναι είτε mobile-based ή web-based. Το crowdsourcing απαιτεί άφθονη, αξιόπιστη και φθηνή πρόσβαση στις επικοινωνιακές ανάγκες, ώστε να εξασφαλιστεί η συμμετοχή του πλήθους. Συνεπώς, είναι σημαντικό οι τεχνολογικές υποδομές να είναι ανεπτυγμένες.

Για την ανάπτυξη των υποδομών, απαιτούνται κατάλληλες πηγές κεφαλαίων. Οι πηγές των κεφαλαίων μπορεί να προέρχονται από έναν συνδυασμό εγχώριων και ξένων επενδύσεων. Με αυτόν τον τρόπο, η επαρκής βελτίωση της τεχνολογικής υποδομής, μπορεί να ενισχύσει αισθητά τη συμμετοχή του πλήθους, προσφέροντας αξιόπιστη και συνεχή ροή επικοινωνίας.

Η σωστή ανάπτυξη των διασυνδέσεων είναι ένας τρόπος για την ελαχιστοποίηση του κόστους και αύξησης της εμπιστοσύνης. Ο Erran Carmel αναφέρει

*“Η διασύνδεση είναι κάτι που προκύπτει μεταξύ ατόμων, ομάδων εργασίας, επιχειρήσεων ή και εθνών λόγω της γεωγραφικής, πολιτιστικής, γλωσσικής ή εθνοτικής σύνδεσης.”*

ERRAN CARMEL  
2003

Οι σωστές διασυνδέσεις είναι απαραίτητες για τα ανθρωποκεντρικά μοντέλα. Μέσω αυτών, η μεταφορά γνώσης γίνεται ευκολότερη. Όλες οι crowdsourcing εφαρμογές για να λειτουργήσουν ορθά, χρειάζονται αξιόπιστες πληροφορίες από το πλήθος, οι οποίες καθιστούν απαραίτητη την ανάπτυξη του παράγοντα *εμπιστοσύνη*. Οι σωστές διασυνδέσεις, καθώς και η ύπαρξη σχέσεων με επιτυχημένες υπάρχουσες ενέργειες, καθιστούν μια crowdsourcing εφαρμογή έμπιστη.

Ένα οικονομικό περιβάλλον προώθησης του επιχειρηματικού πνεύματος βοηθά στην επιτυχία της εφαρμογής. Ιδιαίτερη όμως προσοχή πρέπει να δοθεί στους κινδύνους που ελλοχεύουν για θέματα ασφαλείας, καθώς και ρυθμιστικούς κινδύνους.

*“Μέσω όλων των ανωτέρω παραγόντων, υπάρχει η δυνατότητα επίτευξης συγχρονισμού των κινήτρων του πλήθους”.*

Ο συγχρονισμός των κινήτρων του πλήθους αποτελεί τον καθοριστικό παράγοντα του μοντέλου. Ορίζεται ως “ο βαθμός στον οποίο το πλήθος είναι σε θέση να συνδεθεί με τον μακροπρόθεσμο στόχο της crowdsourcing εφαρμογής, ενθαρρύνοντας με αυτόν τον τρόπο την ευρύτερη συμμετοχή του”. Η επίδοση και η προσπάθεια που αναμένεται, καθώς επίσης η κοινωνική επιρροή και η “χαλάρωση” των προϋποθέσεων, είναι οι άμεσοι καθοριστικοί παράγοντες της αποδοχής των ατόμων έναντι μιας τεχνολογία που βασίζεται σε αυτούς.

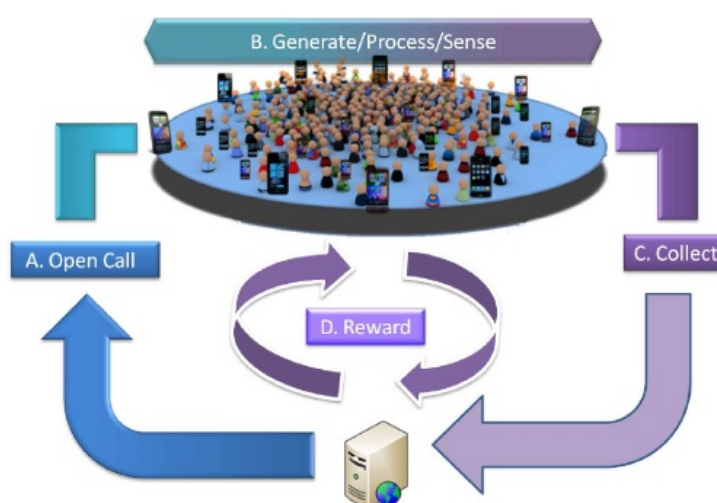
Ο πίνακας του ΣΧΗΜΑΤΟΣ: 2.2 απεικονίζει την επίδραση που φέρουν οι περιφερειακοί παράγοντες στους καθοριστικούς παράγοντες. Οι περιφερειακοί παράγοντες δεν επηρεάζουν όλους τους καθοριστικούς παράγοντες, αλλά κάποιους συγκεκριμένους [2].

	Performance Expectancy	Effort Expectancy	Social Influence	Facilitating Conditions
Vision & Strategy	✓		✓	
Human Capital	✓	✓		
Linkages & Trust	✓		✓	
Infrastructure		✓		✓
External Environment		✓	✓	✓

ΣΧΗΜΑ 2.2: Επίδραση περιφερειακών παραγόντων στους καθοριστικούς παράγοντες [2].

### 2.3 Χρησιμότητα των smartphones στο crowdsourcing μοντέλο

Όπως έχει διατυπωθεί, το crowdsourcing αναφέρεται σε ένα καταναμημένο μοντέλο επίλυσης προβλημάτων στο οποίο το “απροσδιόριστου μεγέθους” πλήθος ασχολείται για την επίλυση ενός πολύπλοκου προβλήματος, μέσω μιας ανοικτής πρόσκλησης ΣΧΗΜΑ: 2.3.



ΣΧΗΜΑ 2.3: Διαδικασία του crowdsourcing [7].

Τα smartphones είναι ευρέως διαδεδομένα, έχουν καθημερινή χρήση και είναι σχεδόν πάντα συνδεδεμένα στο διαδίκτυο. Ως εκ τούτου, προσφέρονται για την επέκταση των *υφιστάμενων web-based crowdsourcing εφαρμογών* σε *mobile-based*. Αυτή η ανάπτυξη, καθιστά την συλλογή πληροφοριών ευκολότερη και επίκαιρη. Επιπλέον οι δυνατότητες των smartphones, με τους εξελιγμένους ενσωματωμένους αισθητήρες που έχουν εγκαταστημένους, συμβάλλουν στην ανάπτυξη νέων αποτελεσματικών εφαρμογών crowdsourcing.

Ένα άλλο βασικό χαρακτηριστικό του crowdsourcing είναι αν η συνεισφορά του πλήθους είναι *ενεργητική* ή *παθητική*. Ειδικότερα, οι υπολογισμοί που εκτελούνται και τα δεδομένα που συλλέγονται από τους χρήστες είναι η είσοδος για *ενεργητικό crowdsourcing*. Οι εφαρμογές crowdsourcing που παρέχονται διαδικτυακά είναι *ενεργητικές*, δεδομένου ότι απαιτούν την ενεργό συμμετοχή του χρήστη. Αντίθετα οι crowdsourcing εφαρμογές που κατηγοριοποιούνται ως παθητικές έχουν σαν είσοδο τα δεδομένα που παράγονται από τους αισθητήρες, και οι υπολογισμοί εκτελούνται από τις συσκευές αυτόματα, κατόπιν αρχικής έγκρισης από τον χρήστη.

Οι Geiger και Quinn κάνουν αναφορά και σε περαιτέρω ταξινομήσεις. Ο Quinn ταξινομεί τις εφαρμογές σύμφωνα με τις δεξιότητες των χρηστών, όπως είναι η οπτική αναγνώριση, η κατανόηση της γλώσσας ή της επικοινωνίας. *Αξίζει να σημειωθεί, ότι η ανθρώπινη δεξιότητα αξιολογείται μόνο σε εφαρμογές με ενεργητική συμμετοχή του χρήστη*. Επιπρόσθετα, γίνεται διαχωρισμός με βάση τα κίνητρα που χρησιμοποιούνται για την προσέλκυση του πλήθους, όπως για παράδειγμα απόδοση αμοιβής ή απλά φήμης. Ο Geiger ταξινομεί τις εφαρμογές σύμφωνα με το πόσο ποιοτική είναι η συνεισφορά του πλήθους, η οποία μπορεί να είναι είτε *ομοιογενής* είτε *ετερογενής*. Στην ομογενή ταξινόμηση κάθε συνεισφορά έχει την ίδια αξία, ενώ στην ετερογενή κάθε συνεισφορά αξιολογείται χωριστά, ή σε σύγκριση με τις υπόλοιπες [7].

Στο επόμενο κεφάλαιο γίνεται αναφορά στο **Spatial Crowdsourcing**. Είναι μια κατηγορία του crowdsourcing, αλλά με χρήση δεδομένων τοποθεσίας και χωροτακτικών εργασιών (*spatial tasks*).

## Κεφάλαιο 3

# Spatial Crowdsourcing

### 3.1 Εισαγωγή

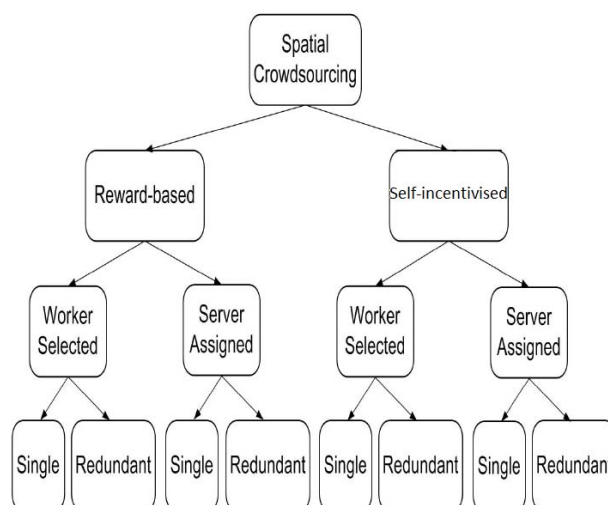
Το πλήθος έχοντας στην κατοχή του κινητά τηλέφωνα με ανεπτυγμένους αισθητήρες, μπορεί να πραγματοποιεί συλλογή διαφόρων τύπων δεδομένων. Για παράδειγμα δύναται να συλλέξει ήχο, εικόνα, βίντεο, πιθανές ταχύτητες αντικειμένων, ακόμα και γεωγραφικές συντεταγμένες. Η αξιοποίηση των δεδομένων τοποθεσίας που συλλέγονται, συνθέτουν μια νέα πλατφόρμα του crowdsourcing, το spatial crowdsourcing.

Στο παρόν κεφάλαιο γίνεται παρουσίαση του γενικού πλαισίου του spatial crowdsourcing, και πώς αυτό λειτουργεί. Στη συνέχεια γίνεται αναφορά στο γενικό πρόβλημα ανάθεσης εργασιών, πώς δύναται να αναχθεί σε διάγραμμα δικτύου ροής και κατόπιν να επιλυθεί.

### 3.2 Κατηγοριοποίηση του Spatial Crowdsourcing

Το spatial crowdsourcing παρουσιάζεται ως ο νέος μηχανισμός για χωρικές εφαρμογές, δηλαδή εφαρμογές που σχετίζονται με την τοποθεσία και τις γεωγραφικές συντεταγμένες. Η εκτέλεση των εφαρμογών πραγματοποιείται από τους χρήστες, και απαιτείται να βρίσκονται σε κατάλληλη τοποθεσία. Στο ΣΧΗΜΑ: 3.1 απεικονίζεται το γενικό πλαίσιο του spatial crowdsourcing. Η ταξινόμηση χωρίζεται κατ' αρχήν με βάση τα κίνητρα των ανθρώπων, κατόπιν ορίζονται δύο τρόποι υποστήριξης των εφαρμογών, και τέλος γίνεται αναφορά σε δύο τρόπους ανάθεσης των εφαρμογών στους χρήστες [24].

*Να σημειωθεί ότι το πλαίσιο που εφαρμόζεται για το spatial crowdsourcing, ακολουθείται από όλες τις εφαρμογές που πραγματοποιούν επεξεργασία δεδομένων που προέρχονται από το πλήθος.*



ΣΧΗΜΑ 3.1: Κατηγοριοποίηση του spatial crowdsourcing [24].

Ο πρώτος διαχωρισμός πραγματοποιείται με βάση τα κίνητρα που παρέχονται στους χρήστες, και ξεχωρίζουν δύο κατηγορίες.

Στην *reward-based* κατηγορία, οι χρήστες επιβραβεύονται για την ορθή εκτέλεση της εφαρμογής, λαμβάνοντας μια ανταμοιβή όπως για παράδειγμα κάποια χρηματική αμοιβή.

Αντίθετα, η κατηγορία *self-incentivised* αφορά χρήστες που είναι πρόθυμοι να εκτελούν εθελοντικά εφαρμογές, χωρίς να προσδοκούν κάποια αμοιβή. Για παράδειγμα, μια ομάδα χρηστών που είναι πρόθυμοι να αναφέρουν κυκλοφοριακά γεγονότα εθελοντικά, με την χρήση των αισθητήρων που είναι εξοπλισμένες οι κινητές συσκευές τους.

Σε δεύτερο στάδιο, η εκτέλεση των εργασιών μπορεί να υλοποιηθεί με δύο τρόπους.

Στην μια περίπτωση, οι εργασίες εκτελούνται από τους χρήστες (*Worker Selected Tasks*). Οι χρήστες, οι οποίοι βρίσκονται συνδεδεμένοι, μπορούν να επιλέξουν οποιαδήποτε εργασία μπορεί να εκτελεστεί κοντά στην περιοχή που βρίσκονται, χωρίς την ανάγκη συντονισμού με τον εξυπηρετητή, ο οποίος δημοσιεύει τις εργασίες. Το θετικό με την αυτόνομη επιλογή εργασιών είναι ότι οι χρήστες δεν χρειάζεται να αποκαλύψουν την πραγματική τους τοποθεσία στον εξυπηρετητή για κάθε ανάθεση. Το αρνητικό όμως είναι ότι ορισμένες εργασίες μπορεί να μην ανατεθούν ποτέ, ενώ άλλες να ανατίθενται σε πλεονασμό.

Αντίθετα, στην άλλη περίπτωση οι εργασίες ανατίθενται από τους εξυπηρετητές (*Server Assigned Tasks*). Ο εξυπηρετητής δεν δημοσιεύει τις εργασίες στους χρήστες. Αντ' αυτού, ο κάθε χρήστης αποστέλλει την τοποθεσία που βρίσκεται στον κεντρικό εξυπηρετητή. Μετά την παραλαβή όλων των θέσεων των συνδεδεμένων χρηστών, ο εξυπηρετητής αντιστοιχίζει σε κάθε χρήστη, εργασίες βάση της τοποθεσίας που έχει. Το πλεονέκτημα της SAT είναι ότι ο εξυπηρετητής έχει όλες τις πληροφορίες, και μπορεί να αποδώσει σε κάθε χρήστη τις πιο κοντινές



εργασίες, μεγιστοποιώντας παράλληλα τη συνολική ανάθεση εργασιών. Ωστόσο, το μειονέκτημα είναι ότι οι χρήστες είναι υποχρεωμένοι να αναφέρουν τις θέσεις τους για κάθε ανάθεση, η οποία αναφορά μπορεί να θέσει σε κίνδυνο την ιδιωτική ζωή [24].

Στο τρίτο και τελευταίο στάδιο, η μέθοδος ανάθεσης των προς εκτέλεση εργασιών μπορεί να διαχωριστεί με δύο τρόπους.

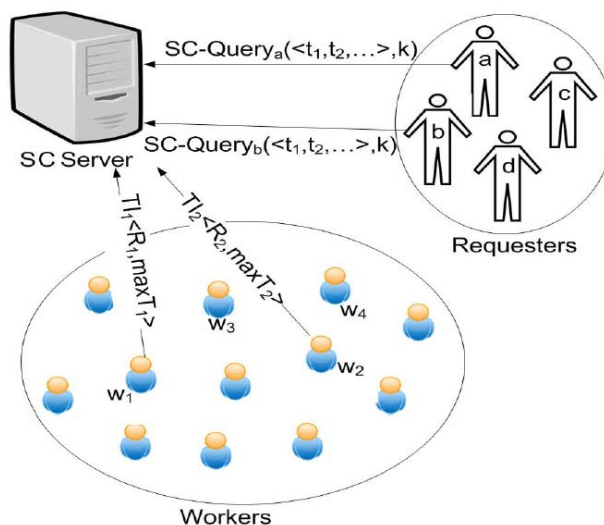
Με ατομική ανάθεση εργασιών (*Single Task Assignment*). Ο κάθε χρήστης θεωρείται αξιόπιστος, και ως εκ τούτου ολοκληρώνει τις εργασίες ορθά, χωρίς κακόβουλες προθέσεις. Κάθε εργασία αποδίδεται μόνο σε έναν χρήστη, κατά προτίμηση στον πλησιέστερο της κάθε εργασίας.

Με ομαδική ανάθεση εργασιών (*Redundant Task Assignment*). Με αυτή την λειτουργία, αντί κάθε εργασία να εκτελείται από έναν συγκεκριμένο χρήστη, εκτελείται από τους  $k$  κοντινούς χρήστες, όπου το  $k$  ορίζεται από τον αιτούντα ο οποίος εξέδωσε την εργασία. Συνεπώς, όσο μεγαλύτερη είναι η τιμή του  $k$ , τόσο μεγαλύτερη είναι και η πιθανότητα ότι θα ολοκληρωθεί σωστά η εργασία, και σε σύντομο χρονικό διάστημα [24].

*Βασική υπόθεση, είναι η εγκυρότητα των δεδομένων που στέλνει ο κάθε χρήστης στον εξυπηρετητή.*

### 3.3 Περιγραφή Ανάθεσης των Εργασιών

Έχοντας το γενικό πλαίσιο ανάθεσης εργασιών του spatial crowdsourcing, όπως απεικονίζεται στο ΣΧΗΜΑ: 3.2, γίνεται περιγραφή πώς πραγματοποιείται η ανάθεση των εργασιών.



ΣΧΗΜΑ 3.2: Πλαίσιο ανάθεσης εργασιών του spatial crowdsourcing [24].

Οι χρήστες, οι οποίοι συμβολίζονται ως  $w$ , φέρουν μια φορητή συσκευή και εκτελούν, εθελοντικά ή μη, χωρικές εργασίες. Όταν βρίσκονται σε online λειτουργία είναι έτοιμοι να δεχτούν εργασίες, ειδάλλως βρίσκονται σε offline λειτουργία.

Μια χωρική εργασία  $t$  είναι της μορφής  $\langle l, q, s, \delta \rangle$ . Το  $l$  είναι μια τοποθεσία στον  $2D$  χώρο, και το  $q$  είναι ένα ερώτημα που πρέπει να απαντηθεί στην θέση  $l$ . Το ερώτημα  $q$  τίθεται σε χρόνο  $s$  και λήγει σε χρόνο  $s + \delta$ . Επομένως, το ερώτημα  $q$  μιας χωρικής εργασίας  $t$  μπορεί να εκτελεστεί, μόνο όταν ο χρήστης βρίσκεται γεωγραφικά στην θέση  $l$ . Για παράδειγμα, έστω ότι η εργασία είναι να ληφθεί μια φωτογραφία από ένα συγκεκριμένο κτίριο. Αυτό σημαίνει ότι ο χρήστης πρέπει να μετακινηθεί στην ακριβή τοποθεσία του κτιρίου, προκειμένου να γίνει λήψη της φωτογραφίας.

Το σύνολο από χωρικές εργασίες συνθέτει ένα χωρικό ερώτημα (*SC-Query*). Το χωρικό ερώτημα είναι της μορφής  $(\langle t_1, t_2, \dots \rangle, k)$ , όπου η παράμετρος  $k$  ορίζεται από τον χρήστη που ζητάει πληροφορίες (*requester*), και κάθε χωρική εργασία  $t_i$  εκτελείται  $k$  φορές.

Οι χρήστες  $w$ , στέλνουν στον *SC-server* ερωτήματα εργασιών (*tasks inquiries*). Το ερώτημα εργασίας είναι ένα ερώτημα που στέλνει ένας online χρήστης στον *SC-εξυπηρετητή*, όταν είναι έτοιμος να εκτελέσει εργασίες. Το ερώτημα περιλαμβάνει την θέση  $l$  του χρήστη  $w$  με δύο, όμως, περιορισμούς. Ο πρώτος είναι ο ορισμός μιας χωρικής περιοχής  $R$ , όπου ο χρήστης μπορεί να δεχθεί χωρικές εργασίες. Η απεικόνισή της παραπέμπει σε ένα εικονικό ορθογώνιο παραλληλόγραμμα. Οποιαδήποτε εργασία υπάρχει έξω από την περιοχή αυτή, δεν γίνεται δεκτή από τον χρήστη. Ο δεύτερος περιορισμός είναι ο μέγιστος αριθμός των εργασιών που μπορεί να εκτελέσει, και συμβολίζεται ως  $maxT$ . Οι χρήστες μπορούν επίσης να καθορίσουν και άλλους περιορισμούς στο ερώτημα εργασίας, όπως για παράδειγμα τον διαθέσιμο χρόνο που διαθέτουν [24].

Μετά από την παραλαβή των *SC-Queries* από όλους τους αιτούντες, και αφού οι χρήστες ενημερώσουν τον εξυπηρετητή ότι είναι διαθέσιμοι, ο *SC-εξυπηρετητής* αναθέτει τις χωρικές εργασίες των *SC-Queries* στους διαθέσιμους χρήστες, λαμβάνοντας υπ' όψιν πιθανούς περιορισμούς. Η ανάθεση μπορεί να γίνει είτε ατομικά είτε ομαδικά.

Ωστόσο, επειδή ο *SC-εξυπηρετητής* δέχεται συνεχώς *SC-Queries* από τους αιτούντες, καθώς και ερωτήματα εργασιών από τους χρήστες, η ανάθεση των εργασιών δεν είναι μια απλή διαδικασία.

Έστω ότι την χρονική στιγμή  $s_i$ , το σύνολο  $W_i = \{w_1, w_2, \dots\}$  είναι οι online χρήστες, και το σύνολο  $T_i = \{t_1, t_2, \dots\}$  είναι οι διαθέσιμες εργασίες. Το σύνολο των ανατεθέντων εργασιών, το οποίο συμβολίζεται ως  $I_i$ , είναι το σύνολο πλειάδων της μορφής  $\langle w, t \rangle$ , όπου μια χωρική εργασία  $t$  έχει ανατεθεί σε έναν χρήστη  $w$ , ενώ ικανοποιεί πιθανούς περιορισμούς που έχουν τεθεί. Επίσης το  $|I_i|$  υποδηλώνει τον αριθμό των εργασιών που έχουν ανατεθεί την χρονική στιγμή  $s_i$ .

Επομένως, αφού το σύνολο των ανατεθέντων εργασιών πρέπει να πλήρη τους περιορισμούς των χρηστών, συνεπάγεται ότι για κάθε πλειάδα  $\langle w, t \rangle$  που ανήκει στο σύνολο  $I_i$ , η χωρική εργασία  $t$  πρέπει να βρίσκεται μέσα στη χωρική περιοχή  $R$  του χρήστη  $w$ . Επιπλέον, σε κάθε χρήστη  $w$  μπορεί να ανατίθενται το πολύ  $maxT$  εργασίες.

Η ιδανική περίπτωση είναι όταν όλες οι εργασίες ανατίθενται σε όλους τους χρήστες. Αυτό όμως μπορεί να μην είναι εφικτό, λόγω των περιορισμών των χρηστών. Ο στόχος λοιπόν είναι να μεγιστοποιηθεί ο αριθμός των ανατεθέντων εργασιών.

Λαμβάνοντας υπ' όψιν ένα χρονικό διάστημα  $\phi = \{s_1, s_2, \dots, s_n\}$ , και ότι  $|I_i|$  είναι ο αριθμός των ανατεθέντων εργασιών την χρονική στιγμή  $s_i$ , το πρόβλημα του μέγιστου καταμερισμού των εργασιών (maximum task assignment problem) είναι η διαδικασία ανάθεσης εργασιών στους χρήστες κατά τη διάρκεια του χρονικού διαστήματος  $\phi$ , όταν ο συνολικός αριθμός των ανατεθέντων εργασιών ( $\sum_{i=1}^n |I_i|$ ) είναι μέγιστος [24].

### 3.4 Μέθοδοι Ανάθεσης των Εργασιών

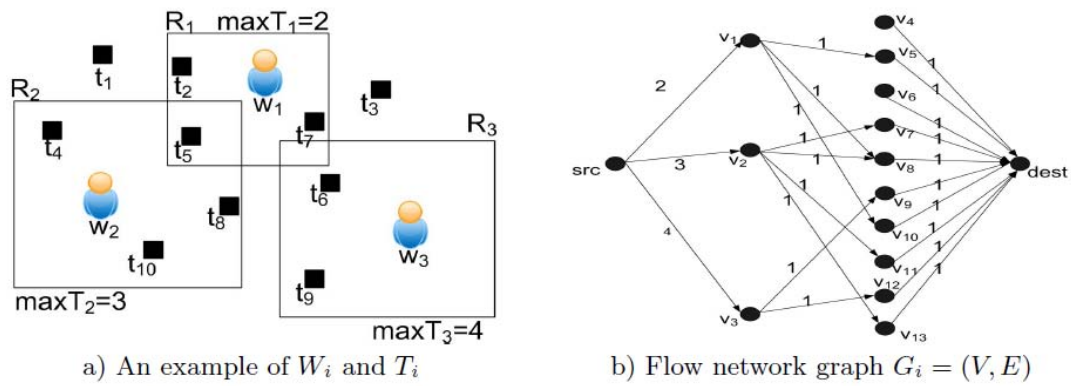
Προκειμένου να γίνει επίλυση του προβλήματος της μέγιστης ανάθεσης των εργασιών, ο SC-server πρέπει να έχει γνώση όλων των χωρικών εργασιών, και όλους τους διαθέσιμους χρήστες. Έτσι επιτρέπεται στον SC-server να αποδώσει βέλτιστα κάθε εργασία σε κάθε χρήστη, έτσι ώστε ο συνολικός αριθμός των ανατεθέντων εργασιών να μεγιστοποιηθεί. Ωστόσο, ο SC-server δεν κατέχει τέτοια γνώση. Σε κάθε χρονική στιγμή, ο SC-server λαμβάνει μια σειρά νέων εργασιών από τους αιτούντες, αλλά και ένα σύνολο από νέα ερωτήματα εργασιών από τους χρήστες. Ως εκ τούτου, ο SC-server έχει μόνο μια μερική άποψη των διαθέσιμων εργασιών και χρηστών, για ένα χρονικό διάστημα.

Μέσω της δυνατότητας μετατροπής του προβλήματος της μέγιστης ανάθεσης των εργασιών σε πρόβλημα εύρεσης μέγιστης ροής (maximum flow problem), έχει αναπτυχθεί η *Greedy στρατηγική*. Με χρήση του γνωστού αλγορίθμου FORD-FULKERSON, πραγματοποιείται ο υπολογισμός της μέγιστης ροής του συστήματος. Ωστόσο η Greedy στρατηγική, λόγω της μεγιστοποίησης της τρέχουσας ανάθεσης, δεν λαμβάνει υπόψη μελλοντικές βελτιστοποιήσεις. Για τον λόγο αυτό, γίνεται χρήση της *εντροπίας της τοποθεσίας*. [24].

#### 3.4.1 Στρατηγική Greedy

Ο λόγος που αυτή η προσέγγιση ονομάζεται Greedy, είναι γιατί σε κάθε χρονικό διάστημα προσπαθεί να μεγιστοποιήσει μόνο την τρέχουσα ανάθεση εργασιών.

Στο ΣΧΗΜΑ: 3.3 γίνεται παρουσίαση πώς μετατρέπεται ένα πρόβλημα μέγιστης ανάθεσης εργασιών σε πρόβλημα εύρεσης μέγιστης ροής του δικτύου, την χρονική στιγμή  $s_i$ . Στο ΣΧΗΜΑ:



ΣΧΗΜΑ 3.3: Παράδειγμα μετατροπής προβλήματος μέγιστης ανάθεσης εργασιών σε πρόβλημα εύρεσης μέγιστης ροής του δικτύου [24].

3.3a εμφανίζεται ένα σύνολο χρηστών  $W_i$  και ένα σύνολο εργασιών  $T_i$ . Κάθε  $w_j$  χρήστης συσχετίζεται με μια χωρική περιοχή  $R_j$ . Το αντίστοιχο γράφημα του δικτύου ροής απεικονίζεται στο ΣΧΗΜΑ: 3.3b.

Όπως φαίνεται στο σχήμα 3.3a, ο χρήστης  $w_1$  μπορεί να δεχτεί εργασίες μέσα στην δική του χωρική περιοχή  $R_1$ . Οι εργασίες αυτές είναι οι  $t_2$ ,  $t_5$  και  $t_7$ . Συνεπώς, η κορυφή  $v_1$  που έχει αντιστοιχηθεί στον  $w_1$ , μπορεί να μεταφέρει ροή μόνο προς τις κορυφές  $v_5$ ,  $v_8$  και  $v_{10}$ , δηλαδή αυτές που έχουν αντιστοιχηθεί με τις αντίστοιχες εργασίες. Επιπλέον, ο  $w_1$  είναι διατεθειμένος να εκτελέσει δύο εργασίες, αφού έχει τεθεί ότι  $\max T_1 = 2$ . Άρα η χωρητικότητα της ακμής ( $src, v_1$ ) ισούται με 2. Τέλος, η χωρητικότητα όλων των ακμών που συνδέουν τις κορυφές  $v_4, v_5, \dots, v_{13}$  με την κορυφή προορισμού  $dest$  ισούται με 1. Αυτό συμβαίνει διότι κάθε χωρική εργασία πρόκειται να ανατεθεί σε έναν μόνο χρήστη.

Η μεθοδολογία αυτή επαναλαμβάνεται για τους χρήστες  $w_2$  και  $w_3$ , έως ότου το πρόβλημα ανάθεσης εργασιών, αναχθεί σε πρόβλημα εύρεσης μέγιστης ροής. Κατόπιν δύναται να γίνει χρήση οποιουδήποτε αλγορίθμου, που υπολογίζει την μέγιστη ροή του δικτύου, ώστε να επιλυθεί το πρόβλημα. Μια από τις γνωστές τεχνικές για τον υπολογισμό της μέγιστης ροής είναι ο αλγόριθμος *Ford-Fulkerson* [5, 25]. Ο αλγόριθμος *Ford-Fulkerson* ξεκινά στέλνοντας ροή από την κορυφή πηγή  $src$ , προς την κορυφή προορισμού  $dest$ , όσο υπάρχει μονοπάτι με διαθέσιμη χωρητικότητα [24].

Παρόλα αυτά, ένας άλλος ευρετικός μηχανισμός που μπορεί να βελτιώσει την διαδικασία ανάθεσης των εργασιών, βασίζεται στην αξιοποίηση των χωρικών πληροφοριών της τοποθεσίας. Δεδομένου ότι κάθε χωρική εργασία είναι συνδεδεμένη με μια συγκεκριμένη τοποθεσία, είναι πιθανότερο να ολοκληρωθεί όταν οριοθετείται σε περιοχές με υψηλό αριθμό χρηστών. Επομένως, η ανάθεση των εργασιών που βρίσκονται σε περιοχές με μικρό αριθμό χρηστών έχει μεγαλύτερη προτεραιότητα.

### 3.4.2 Στρατηγική Least Location Entropy Priority (LLEP)

Για την λύση του προβλήματος, γίνεται χρήση της εντροπίας της τοποθεσίας. Πραγματοποιείται μέτρηση του αριθμού των διαθέσιμων χρηστών, καθώς επίσης πόσο συχνά πρόκειται να επισκεφθούν τη συγκεκριμένη τοποθεσία. Μια τοποθεσία έχει υψηλή εντροπία, εάν πολλοί χρήστες την επισκεφθούν με ίσες αναλογίες. Αντίθετα, μια τοποθεσία έχει χαμηλή εντροπία, εάν η κατανομή των επισκέψεων περιορίζεται σε λίγους μόνο χρήστες.

Επομένως, η στρατηγική που ακολουθείται είναι να δοθεί μεγαλύτερη προτεραιότητα στις εργασίες που βρίσκονται σε περιοχές με χαμηλή εντροπία, διότι οι εργασίες αυτές έχουν μικρότερη πιθανότητα να ολοκληρωθούν από τους χρήστες.

Η εντροπία μιας τοποθεσίας ορίζεται ως εξής:

$$Entropy(l) = - \sum_{w \in W_l} P_l(w) \times \log P_l(w) \quad (3.1)$$

Το σύνολο  $O_l$  περιλαμβάνει τον αριθμό των επισκέψεων για μια δοθείσα τοποθεσία  $l$ . Επομένως, το  $|O_l|$  προσδιορίζει τον συνολικό αριθμό των επισκέψεων. Το σύνολο  $W_l$ , είναι το σύνολο των χρηστών που επισκέφθηκαν την τοποθεσία  $l$ , και το  $|O_{w,l}|$  είναι το σύνολο των επισκέψεων που έκανε ο κάθε χρήστης  $w$ , στην τοποθεσία. Η πιθανότητα μια τυχαία επιλογή από το σύνολο  $O_l$  να ανήκει στο  $O_{w,l}$  είναι  $P_l(w) = \frac{|O_{w,l}|}{|O_l|}$ , και είναι το κλάσμα του συνόλου των επισκέψεων του χρήστη  $w$  στην τοποθεσία  $l$ .

Για κάθε εργασία  $t_i$  της μορφής  $\langle l_i, q_i, s_i, \delta_i \rangle$  δύναται να συσχετιστεί ένα συγκεκριμένο κόστος, που είναι η εντροπία της τοποθεσίας  $l_i$ . Ως εκ τούτου, οι εργασίες με χαμηλό κόστος έχουν υψηλή προτεραιότητα σε αντίθεση με αυτές που έχουν υψηλό κόστος, δεδομένου ότι έχουν μικρότερη πιθανότητα να ολοκληρωθούν. Στόχος είναι η πραγματοποίηση της μέγιστης ανάθεσης των εργασιών για κάθε χρονικό διάστημα, ενώ παράλληλα το συνολικό κόστος που συσχετίζεται με τις ανατιθέμενες εργασίες είναι χαμηλό [24].

Αφού κατασκευαστεί το γράφημα  $G_i = (V, E)$  για την χρονική στιγμή  $s_i$ , στόχος είναι να βρεθεί το ελάχιστο συνολικό κόστος της ροής. Κάθε ακμή  $(u, v) \in E$  έχει χωρητικότητα  $c(u, v) > 0$ , ροή  $f(u, v) \geq 0$ , και κόστος  $a(u, v) \geq 0$ . Το κόστος της ροής  $f(u, v)$  είναι  $f(u, v) \times a(u, v)$ . Έστω ότι  $f_{max}$ , είναι η μέγιστη ροή που στέλνεται από την πηγή  $src$  προς τον προορισμό  $dest$ . Τότε το ελάχιστο συνολικό κόστος της ροής ορίζεται ως:

$$\sum_{(u,v) \in E} f(u, v) \times a(u, v) \quad (3.2)$$

με τους εξής περιορισμούς:

$$f(u, v) \leq c(u, v) \quad (3.3)$$

$$f(u, v) = -f(u, v) \quad (3.4)$$

$$\sum_{w \in V} f(u, w) = 0, \forall u \neq src, \forall u \neq dest \quad (3.5)$$

$$\sum_{w \in V} f(src, w) = f_{max}, \sum_{w \in V} f(w, dest) = f_{max} \quad (3.6)$$

Αφού όλοι οι υπολογισμοί είναι γραμμικοί, το σύστημα λύνεται με γραμμικό προγραμματισμό.

Στο παράδειγμα του σχήματος 3.3, έστω ότι η χωρική εργασία  $t_5$  έχει εντροπία τοποθεσίας  $Entropy(l_5)$ . Εφόσον βρίσκεται στις περιοχές των χρηστών  $w_1$  και  $w_2$ , το κόστος των δύο ακμών  $(u_1, u_8)$  και  $(u_2, u_8)$  ορίζεται ίσο με  $Entropy(l_5)$ . Αφού υπολογιστούν όλες οι τιμές των ακμών, δύναται να βρεθεί η μέγιστη ροή του γραφήματος 3.3b, εφαρμόζοντας οποιονδήποτε αλγόριθμο εύρεσης μέγιστης ροής [24].

### 3.5 Επίλογος

Στο παρόν κεφάλαιο γίνεται αναφορά στο γενικό πλαίσιο του spatial crowdsourcing. Βασικός διαχωρισμός είναι εάν η εφαρμογή προσφέρεται με υλική ανταμοιβή στους συμμετέχοντες ή απλά οι χρήστες συμμετέχουν σε αυτήν εθελοντικά. Το δεύτερο επίπεδο διαχωρισμού προσδιορίζει εάν οι εργασίες τις εφαρμογής επιλέγονται να εκτελεστούν από τους χρήστες, ή εάν ο κεντρικός εξυπηρετητής διαμοιράζει στο πλήθος τις εργασίες. Τέλος, γίνεται αναφορά εάν οι εργασίες εκτελούνται ανά ομάδες ή ατομικά.

Το βασικό πρόβλημα που ισχύει για όλες τις εφαρμογές crowdsourcing, είναι πώς μπορεί να επιτευχθεί η μέγιστη ανάθεση των εργασιών στο πλήθος. Για τον λόγο αυτό, έχουν αναπτυχθεί πρωτόκολλα, για την επίλυση του προβλήματος. Η μια στρατηγική που αναπτύχθηκε ονομάζεται Greedy. Έχει όμως ένα μειονέκτημα, δεν πραγματοποιεί ολική βελτιστοποίηση αλλά τοπική. Μια άλλη στρατηγική που ακολουθείται, κάνει χρήση της εντροπίας της τοποθεσίας. Αυτή θέτει σαν στόχο να υλοποιούνται πρώτα εργασίες που έχουν χαμηλή εντροπία, δηλαδή έχουν μικρότερη πιθανότητα να εκτελεστούν από το πλήθος.

Οι στρατηγικές αυτές, αναγάγουν το πρόβλημα επίτευξης μέγιστης ανάθεσης των εργασιών σε πρόβλημα εύρεσης της μέγιστης ροής. Έτσι, αφού κατασκευαστεί το γράφημα ροής  $G_i = (V, E)$ , με εφαρμογή γνωστών αλγορίθμων όπως είναι ο αλγόριθμος *Ford–Fulkerson*, υπάρχει η δυνατότητα να επιτευχθεί λύση του προβλήματος.

## Κεφάλαιο 4

# Η εξέλιξη και οι δυνατότητες των κινητών τηλεφώνων

### 4.1 Εισαγωγή

Τα κινητά τηλέφωνα τείνουν να καθιερωθούν ως το κύριο μέσο επικοινωνίας του πλήθους. Κανάλια διανομής εφαρμογών, όπως το AppStore της Apple, καθιστούν τα κινητά τηλέφωνα ικανά να ανακτούν εφαρμογές, μέσα σε σύντομο χρονικό διάστημα. Η τεχνολογική ανάπτυξη των κινητών συσκευών, παρουσιάζει ταχεία ανάπτυξη στην ανάπτυξη εφαρμογών και υπηρεσιών, οι οποίες βασίζονται σε πληροφορίες τοποθεσιών.

Στο παρόν κεφάλαιο γίνεται αναφορά στην τεχνολογική εξέλιξη των κινητών τηλεφώνων. Γίνεται επίσης, εκτενή αναφορά στους αισθητήρες που βρίσκονται εγκατεστημένοι στα τηλέφωνα και τις χρήσεις που παρέχουν. Το κινητό τηλέφωνο, σε συνδυασμό με το Global Positioning System (GPS) και τους υπόλοιπους αισθητήρες, μπορεί να λειτουργήσει σαν τοπικός σταθμός βάσης και να παρέχει υπηρεσίες που σχετίζονται με την γεωγραφική τοποθεσία.

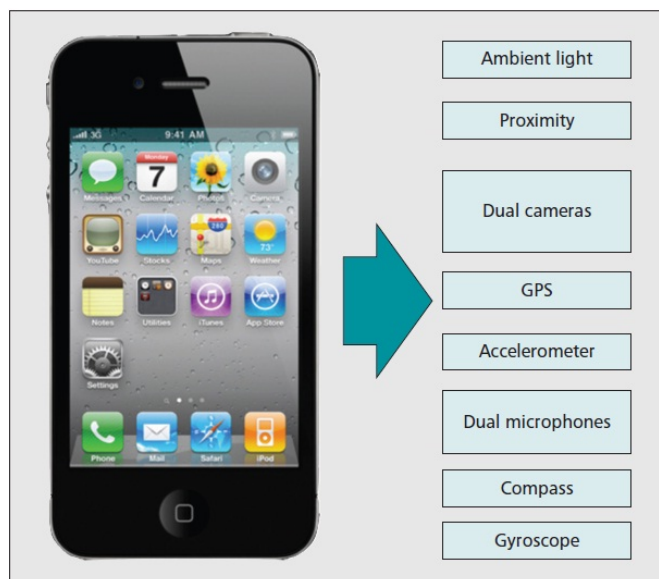
### 4.2 Εξέλιξη των Κινητών Τηλεφώνων

Η τεχνολογική εξέλιξη των τελευταίων χρόνων, έχει οδηγήσει στην κατασκευή νέων αισθητήρων. Η εγκατάστασή τους στα κινητά τηλέφωνα, δύναται να τα μετατρέψει σε υπολογιστική πλατφόρμα με πλούσια λειτουργικότητα.

Στο ΣΧΗΜΑ: 4.1, παρουσιάζεται η σουίτα των αισθητήρων που υπάρχουν στο κινητό τηλέφωνο iPhone 4 της εταιρίας Apple. Μέσω του συγκεκριμένου smartphone παρουσιάζονται οι κύριοι αισθητήρες που έχουν αναπτυχθεί και υπάρχουν σε όλα σχεδόν τα κινητά τηλέφωνα.

Τα σημερινά κινητά τηλέφωνα περιλαμβάνουν γυροσκόπιο, πυξίδα, επιταχυνσιόμετρο, αισθητήρα μέτρησης απόστασης, καθώς επίσης και αισθητήρα μέτρησης φωτισμού του περιβάλλοντος. Επίσης έχουν εγκατεστημένη εμπρός και πίσω κάμερα, μικρόφωνο, GPS, WiFi και

Bluetooth. Το σύνολο των νέων αυτών αισθητήρων, συνθέτει το περιβάλλον της εργασίας του χρήστη [27].



ΣΧΗΜΑ 4.1: Αισθητήρες που είναι εγκατεστημένα στο κινητό iPhone 4 της Apple [27].

Η βασική λειτουργία του *επιταχυνσιομέτρου* (accelerometer) είναι να ενισχύσει τη διεπαφή του χρήστη και τη χρήση της κάμερας. Χρησιμοποιείται για να προσδιοριστεί αυτόματα ο προσανατολισμός στον οποίο ο χρήστης κρατά το τηλέφωνο, ώστε να χρησιμοποιηθούν οι πληροφορίες αυτές εκ νέου, για τον προσανατολισμό της οθόνης μεταξύ ενός οριζόντιου ή κατακόρυφου προσανατολισμού.

Το *GPS* δίνει την δυνατότητα στο τηλέφωνο να αποκτήσει πρόσβαση σε γεωγραφικές συντεταγμένες και σε συνδυασμό με την *πιξίδα* και το *γυροσκόπιο*, το μετατρέπουν σε έναν τοπικό σταθμό βάσης (location-based station). Η ανάπτυξη των εν λόγω αισθητήρων, προσφέρει στο τηλέφωνο αυξημένη γνώση της θέσης που βρίσκεται, σε σχέση με τον φυσικό κόσμο (κατεύθυνση και προσανατολισμός), ενισχύοντας με αυτόν τον τρόπο την δημιουργία εφαρμογών.

Οι αισθητήρες, εκτός της λειτουργικότητας που προσφέρουν και της παροχής χωρικών πληροφοριών, αποτελούν ένα σημαντικό μέσο συγκέντρωσης στοιχείων των χρηστών των τηλεφώνων. Ο συνδυασμός των *δεδομένων του επιταχυνσιομέτρου* με την εκτίμηση της τοποθεσίας από τον αισθητήρα *GPS*, είναι ικανός να προσδιορίσει τις φυσικές κινήσεις του χρήστη. Άμεσα μπορούν να αξιοποιηθούν για να αναγνωριστούν ποικίλες δραστηριότητες, όπως για παράδειγμα εάν ο χρήστης κινείται αργά ή γρήγορα.

Η *κάμερα* και το *mikρόφωνο* αποτελούν τους πιο ισχυρούς και διαδεδομένους αισθητήρες. Με τη συλλογή ήχου από το μικρόφωνο του τηλεφώνου, υπάρχει η δυνατότητα να χαρακτηριστεί ένα σύνολο από διακριτούς ήχους που σχετίζονται με ένα συγκεκριμένο πλαίσιο ή δραστηριότητα. Η κάμερα του τηλεφώνου αντίστοιχα, μπορεί να χρησιμοποιηθεί για παραδοσιακές εργασίες όπως είναι το *blogging* φωτογραφιών. Σε πιο εξειδικευμένες δραστηριότητες τηλεπισκόπησης, μπορεί να λειτουργήσει για την παρακολούθηση της κίνησης των ματιών



του χρήστη σε όλη την οθόνη του τηλεφώνου ή να γίνει ενεργοποίηση εφαρμογών, κάνοντας χρήση της εμπρόσθιας κάμερας [27].

### 4.3 Πεδία και παροχή εφαρμογών

Οι crowdsourcing εφαρμογές, οι οποίες δύνανται να επεξεργαστούν δεδομένα από όλους τους τύπους των αισθητήρων, μπορούν να κατηγοριοποιηθούν ανάλογα με τις δυνατότητες που προσφέρουν. Οι τέσσερις πιο σημαντικοί τομείς είναι η *παρακολούθηση της οδικής κυκλοφορίας*, τα *κοινωνικά δίκτυα*, η *παρακολούθηση του περιβάλλοντος*, καθώς επίσης και η παροχή *υπηρεσιών υγείας*. Επιπλέον, ο σχεδιασμός όλων των εφαρμογών, καθιστά το πλήθος ως τον κύριο άξονα λειτουργίας και η ενεργητική συμμετοχή του αποτελεί επιτακτική ανάγκη. Λόγω της αυξημένης συμμετοχής, μπορούμε να τις κατηγοριοποιήσουμε είτε ως *ανθρωποκεντρικές* είτε ως *περιβαλλοντοκεντρικές* εφαρμογές.

Οι *ανθρωπο-κεντρικές* εφαρμογές, επικεντρώνονται κυρίως στην κατανόηση της συμπεριφοράς των ατόμων. Κάνουν χρήση των ενσωματωμένων αισθητήρων του τηλεφώνου, για να συλλέξουν δεδομένα που αφορούν τους χρήστες. Οι τομείς στους οποίους αναπτύχθηκαν τέτοιου είδους εφαρμογές αφορούν είτε την παροχή υπηρεσιών για την πρόληψη της υγείας των ατόμων, είτε για αθλητικές δραστηριότητες, είτε ακόμη και για κοινωνική δικτύωση μεταξύ των χρηστών.

Αντιθέτως, οι *περιβαλλοντο-κεντρικές* εφαρμογές, αναπτύχθηκαν για την συλλογή περιβαλλοντικών παραμέτρων, όπως για παράδειγμα για την μέτρηση της ποιότητας του αέρα του περιβάλλοντος. Οι τομείς στους οποίους αναπτύχθηκαν τέτοιου είδους εφαρμογές, αφορούν είτε την εύρεση της ποιότητας του περιβάλλοντος, είτε την εύρεση της έντασης πιθανού θορύβου, είτε τις συνθήκες της οδικής κυκλοφορίας [10].

*Φυσικά, για την σωστή λειτουργία και την επιτυχή χρήση των εφαρμογών, είναι απαραίτητα τα αναπτυσσόμενα δίκτυα διανομής εφαρμογών, τα οποία προσφέρονται από όλους τους μεγάλους προμηθευτές.*

#### Παρακολούθηση οδικής κυκλοφορίας

Η κυκλοφοριακή συμφόρηση ανέκαθεν αποτελεί ένα σοβαρό παγκόσμιο πρόβλημα. Τα τηλέφωνα κάνοντας χρήση του εγκατεστημένου GPS, δύνανται να συλλέγουν πληροφορίες σχετικά με την κατάσταση της κυκλοφορίας των οδικών αρτηριών. Με αυτόν τον τρόπο οι οδηγοί, για να φτάσουν στον επιθυμητό προορισμό τους, λαμβάνουν υπ' όψιν τις ενημερώσεις, και μπορούν να επιλέξουν μια συντομότερη και χωρίς εμπόδια διαδρομή.

Εφαρμογές που έχουν αναπτυχθεί για τον σκοπό αυτό, είναι το Mobile Millennium Project, το VTrack, το Nericell, καθώς και το GreenGPS, οι οποίες κάνουν πρόβλεψη της κυκλοφορίας σε

Type	Monitored subject	Application	Time	Location	Pictures	Sound	Acceleration	Pollution	Biometric data	Barometric data
People-centric sensing	User health	DietSense	x	x	x	x				
		Pediatric obesity	x	x	x		x		x	
		BALANCE	x				x		x	x
		Jog Falls	x		x		x		x	
		HealthSense	x				x			
		MobAsthma	x	x				x	x	
	Personal impact	SenSay	x	x			x			
		PEIR	x	x						
	Sport experiences	BikeNet	x	x		x	x	x	x	
		Biketastic	x	x		x	x			
		SkiScape	x	x		x	x			
	Social media	Micro-Blog	x	x	x	x	x		x	
		CenceMe	x	x	x	x	x			
	Price auditing	LiveCompare	x	x	x					
		PetrolWatch	x	x						
Environment-centric sensing	Air pollution	Haze Watch	x	x				x		
		PollutionSpy	x	x				x		
		Paulos et al.	x	x				x		
	Thermal columns	ikarus	x	x		x				x
		NoiseTube	x	x		x				
	Noise and ambiance	Ear-Phone	x	x		x				
		NoiseSpy	x	x		x				
		SoundSense	x	x		x				
		MoVi	x	x		x	x			
		MetroTrack	x	x		x				
		Nericell	x	x		x	x			
	Road conditions	Virtual Trip Lines	x	x						
		VTrack	x	x						
		Transit tracking	x	x				x		
		CarTel	x	x	x			x	x	
GreenGPS		x	x							

ΣΧΗΜΑ 4.2: Διαχωρισμός εφαρμογών ανά κατηγορίες [10].

πραγματικό χρόνο. Οι εφαρμογές κάνουν χρήση των αισθητήρων για εύρεση των γεωγραφικών πληροφοριών, και κατασκευάζουν εικονικούς χάρτες. Οι χρήστες επιλέγουν, βασιζόμενοι στις επεξεργασμένες πληροφορίες, την διαδρομή που είναι πιο προσιτή για αυτούς.

### Κοινωνική δικτύωση

Εκατομμύρια άνθρωποι συμμετέχουν τακτικά σε online κοινωνικά δίκτυα. Λόγου χάριν το Facebook, το Twitter ή το Foursquare είναι εφαρμογές που χρησιμοποιούνται καθημερινά. Μέσω της IP, η οποία ανατίθεται με βάση γεωγραφικά κριτήρια, ή του αισθητήρα GPS, μπορεί να προσδιοριστεί η τοποθεσία που βρίσκεται το άτομο που έχει υπό την κατοχή του το τηλέφωνο. Για παράδειγμα εάν η IP που έχει ανατεθεί, ανήκει σε δίκτυο πανεπιστημίου, τότε ο χρήστης είναι πιθανό να βρίσκεται εντός των ορίων του. Επιπλέον μέσω της κάμερας είναι δυνατή η λήψη φωτογραφιών και η διαμοίρασή τους σε όλους τους συνδεδεμένους χρήστες.

## Παρακολούθηση του Περιβάλλοντος

Συμβατικοί τρόποι μέτρησης και αναφοράς της περιβαλλοντικής ρύπανσης, βασίζονται σε συγκεντρωτικά στατιστικά στοιχεία που ισχύουν σε μια κοινότητα ή μια ολόκληρη πόλη. Οι αισθητήρες των τηλεφώνων προσφέρονται για τέτοιου είδους μετρήσεις, προσδιορίζοντας σε πραγματικό χρόνο τις συνθήκες που επικρατούν στο περιβάλλον. Μέσω εφαρμογών, όπως είναι το PEIR project [29], το πλήθος μπορεί να ρυθμίσει το κινητό του ώστε να λειτουργεί σαν τοπικός σταθμός λήψης μετρήσεων.

## Πρόληψη υγείας

Οι πληροφορίες που χρησιμοποιούνται για την προσωπική φροντίδα της υγείας σήμερα, προέρχονται σε μεγάλο βαθμό από έρευνες που πραγματοποιούνται σε προσωπικό επίπεδο, και πάντα με τις υποδείξεις των ειδικών ιατρών. Τα τηλέφωνα με αισθητήρες έχουν τη δυνατότητα να συλλέγουν δεδομένα, που μπορούν κυρίως να αλλάξουν ριζικά τον τρόπο που η υγεία και η ευεξία αξιολογούνται, αλλά να γίνει και εύρεση μεθόδων θεραπείας και φροντίδας. Εφαρμογές, όπως για παράδειγμα το UbiFit Garden, αποτυπώνουν τα επίπεδα της φυσικής δραστηριότητας, και συσχετίζουν τις πληροφορίες αυτές με προσωπικούς για την υγεία στόχους.

Στο κεφάλαιο 8 γίνεται εκτενή αναφορά στις εφαρμογές που αναπτύχθηκαν, για κάθε έναν τομέα ξεχωριστά.

## Δίκτυα διανομής εφαρμογών

Οι προγραμματιστές, τίθενται συνεχώς αντιμέτωποι με το πρόβλημα της διανομής των εφαρμογών στο πλήθος. Τα σύγχρονα τηλέφωνα έχουν ένα αποτελεσματικό κανάλι διανομής εφαρμογών. Πρώτη που το διέθεσε ήταν η Apple με το App Store για το τηλέφωνο τύπου iPhone. Κάθε προμηθευτής τηλεφώνων πλέον διαθέτει ένα App Store (Apple AppStore, Android Market, Microsoft Mobile Marketplace, Nokia Ovi).

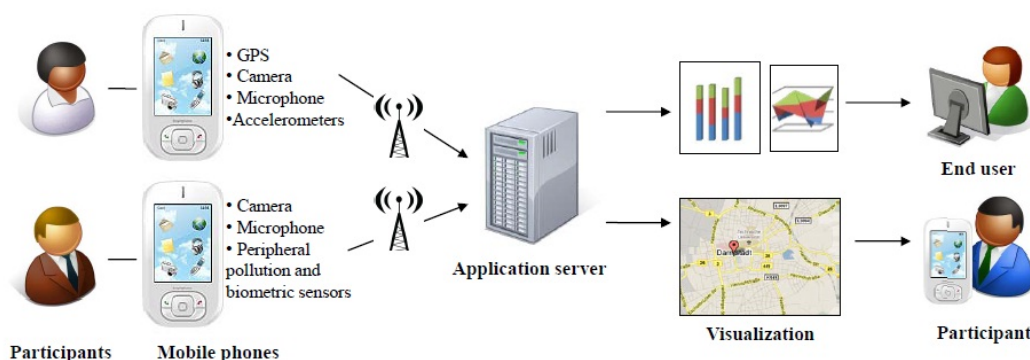
Τα δίκτυα διανομής εφαρμογών, κατέστησαν την δυνατότητα για εύκολη και γρήγορη χρήση των εφαρμογών από το πλήθος. Η ταχεία εγκατάστασή τους αποτελεί σημαντικό πλεονέκτημα, καθώς πλέον ένας μεγάλος αριθμός χρηστών παγκοσμίως, μπορεί να συμμετέχει, εθελοντικά ή μη, σε μελέτες ή πειράματα μόνο κάνοντας χρήση του τηλεφώνου.

Συνοψίζοντας, με τα νέου τύπου τηλέφωνα προσφέρεται μια σειρά από πλεονεκτήματα, που έχουν σαν αποτέλεσμα την ανάπτυξη ενός μεγάλου δικτύου από ασύρματες συσκευές, ιδιαίτερα στις αστικές περιοχές. Πρώτον, οι εγκατεστημένοι αισθητήρες που υπάρχουν στα τηλέφωνα, εξαλείφουν την ανάγκη επιπλέον εγκατάστασης αισθητήρων. Δεύτερον, τα τηλέφωνα προγραμματίζονται εύκολα. Για παράδειγμα τα τηλέφωνα με λειτουργικό σύστημα iOS ή Android, αναβαθμίζονται αυτόματα, κατόπιν επιτυχούς σύνδεσης με τον κεντρικό εξυπηρετητή

του παρόχου. Τρίτον, κάθε προμηθευτής τηλεφώνου προσφέρει ένα δίκτυο διανομής εφαρμογών, επιτρέποντας στους προγραμματιστές να προσφέρουν νέες εφαρμογές σε μεγάλους πληθυσμούς χρηστών παγκοσμίως. Με αυτόν τον τρόπο επιτρέπεται η συλλογή και ανάλυση των δεδομένων, πολύ πέρα από την κλίμακα του τι ήταν μέχρι πρότινος εφικτό. Τέλος, είναι πλέον δυνατός ο σχεδιασμός εφαρμογών που μπορούν να βελτιώσουν θεαματικά τη ζωή του πλήθους, αφού πρώτα συμπεριληφθεί στην διαδικασία επεξεργασίας των δεδομένων και διαπιστωθούν οι ανάγκες του [10, 27].

#### 4.4 Γενική Αρχιτεκτονική των Εφαρμογών

Η δομή μιας τυπικής crowdsourcing εφαρμογής παρουσιάζεται στο ΣΧΗΜΑ: 4.3. Τα δεδομένα των αισθητήρων, τα οποία συλλέγονται αυτόματα ή χειροκίνητα από τα τηλέφωνα των συμμετεχόντων, στέλνονται χρησιμοποιώντας ασύρματες επικοινωνίες δεδομένων, σε έναν κεντρικό εξυπηρετητή (application server) για επεξεργασία.



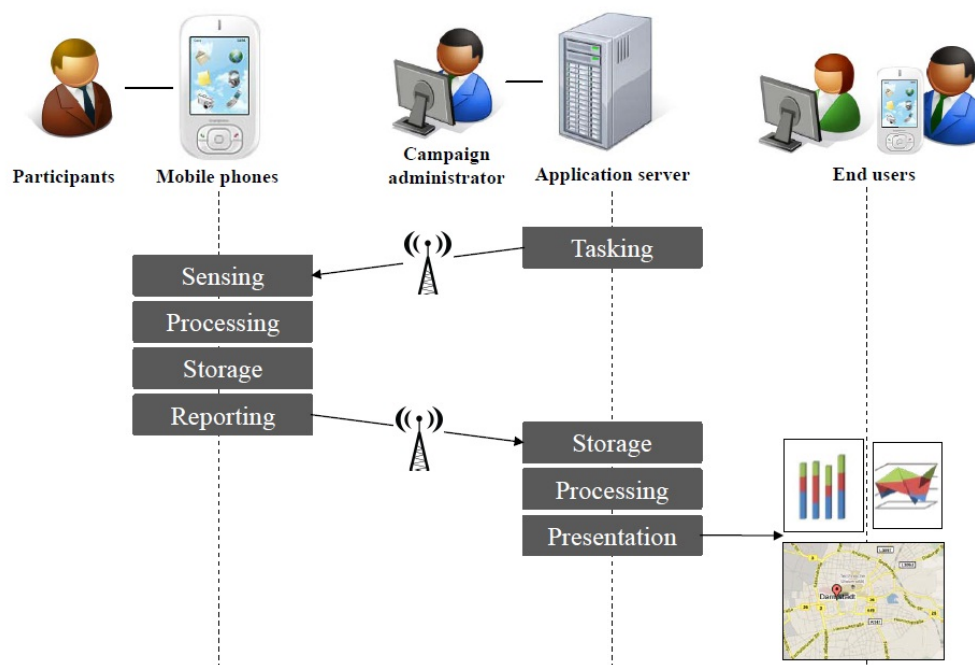
ΣΧΗΜΑ 4.3: Δομή της αρχιτεκτονικής μιας τυπικής crowdsourcing εφαρμογής [10].

Κατόπιν, στον εξυπηρετητή τα δεδομένα αναλύονται και διατίθενται σε διάφορες μορφές, όπως για παράδειγμα γραφικές αναπαραστάσεις ή χάρτες, απεικονίζοντας τα αποτελέσματα σε μεμονωμένους χρήστες ή σε ομάδες χρηστών. Επιπρόσθετα, τα αποτελέσματα δύναται να εμφανίζονται στα κινητά τηλέφωνα των συμμετεχόντων ή να προσπελαύνονται από το ευρύτερο κοινό μέσω διαδικτυακών πυλών, ανάλογα με τον σχεδιασμό της εφαρμογής [10].

Το αναλυτικό μοντέλο του συστήματος των crowdsourcing εφαρμογών, απεικονίζεται στο ΣΧΗΜΑ: 4.4, όπου παρουσιάζεται ο ρόλος των ενδιαφερομένων μερών, και των αντίστοιχων βασικών λειτουργιών των υποδομών της εφαρμογής [10].

##### 4.4.1 Συμμετέχοντες

Από το ΣΧΗΜΑ: 4.4 γίνεται προσδιορισμός των ακόλουθων ενδιαφερομένων μερών:



ΣΧΗΜΑ 4.4: Απεικόνιση συμμετεχόντων και βασικών λειτουργιών των υποδομών μιας crowdsourcing εφαρμογής [10].

- Campaign administrators:** Τα άτομα αυτά είναι υπεύθυνα για τον σχεδιασμό, την υλοποίηση και την ανάπτυξη της αρχιτεκτονικής του συστήματος των εφαρμογών. Είναι αρμόδιοι για τη συντήρηση και τη διαχείριση των υποδομών. Για παράδειγμα, μπορούν να διαθέσουν την εφαρμογή για λήψη από δικτυακούς τόπους ή από δίκτυα διανομής εφαρμογών, καθώς επίσης να ρυθμίσουν τον κεντρικό εξυπηρετητή, για συλλογή και επεξεργασία των δεδομένων.
- Participants:** Οι συμμετέχοντες πραγματοποιούν εγκατάσταση της εκάστοτε εφαρμογής στο τηλέφωνό τους, και συλλέγουν μετρήσεις μέσω των διαθέσιμων αισθητήρων. Η συνεισφορά τους στις εφαρμογές αποτελεί βασικό παράγοντα επιτυχούς λειτουργίας. Ο βαθμός στον οποίο εμπλέκονται οι συμμετέχοντες στην εφαρμογή, μπορεί να υποκινηθεί από διάφορους παράγοντες. Για παράδειγμα, υπάρχει το κίνητρο να επωφεληθούν από τα δεδομένα που παρέχονται από τους ίδιους και τους άλλους συμμετέχοντες.
- End users:** Οι τελικοί χρήστες έχουν πρόσβαση και συμβουλεύονται τα επεξεργασμένα δεδομένα. Αφού γίνει η συλλογή των δεδομένων από τους συμμετέχοντες, και υποστούν επεξεργασία από τον κεντρικό εξυπηρετητή, διαμοιράζονται ανάλογα με τα ενδιαφέροντα και τις προτιμήσεις των χρηστών. Για παράδειγμα, υπάρχει ενδεχόμενο οι οδηγοί να θέλουν να ενημερώνονται για την κυκλοφορία, ή επαγγελματίες γιατροί να ελέγχουν δεδομένα των ασθενών τους. Οι τελικοί χρήστες μπορεί να συμμετέχουν ενεργά στην εφαρμογή, μπορεί όμως και να μην λαμβάνουν μέρος [10].

#### 4.4.2 Κύριες λειτουργίες των υποδομών των εφαρμογών

Στο ΣΧΗΜΑ: 4.4 παρουσιάζονται οι βασικές λειτουργίες, που συνθέτουν το γενικό μοντέλο του συστήματος των crowdsourcing εφαρμογών. Οι λειτουργίες οργανώνονται σε μια αρχιτεκτονική πελάτη – εξυπηρετητή. Αφού πρώτα πραγματοποιηθεί η λήψη των δεδομένων από τους αισθητήρες, στην συνέχεια παρουσιάζονται τα αποτελέσματα στους τελικούς χρήστες.

- **Λειτουργία Sensing:** Η λειτουργία ανίχνευσης εντοπίζεται στα τηλέφωνα των συμμετεχόντων. Βασικός ρόλος της είναι να λαμβάνει, και να αποθηκεύει διαφορετικά είδη δεδομένων από τους αισθητήρες. Πιο συνηθισμένα δεδομένα είναι η ώρα, η γεωγραφική τοποθεσία, οι φωτογραφίες, τα ηχητικά δείγματα, τα δεδομένα του επιταχυνσιόμετρου, τα δεδομένα του επιπέδου της μόλυνσης του περιβάλλοντος, ή τα βιομετρικά δεδομένα. *Τα δεδομένα του αισθητήρα, μπορεί να ληφθούν σύμφωνα με μια από τις ακόλουθες μεθόδους ανίχνευσης. Είτε χειροκίνητα, είτε αυτόματα, είτε καιροσκοπικά.*

Στη χειροκίνητη μέθοδο, οι συμμετέχοντες συμμετέχουν άμεσα στη διαδικασία λήψης δεδομένων, ενεργοποιώντας οι ίδιοι τη συλλογή μετρήσεων από τους αισθητήρες. Για παράδειγμα, όταν βρεθούν σε κυκλοφοριακή συμφόρηση, στέλνουν τα δεδομένα της κίνησης, της ταχύτητας, και της τοποθεσίας που βρίσκονται. Αντίθετα, στην αυτόματη και στην καιροσκοπική μέθοδο, οι συμμετέχοντες δεν εμπλέκονται άμεσα στην διαδικασία. Οι αισθητήρες συλλέγουν σε μια σταθερή συχνότητα δειγματοληψίας, και για αυτό είναι επίσης γνωστή και ως συνεχή μέθοδο. Στην καιροσκοπική λειτουργία οι εγκατεστημένοι αισθητήρες, παρακολουθούν το περιβάλλον και η εφαρμογή ενεργοποιεί τη λειτουργία των αισθητήρων όταν σημειώνεται υπέρβαση ενός κατώτατου ορίου, που έχει αρχικά οριστεί.

- **Λειτουργία Tasking:** Υποστηρίζει την λειτουργία της *ανίχνευσης*, διανέμοντας τις εργασίες στα τηλέφωνα. Οι εν λόγω εργασίες προσδιορίζονται από λεπτομέρειες και απαιτήσεις, που βασίζονται στην λειτουργία και το είδος της εκάστοτε εφαρμογής, συμπεριλαμβάνοντας κριτήρια όπως είναι η έναρξη της συλλογής των δεδομένων ή η τιμή της συχνότητας δειγματοληψίας. Ένα παράδειγμα αυτών των απαιτήσεων θα μπορούσε να είναι ότι τα τηλέφωνα πρέπει να είναι εξοπλισμένα με αισθητήρα GPS, και με ενσωματωμένες κάμερες που μπορεί να καταγράψουν έως 3 Megapixel εικόνες. Οι εργασίες περιέχουν επίσης πληροφορίες σχετικά με τη θέση ή το χρονικό διάστημα λήψης των δεδομένων.
- **Λειτουργία Reporting:** Μέσω της λειτουργίας αναφοράς, διασφαλίζεται η μετάδοση των ενδείξεων που συλλέγονται από τους αισθητήρες προς τον εξυπηρετητή. Για την μεταφορά των δεδομένων, κυρίως γίνεται χρήση των υποδομών του δικτύου επικοινωνίας που διαθέτει το τηλέφωνο, όπως είναι το ασύρματα δίκτυο ή η GSM/GPRS/3G/4G συνδεσιμότητα. Για παράδειγμα, οι μετρήσεις των αισθητήρων μπορεί να μεταδοθούν στον εξυπηρετητή με χρήση SMS ή TCP συνδέσεων.

- **Λειτουργία Storage:** Η λειτουργία αυτή είναι βασική στην αρχιτεκτονική, διότι διασφαλίζεται η αποθήκευση των δεδομένων που συλλέγονται από τους αισθητήρες των τηλεφώνων, καθώς και των δεδομένων του εξυπηρετητή. Αυτό συμβαίνει λόγω του ότι ο εξυπηρετητής παρέχει μακροχρόνια αποθήκευση των δεδομένων, ενώ τα τηλέφωνα εξασφαλίζουν βραχυπρόθεσμη αποθήκευση των δεδομένων τα οποία υποβάλλονται για επεξεργασία ή μετάδοση στον εξυπηρετητή. Όσον αφορά στην λειτουργία του εξυπηρετητή, τα δεδομένα αποθηκεύονται συνήθως σε σχεσιακές βάσεις δεδομένων ή βάσεις δεδομένων ειδικά προσαρμοσμένες για τη διαχείριση των μετρήσεων των αισθητήρων.
- **Λειτουργία Processing:** Η παρούσα λειτουργία επεξεργάζεται τα δεδομένα που συλλέχθηκαν από τους αισθητήρες, είτε απευθείας στα τηλέφωνα σε μερική κλίμακα, είτε στον εξυπηρετητή σε ευρύτερη κλίμακα. Η λειτουργία αυτή αναλύει τα δεδομένα που εστάλησαν στον εξυπηρετητή, και τα ετοιμάζει για να γίνει η παρουσίασή τους.
- **Λειτουργία Presentation:** Μέσω αυτής της λειτουργίας πραγματοποιείται η παρουσίαση των αποτελεσμάτων που προέκυψαν από την επεξεργασία, στους τελικούς χρήστες. Τα αποτελέσματα εμφανίζονται είτε τοπικά στα τηλέφωνα των συμμετεχόντων, είτε παρουσιάζονται μέσω διαδικτυακών πυλών σε ένα ευρύτερο κοινό. Τα αποτελέσματα παρουσιάζονται επίσης και σε φόρμες ακατέργαστων δεδομένων, σε μορφές γραφημάτων, χαρτών, ή γεωγραφικών επικαλύψεων για να επιτραπεί στους τελικούς χρήστες να αναλύσουν οι ίδιοι τα δεδομένα.

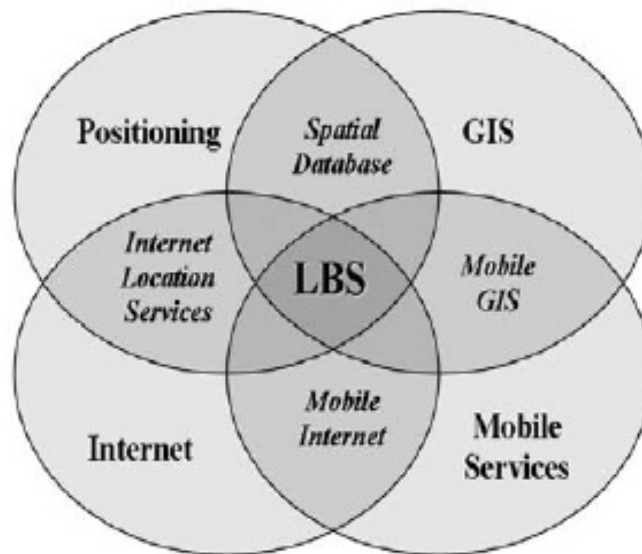
#### 4.5 Location-Based Υπηρεσίες

Η γεωγραφική θέση του χρήστη είναι δυνατόν να συνδυαστεί με επιπλέον πληροφορίες, έτσι ώστε να επιτευχθεί ο συνδυασμός ενός συστήματος Location-based υπηρεσιών με το Σύστημα Γεωγραφικών Πληροφοριών, ή οποιοδήποτε άλλο σύστημα πληροφοριών θέσης.

Οι Location-Based υπηρεσίες όπως απεικονίζονται στο ΣΧΗΜΑ: 4.5, μπορούν να θεωρηθούν ως η σύγκλιση:

- των υπηρεσιών της κινητής τηλεφωνίας,
- των τεχνολογιών γνώσης της τοποθεσίας που βρίσκεται ο χρήστης,
- του Διαδικτύου,
- και του Γεωγραφικού Συστήματος Πληροφοριών (GIS).

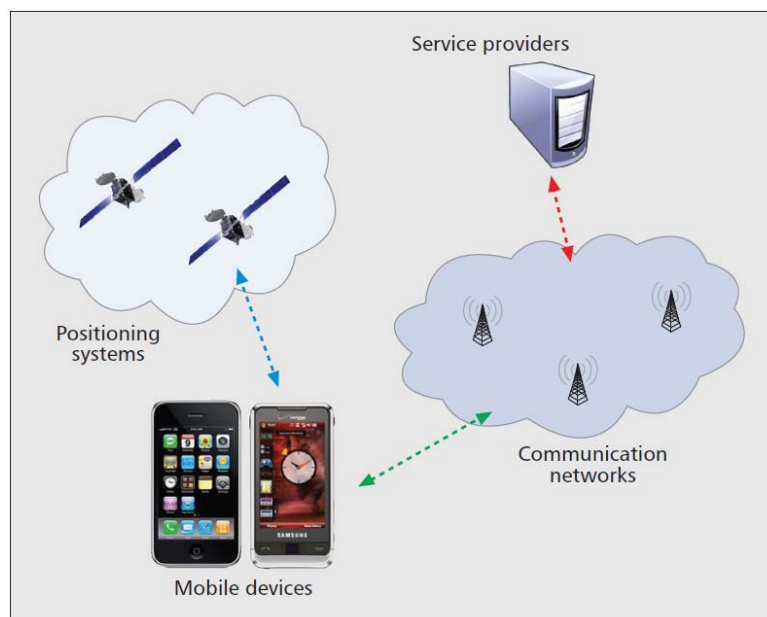
Τα συστήματα που προσφέρουν LBS's, πρέπει να μπορούν να επεξεργάζονται ως σύνολο τις πληροφορίες και τα δεδομένα που σχετίζονται με τη γεωγραφική θέση, τη δρομολόγηση, την αναζήτηση, το περιεχόμενο που έχουν τα πολυμέσα και τη θέση διεύθυνσης. [8].



ΣΧΗΜΑ 4.5: Σύγκλιση τεχνολογιών για την υποστήριξη Location-Based Υπηρεσιών [8].

#### 4.5.1 Αρχιτεκτονική των Location-Based Υπηρεσιών

Η δομή της αρχιτεκτονικής των LBS's απεικονίζεται στο ΣΧΗΜΑ: 4.6.



ΣΧΗΜΑ 4.6: Δομή της αρχιτεκτονικής των location-based υπηρεσιών [23].

Τα κινητά τηλέφωνα χρησιμοποιούνται για την αποστολή ερωτημάτων στους LBS εξυπηρετητές. Η γεωγραφική τοποθεσία λαμβάνεται μέσω των συστημάτων εντοπισμού θέσης, όπως είναι το GPS. Τα ερωτήματα, καθώς και οι απαντήσεις από τους κεντρικούς διακομιστές, μεταδίδονται μέσω των δικτύων επικοινωνίας, όπως είναι τα τρίτης γενιάς (3G) δίκτυα ή μέσω WiFi. Οι LBS εξυπηρετητές αποτελούν φορείς παροχής υπηρεσιών, απαντώντας στα ερωτήματα με



καλώς προσαρμοσμένες απαντήσεις, με βάση τις πληροφορίες τοποθεσίας των ερωτημάτων [23].

#### 4.5.2 Προσδιορισμός της τοποθεσίας των χρηστών

Οι LBS's μπορούν να προσδιορίσουν την γεωγραφική τοποθεσία, κάνοντας χρήση μιας από τις ακόλουθες υπηρεσίες:

- **Αναγνώριση μέσω των πύργων αναμετάδοσης:** Τα τηλέφωνα μπορούν να προσδιορίσουν τη θέση που βρίσκονται, βασιζόμενα στους πύργους αναμετάδοσης των δικτύων κινητής τηλεφωνίας. Εφόσον βρίσκονται σε κοντινή απόσταση με αυτούς, έχουν την δυνατότητα να παρέχουν πληροφορίες στις LBS's. Οι πληροφορίες έχουν ακρίβεια έως 100 μέτρα, και όσο πιο εξελιγμένοι είναι οι πύργοι αναμετάδοσης, τόσο περισσότερη ακρίβεια μπορεί να επιτευχθεί.
- **Μέσω του Γεωγραφικού Συστήματος Πληροφοριών:** Το GPS δίνει την δυνατότητα στις συσκευές να λαμβάνουν σήματα από το δίκτυο των δορυφόρων, ώστε να καθορίζεται η τοποθεσία τους. Οι πληροφορίες εντοπισμού πραγματοποιούνται με ακρίβεια έως 20 μέτρων, οι οποίες μπορούν να τοποθετήσουν τη συσκευή σε μια συγκεκριμένη θέση.
- **Μέσω της σύγκλισης των WiFi δικτύων:** Ορισμένες συσκευές προσδιορίζουν τη τοποθεσία που βρίσκονται, μελετώντας τα σήματα των κοντινών ασύρματων δικτύων. Τα σήματα αυτά γίνονται σύγκριση από μια λίστα με γνωστά σημεία ασύρματης πρόσβασης. Ο τρόπος αυτός είναι ακριβής σε απόσταση έως 200 μέτρων.
- **Μέσω του IP πρωτοκόλλου:** Οποιαδήποτε ιστοσελίδα ή διαδικτυακή υπηρεσία μπορεί, κατά προσέγγιση, να εντοπίσει τη θέση μιας συσκευής με βάση τη διεύθυνση IP. Οι διευθύνσεις IP μπορούν να προσδιορίσουν μια δεδομένη περιοχή, ενώ συγκεκριμένες διευθύνσεις IP μπορεί να προσδιορίζουν συγκεκριμένες τοποθεσίες.
- **Μέσω παροχής πληροφοριών από τον χρήστη:** Οι LBS's μπορούν απλά να ζητήσουν από τον χρήστη να παράσχει χειροκίνητα την τρέχουσα τοποθεσία του. Η ακρίβεια της μεθόδου αυτής εξαρτάται από την υπηρεσία και τον συμμετέχοντα [34].

#### 4.6 Επίλογος

Η τεχνολογική εξέλιξη επιφέρει στις συσκευές δυνατότητες, που τα προηγούμενα χρόνια δεν μπορούσαν να έχουν. Μέσω των αισθητήρων που διατίθενται, το κινητό τηλέφωνο μπορεί να αποκτήσει πρόσβαση σε χωρικά δεδομένα και γεωγραφικές συντεταγμένες. Οι προγραμματιστές αφού υλοποιήσουν εφαρμογές, τις διαμοιράζουν μέσω των δικτύων διανομής

εφαρμογών. Τα τηλέφωνα αποκτούν αυτόματα την λειτουργία ενός τοπικού σταθμού εκπομπής δεδομένων, και το πλήθος με τις γνώσεις που έχει δύναται να αποκτήσει σημαντικές πληροφορίες.

Στο επόμενο κεφάλαιο αναπτύσσεται το θέμα της ασφάλειας, και πώς τα δεδομένα των χρηστών διαφυλάσσονται.

## Κεφάλαιο 5

# Ασφάλεια προσωπικών δεδομένων

### 5.1 Εισαγωγή

Η δυνατότητα εντοπισμού της γεωγραφικής θέσης από τις φορητές συσκευές, αποτέλεσε το έναυσμα για ανάπτυξη πολλών εφαρμογών. Ωστόσο έχει εγείρει σοβαρές ανησυχίες σχετικά με τους κινδύνους της αποκάλυψης ευαίσθητων πληροφοριών και δεδομένων. [14]. Η ανωνυμία μπορεί να προσφέρει έναν υψηλό βαθμό προστασίας της ιδιωτικής ζωής του πλήθους, και να μειώσει τις απαιτήσεις για διαφύλαξη ιδιωτικών πληροφοριών, από τους φορείς παροχής υπηρεσιών. Ενώ η ανωνυμία ετυμολογικά σημαίνει “χωρίς ταυτότητα”, οι ερευνητές του απορρήτου των πληροφοριών, την ερμηνεύουν ως μια ισχυρότερη έννοια [16]. Σύμφωνα με τους *Pfitzmann* και *Koehntopp*,

*“Η ανωνυμία είναι η κατάσταση, η οποία δεν είναι αναγνωρίσιμη μέσα σε ένα σύνολο γεγονότων”*

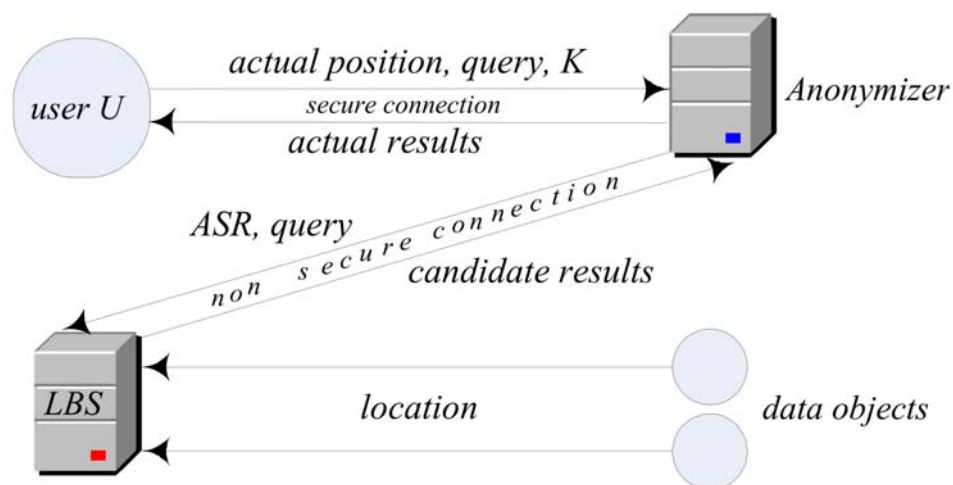
ANDREAS PFITZMANN & MARIT KOEHNTOPP

Για την προστασία των δεδομένων, και της γεωγραφικής τοποθεσίας που βρίσκονται οι συμμετέχοντες, έχουν αναπτυχθεί μηχανισμοί που στηρίζονται στην τεχνολογία *k-anonymity* [36]. Με την τεχνολογία αυτή, οι αναφορές που γίνονται upload στους εξυπηρετητές, πραγματοποιούνται με ασφάλεια. Η βασική ιδέα της *k-anonymity*, είναι η δημιουργία ομάδων από *k* συμμετέχοντες ή *k* αναφορές, ώστε να μοιράζονται ένα κοινό χαρακτηριστικό. Για παράδειγμα, για *k* συμμετέχοντες το κοινό χαρακτηριστικό μπορεί να είναι η κοινή τους τοποθεσία [10].

Στο παρόν κεφάλαιο, γίνεται αναφορά στην δομή της αρχιτεκτονικής που αξιοποιεί την τεχνολογία της *k-anonymity*, προκειμένου να προστατευθεί η ταυτότητα των χρηστών από τυχόν υποκλοπές δεδομένων. Κύρια λειτουργία της είναι η αντικατάσταση της ακριβούς τοποθεσίας ενός χρήστη *U*, με μια ανώνυμη χωρική τοποθεσία *anonymizing spatial region (ASR)*, στην οποία βρίσκονται τουλάχιστον *k-1* άλλοι χρήστες.

## 5.2 Δομή της Αρχιτεκτονικής για Επίτευξη Ανωνυμίας

Για να επιτευχθεί η ανωνυμία των χρηστών και να προστατευτούν τα προσωπικά δεδομένα, γίνεται χρήση του πλαισίου του ΣΧΗΜΑΤΟΣ: 5.1. [14, 31].



ΣΧΗΜΑ 5.1: Δομή της αρχιτεκτονικής των συστημάτων για επίτευξη ανωνυμίας [14].

Ο χρήστης  $U$  στέλνει το ερώτημά του και τον βαθμό της ανωνυμίας  $K$ , μέσω ασφαλούς σύνδεσης στον *anonymizer server*, ο οποίος διατηρεί τις τρέχουσες θέσεις όλων των χρηστών, και είναι ένας αξιόπιστος εξυπηρετητής.

Ο *anonymizer server* αφαιρεί το αναγνωριστικό του χρήστη, και επιλέγει ένα σύνολο ανωνυμίας (AS) το οποίο περιέχει τον χρήστη  $U$  και τουλάχιστον  $K - 1$  άλλους χρήστες από τις κοντινές περιοχές. Η  $K$ -*anonymizing spatial region* ( $K$ -ASR ή ASR) είναι μια περιοχή που περικλείει πολλά σύνολα ανωνυμίας.

Ο *anonymizer server* προωθεί την ASR περιοχή στους εξυπηρετητές που εκτελούν location-based υπηρεσίες, και είναι αποθηκευμένα τα δεδομένα.

Οι εξυπηρετητές επεξεργάζονται το ερώτημα, και επιστρέφουν στον *anonymizer server* ένα σύνολο υποψηφίων αποτελεσμάτων.

Ο *anonymizer server* αφαιρεί τα εσφαλμένα αποτελέσματα, και προωθεί το πραγματικό αποτέλεσμα στον χρήστη  $U$ .

Μέσω αυτής της αρχιτεκτονικής, οι χρήστες δεν επικοινωνούν με τις εφαρμογές άμεσα, με συνέπεια να μην αποκαλύπτουν την ταυτότητά τους. Ο *anonymizing server* ενεργεί ως ο διαμεσολαβητής για κάθε είδους επικοινωνία μεταξύ των χρηστών και των εφαρμογών.

### 5.3 Πληροφορίες Θέσεως μέσω της έννοιας $k$ —anonymity

Οι Gruteser και Grunwald [16] παρουσιάζουν στην επιστημονική κοινότητα την έννοια της  $k$ —anonymity. Ένας χρήστης θεωρείται ως *anonymous*, αν και μόνο αν οι πληροφορίες τοποθεσίας που παρουσιάζονται, δεν διαφέρουν από τις πληροφορίες τοποθεσίας τουλάχιστον  $k - 1$  άλλων χρηστών.

Πιο συγκεκριμένα, οι πληροφορίες τοποθεσίας παρίστανται από μια πλειάδα που περιέχει τρία διαστήματα  $([x_1, x_2], [y_1, y_2], [t_1, t_2])$ . Τα διαστήματα  $[x_1, x_2]$  και  $[y_1, y_2]$  περιγράφουν έναν δισδιάστατο χώρο στον οποίο βρίσκεται ο χρήστης. Το διάστημα  $[t_1, t_2]$  περιγράφει την χρονική περίοδο κατά την οποία ο χρήστης ήταν παρόν στον χώρο αυτόν. Να σημειωθεί ότι, τα διαστήματα αυτά αντιπροσωπεύουν περιοχές αβεβαιότητας, διότι γνωρίζουμε μόνο ότι ο χρήστης ήταν παρόν σε κάποιο σημείο του χώρου για κάποια χρονική στιγμή. Επομένως, μια πλειάδα τοποθεσίας για έναν χρήστη είναι  $k$ —*anonymous*, όταν περιγράφει τη θέση του χρήστη και τις θέσεις των  $k - 1$  άλλων χρηστών. Δηλαδή οι  $k - 1$  άλλοι χρήστες, θα πρέπει να ήταν παρόντες στην περιοχή, καθώς επίσης και το χρονικό διάστημα να περιγράφεται από την πλειάδα του χρόνου [16, 23].

*Όσο μεγαλύτερο είναι το σύνολο ανωνυμίας  $k$  που προσφέρεται, τόσο μεγαλύτερος είναι και ο βαθμός της ανωνυμίας.*

Στις επόμενες δύο υπο-ενότητες (5.3.1, 5.3.2), ακολουθεί περιγραφή δύο αλγορίθμων — μεθόδων, οι οποίοι προσφέρουν ανωνυμία.

#### 5.3.1 Αλγόριθμος Adaptive-Interval

Ένας δεδομένος βαθμός ανωνυμίας μπορεί να διατηρείται σε οποιαδήποτε θέση, ανεξάρτητα από τον αριθμό του πληθυσμού, μειώνοντας με αυτόν τον τρόπο τα ακριβή αποκαλυφθέντα χωρικά δεδομένα. Ο adaptive-interval αλγόριθμος επιλέγει μια επαρκώς μεγάλη περιοχή, έτσι ώστε να υπάρχουν αρκετοί χρήστες.

Ο επιθυμητός βαθμός της ανωνυμίας καθορίζεται από την παράμετρο  $k_{min}$  η οποία, είναι το ελάχιστο αποδεκτό μέγεθος του συνόλου της ανωνυμίας. Ο adaptive-interval αλγόριθμος προσδιορίζει μια επαρκώς μεγάλη περιοχή για μια δεδομένη τιμή  $k_{min}$ . Ο αλγόριθμος περιγράφεται με περισσότερες λεπτομέρειες στο ΣΧΗΜΑ: 5.2.

Ο αλγόριθμος λαμβάνει ως είσοδο την θέση που βρίσκεται ο χρήστης που ζητάει πληροφορίες, τις γεωγραφικές συντεταγμένες της περιοχής που καλύπτει ο *anonymizer server*, και τις θέσεις όλων των άλλων χρηστών της ευρύτερης περιοχής. Πραγματοποιείται διαίρεση της περιοχής, πέριξ της τοποθεσίας όπου βρίσκεται ο χρήστης, έως ότου ο αριθμός των χρηστών της περιοχής γίνει χαμηλότερος από την σταθερά  $k_{min}$ . Η περιοχή η οποία εξακολουθεί να ικανοποιεί την σταθερά  $k_{min}$ , στη συνέχεια επιστρέφεται. Η έξοδος του αλγορίθμου είναι οι πληροφορίες θέσης, οι οποίες στέλνονται στον εξυπηρετητή που εκτελεί τις location-based υπηρεσίες [16, 23].

1. Initialize the quadrants  $q$  and  $q_{prev}$  as the total area covered by the anonymizer
2. Initialize a traffic vector with the current positions of all known users
3. Initialize  $p$  as the position of requestor user
4. If the number of users in traffic vector  $< k_{min}$ , then return the previous quadrant  $q_{prev}$
5. Divide  $q$  into quadrants of equal size
6. Set  $q_{prev}$  to  $q$
7. Set  $q$  to the quadrant that includes  $p$
8. Remove all users outside  $q$  from the traffic vector
9. Repeat from Step 2

ΣΧΗΜΑ 5.2: Ο adaptive-interval αλγόριθμος υπολογίζει μια περιοχή, περιλαμβάνοντας τον χρήστη  $U$  και αρκετούς ακόμα χρήστες, έτσι ώστε να ικανοποιηθεί ο βαθμός της ανωνυμίας  $k_{min}$  [16].

Μια εναλλακτική προσέγγιση του αλγορίθμου είναι η παράλληλη μείωση της ακρίβειας του χρόνου. Δηλαδή να πραγματοποιείται καθυστέρηση της αίτησης του χρήστη, μέχρι να έχουν επισκεφθεί την τοποθεσία έως και  $k_{min}$  άλλοι χρήστες. Ο αλγόριθμος του ΣΧΗΜΑΤΟΣ: 5.2 τροποποιείται κατάλληλα, ώστε να λάβει και αυτήν την παράμετρο ανάλυσης ως είσοδο. Επίσης, καθορίζεται και η περιοχή παρακολούθησης, διαιρώντας το χώρο μέχρι η καθορισμένη ανάλυση να επιτευχθεί. Ο αλγόριθμος παρακολουθεί τις κινήσεις των χρηστών σε όλη αυτή την περιοχή. Όταν υπάρχουν  $k_{min}$  διαφορετικοί χρήστες, οι οποίοι έχουν επισκεφθεί την τοποθεσία, ένα χρονικό διάστημα  $[t_1, t_2]$  υπολογίζεται ως εξής: ο χρόνος  $t_1$  ισούνται με την στιγμή της αίτησης αφαιρώντας ένα τυχαίο παράγοντα απόκρυψης, και ο χρόνος  $t_2$  με την τρέχουσα ώρα [16].

Η πιθανότητα υποκλοπής των δεδομένων και η αποκάλυψη της τοποθεσίας του χρήστη  $U$  ισούται με  $\frac{1}{k}$ . Για να λειτουργήσει σωστά ο αλγόριθμος, βασική προϋπόθεση είναι η ύπαρξη τουλάχιστον δύο χρηστών. Το μειονέκτημα που εντοπίζεται αφορά το σύνολο ανωνυμίας. Όσο μικρότερος είναι ο αριθμός χρηστών που υπάρχει στην περιοχή, τόσο μεγαλύτερος είναι ο κίνδυνος να μην προφυλαχθούν σωστά τα δεδομένα. Επιπλέον, υπάρχει και το πρόβλημα της επικάλυψης του χρόνου ή του διαστήματος χώρου που βρίσκονται οι χρήστες. Κατά την τμηματοποίηση, υπάρχει ο κίνδυνος οποιουδήποτε χρήστη, να βρεθεί να μην ανήκει σε κάποιο σύνολο ανωνυμίας.

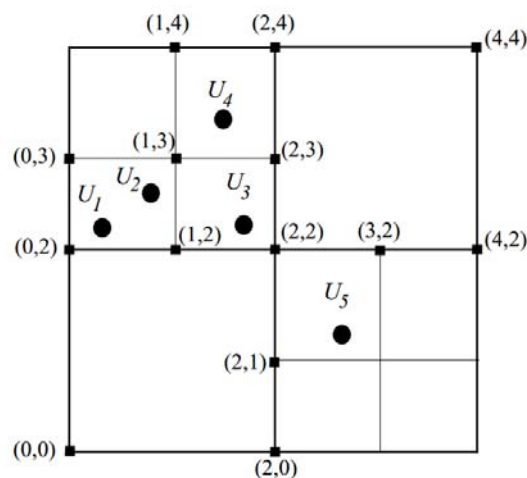
### 5.3.2 Μέθοδος Casper

Το Casper είναι ένα νέο πλαίσιο το οποίο, για την επεξεργασία των αιτημάτων, κάνει χρήση των υπηρεσιών τοποθεσίας διαφυλάσσοντας την ιδιωτική ζωή των χρηστών [31]. Ο **anonymizer server** διατηρεί τις θέσεις των χρηστών, χρησιμοποιώντας μια δομή δεδομένων σε μορφή πυραμίδας, παρόμοια με την Quad-tree, όπου το ελάχιστο μέγεθος του κελιού – τοποθεσίας, αντιστοιχεί στην ανάλυση της ανωνυμίας. Μόλις ο anonymizer server λάβει ένα ερώτημα από τον χρήστη  $U$ , χρησιμοποιεί έναν hash table για το αναγνωριστικό του χρήστη, δείχνοντας το

κελί  $c$  του χαμηλότερου επιπέδου που βρίσκεται ο  $U$ . Εάν το κελί  $c$  περιέχει αρκετούς χρήστες ( $|c| \leq K$ ) για τις απαιτήσεις της ανωνυμίας, διαμορφώνεται η ASR. Σε αντίθετη περίπτωση, εάν  $|c| > K$ , τότε γίνεται ανάκτηση των γειτονικών κελιών του  $c$ , δηλαδή του οριζοντίου  $c_h$  και του κάθετου  $c_v$ .

Εάν  $|c \cup c_h| \leq K$  ή  $|c \cup c_v| \leq K$ , η ένωση των κελιών γίνεται η ASR περιοχή. Αν και οι δύο ενώσεις περιέχουν τουλάχιστον  $K$  χρήστες, η ASR είναι αυτή με τους ελάχιστους χρήστες. Αντίθετα, εάν  $|c \cup c_h| > K$  και  $|c \cup c_v| > K$ , ο *anonymizer server* ανακτά το αρχικό κελί  $c$  και επαναλαμβάνει τη διαδικασία αυτή αναδρομικά [14, 16, 23].

Ένα παράδειγμα του μοντέλου Casper παρουσιάζεται με χρήση του ΣΧΗΜΑΤΟΣ: 5.3.



ΣΧΗΜΑ 5.3: Ανωνυμία με την μέθοδο Casper [14].

Τα κελιά δηλώνονται από τις συντεταγμένες των κάτω αριστερών και επάνω δεξιών άκρων τους. Έστω ένα ερώτημα  $q$  με  $K = 2$ . Αν ο χρήστης  $U_1$  ή ο χρήστης  $U_2$  έχει θέσει το ερώτημα  $q$ , τότε η περιοχή ASR είναι το κελί  $\langle (0, 2), (1, 3) \rangle$ . Αν το ερώτημα  $q$  έχει τεθεί από τον χρήστη  $U_3$  ή τον χρήστη  $U_4$ , η περιοχή ASR είναι η ένωση των κελιών  $\langle (1, 2), (2, 3) \rangle \cup \langle (1, 3), (2, 4) \rangle$ . Τέλος αν το ερώτημα  $q$  έχει τεθεί από τον χρήστη  $U_5$ , η περιοχή ASR είναι όλο το σύνολο του χώρου. [14].

Παρόμοιο πρόβλημα υποκλοπής των δεδομένων και αποκάλυψης της τοποθεσίας, έχει και η μέθοδος Casper. Για παράδειγμα, ο χρήστης  $U_5$  βρίσκεται σε μεγάλη απόσταση από τους υπόλοιπους χρήστες, άρα σε μια προσπάθεια υποκλοπής ο χρήστης δεν είναι προστατευμένος.

## 5.4 Η μετρική των mix zones

Μια mix zone για μια ομάδα χρηστών ορίζεται η χωρική περιοχή, η οποία έχει αρκετό αριθμό χρηστών, και ικανοποιεί τις ακόλουθες δύο συνθήκες:

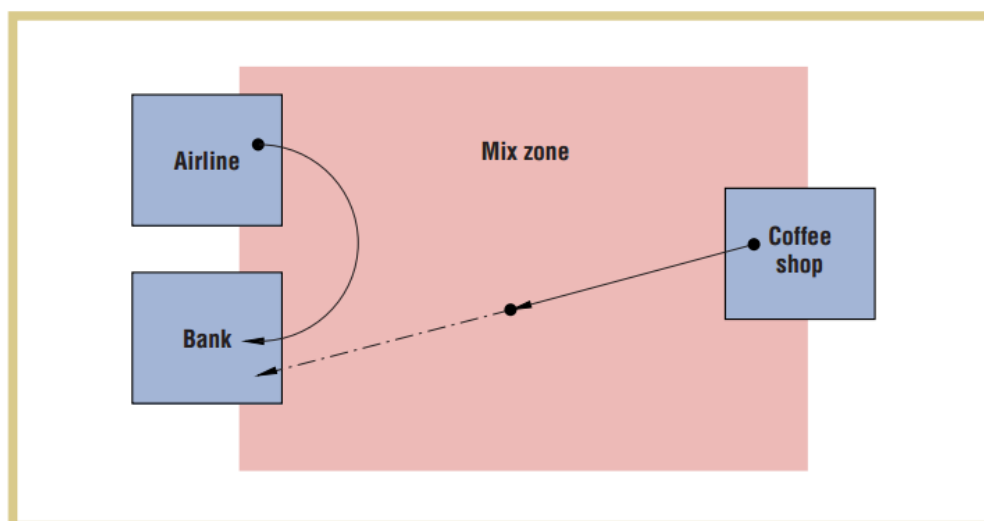
- Κανένας χρήστης δεν αποστέλλει ενημερώσεις των πληροφοριών της τοποθεσίας του στις LBSs, εντός της mix zone.
- Σε κάθε χρήστη εκχωρείται μια νέα ανώνυμη πληροφορία, κατά την έξοδο του από την mix zone.

Φυσικά, για μια ομάδα χρηστών υπάρχει περίπτωση να υπάρχουν περισσότερες από μια mix zone.

Με βάση τους δύο περιορισμούς, όταν οι χρήστες βρίσκονται μέσα σε μια mix zone, οι ταυτότητές τους διαπλέκονται. Με αυτόν τον τρόπο οι χρήστες καθώς εισέρχονται σε μια mix zone, αποκτούν μια νέα ταυτότητα ανωνυμίας σε σύγκριση με αυτήν που είχαν αρχικά. Επομένως υπάρχει περίπτωση να μην γίνεται αναγνώριση του ίδιου χρήστη, όταν αυτός εξέρχεται από την περιοχή σε σχέση με τους άλλους χρήστες, την ίδια χρονική στιγμή.

Ωστόσο, εάν μια mix zone έχει διάμετρο πολύ μεγαλύτερη από την απόσταση που ο χρήστης μπορεί να καλύψει, κατά το χρονικό διάστημα που χρειάζεται για πιθανή ενημέρωση της τοποθεσίας, υπάρχει πιθανότητα να μην γίνει επαρκής ανάμιξη μεταξύ των χρηστών.

Ένα παράδειγμα της εφαρμογής των mix zones, παρουσιάζεται στο ΣΧΗΜΑ: 5.4



ΣΧΗΜΑ 5.4: Παράδειγμα εφαρμογής της mix zones [4].

Το ΣΧΗΜΑ: 5.4 παρέχει την κάτοψη μιας mix zone. Διακρίνονται τρεις μικρότερες περιοχές, μια αεροπορική εταιρία (A), μια τράπεζα (B), και ένα κατάστημα καφέ (C). Η περιοχή A είναι πιο κοντά στην B, σε σύγκριση με την περιοχή C. Στην περίπτωση που δύο χρήστες αφήσουν την περιοχή A και την C την ίδια χρονική στιγμή, και ένας χρήστης φτάσει στην B, το πιο πιθανό είναι να είναι ο χρήστης που εξήλθε από την A περιοχή.

Επομένως σημαντικό είναι η mix zone να είναι μια μεγάλη περιοχή που να περιέχει μεγάλο αριθμό χρηστών. Για κάθε mix zone που ένας χρήστης  $U$  επισκέπτεται σε χρόνο  $t$ , ως σύνολο ανωνυμίας ορίζεται το σύνολο των χρηστών που εισέρχεται σε αυτήν κατά τον ίδιο χρόνο  $t$  [4].



## 5.5 Επίλογος

Λόγω των πιθανών υποκλοπών των προσωπικών δεδομένων των χρηστών, που μπορεί να υπάρξουν από τα συστήματα που προσφέρουν υπηρεσίες, οι προγραμματιστές οφείλουν να λάβουν υπ' όψιν και την παροχή ασφάλειας. Για τον σκοπό αυτό έχουν αναπτυχθεί μέθοδοι, ώστε να μην τίθεται κίνδυνος απειλής των ευαίσθητων δεδομένων.

Στο παρόν κεφάλαιο αναφέρθηκαν σημαντικοί αλγόριθμοι, για την προστασία των δεδομένων. Όταν ο βαθμός του συνόλου ανωνυμίας είναι μεγάλος, τότε δεν είναι εύκολο να αναγνωρισθεί το σύνολο των χρηστών. Ο αλγόριθμος διακριτοποίησης διαιρεί τις περιοχές των χρηστών, προσφέροντας έτσι μικρότερο αριθμό δεδομένων. Η μέθοδος Casper κάνει χρήση μιας δομής δεδομένων σε μορφή πυραμίδας, ενώ στην μέθοδο των mix zones η ταυτότητα του κάθε χρήστη διαπλέκεται.

## **Μέρος II**

# **ΕΦΑΡΜΟΓΕΣ**

## Κεφάλαιο 6

# VTrack Project

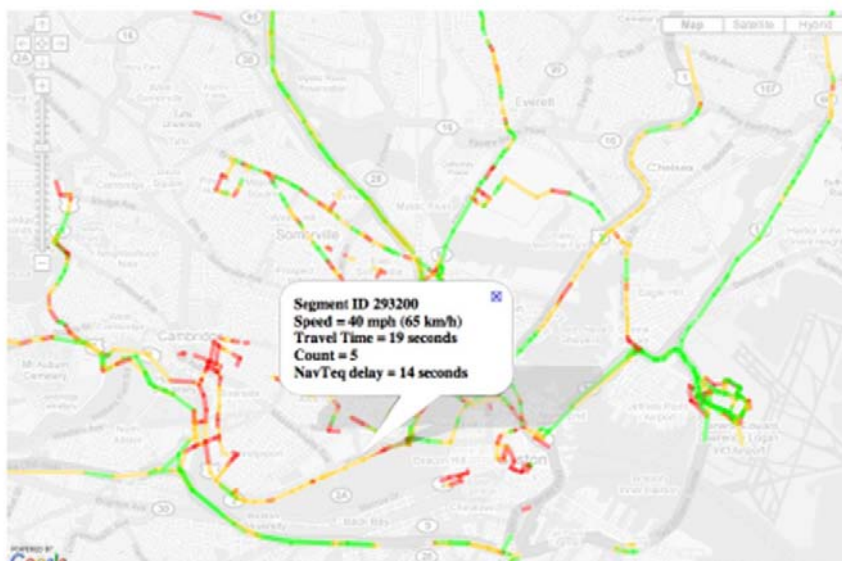
### 6.1 Εισαγωγή

Η υποδομή του οδικού συστήματος, λόγω της αυξημένης χρήσης των αυτοκινήτων και των μέσων μαζικής μεταφοράς, αντιμετωπίζει προβλήματα κυκλοφοριακής συμφόρησης. Η εξάπλωση των τηλεφώνων διανοίγει την ελκυστική δυνατότητα να αποκτηθεί ένας τεράστιος όγκος δεδομένων, απευθείας από τους χρήστες. Το πλήθος που στην κατοχή του έχει κινητά τηλέφωνα με προηγμένους αισθητήρες (γίνεται αναφορά στο [κεφάλαιο 4](#)) έχει την δυνατότητα, ανά τακτά χρονικά διαστήματα, να συλλέγει δεδομένα και πληροφορίες. Ο συνδυασμός των πληροφοριών αυτών, έχει σαν αποτέλεσμα πιθανή εξάλειψη του προβλήματος.

Το σύστημα VTrack είναι ένα σύστημα παρακολούθησης της κυκλοφορίας σε πραγματικό χρόνο. Μπορεί να επεξεργάζεται δείγματα τοποθεσίας με βάση τον χρόνο, που σαν αποτέλεσμα έχει την προσφορά πρόβλεψης και απεικόνισης της κυκλοφορίας. Επιπρόσθετα, κάνει αναφορά στα σημεία που εντοπίζεται κυκλοφοριακή συμφόρηση. Η διαδικασία της map-matching μεθόδου, δημιουργεί έναν χάρτη της κυκλοφορίας σε πραγματικό χρόνο. Επίσης η χρήση του αλγορίθμου Viterbi βοηθάει στην εύρεση βέλτιστου μονοπατιού. Το [ΣΧΗΜΑ: 6.1](#) απεικονίζει ένα στιγμιότυπο της εφαρμογής.

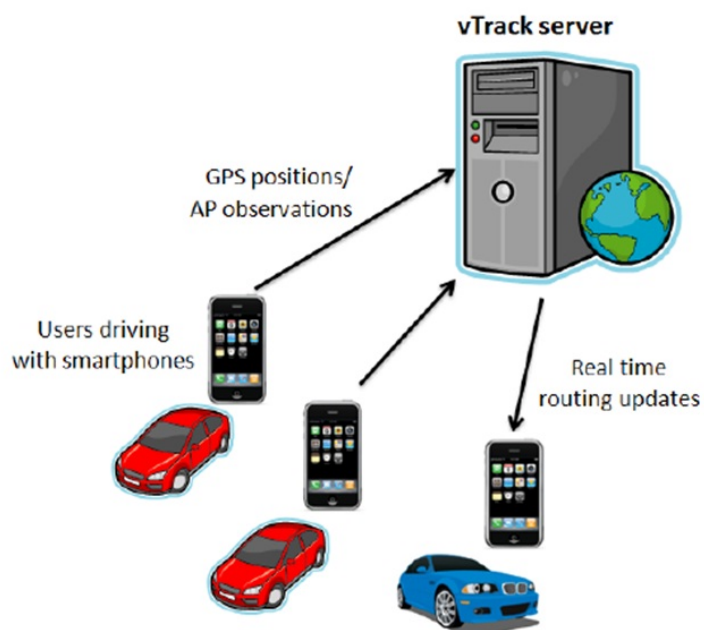
### 6.2 Αρχιτεκτονική του Συστήματος VTrack

Στο [ΣΧΗΜΑ: 6.2](#) παρουσιάζεται η δομή λειτουργίας του συστήματος VTrack. Ο χρήστης εκτελεί την εν' λόγω εφαρμογή, μέσω του τηλεφώνου του. Οι αισθητήρες τύπου GPS και WiFi, συλλέγουν και αναφέρουν τα δεδομένα στον εξυπηρετητή ανά τακτά χρονικά διαστήματα, κατά μήκος της διαδρομής που ακολουθείται. Ο εξυπηρετητής εκτελεί έναν αλγόριθμο εκτίμησης του χρόνου ταξιδιού κάνοντας χρήση των δειγμάτων θέσεως. Στόχος είναι να προσδιοριστούν τα τμήματα του οδικού δικτύου στο οποίο ο χρήστης βρίσκεται, και να εκτιμηθεί ο χρόνος



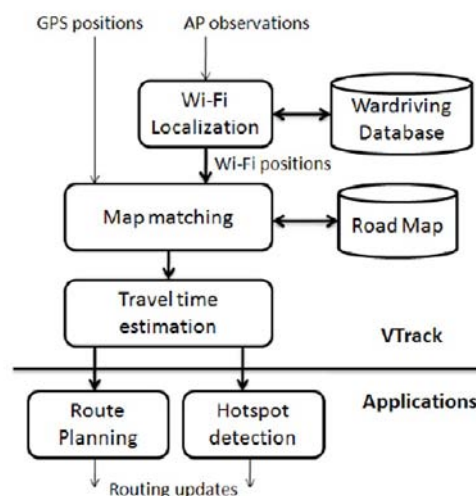
ΣΧΗΜΑ 6.1: Η εφαρμογή VTrack απεικονίζει σημεία, που εντοπίζεται κυκλοφοριακή συμφόρηση [3].

κίνησης σε αυτά τα τμήματα. Οι εκτιμήσεις στη συνέχεια χρησιμοποιούνται για τον προσδιορισμό ενεργών σημείων (hotspots), δηλαδή εκεί που υπάρχει αυξημένη κινητικότητα, και σε πραγματικό χρόνο να γίνεται σχεδιασμός της διαδρομής [3].



ΣΧΗΜΑ 6.2: Αρχιτεκτονική του VTrack project [3].

Στις περιπτώσεις που ο αισθητήρας του GPS δεν μπορεί να συνδεθεί με τον δορυφόρο για την λήψη των συντεταγμένων, είτε δεν τίθεται σε λειτουργία, είτε χρειάζεται πολύ ενέργεια για να λειτουργήσει, το σύστημα κάνει χρήση τον αισθητήρα WiFi για την εκτίμηση της τοποθεσίας. Παρατηρηθέντα σημεία πρόσβασης (AP) από το ασύρματο δίκτυο των κινητών, μετατρέπονται σε εκτιμήσεις τοποθεσίας, κάνοντας χρήση μιας “wardriving” βάσης δεδομένων των συντεταγμένων του GPS, αναφέροντας τα σημεία πρόσβασης WiFi που είχαν παρατηρηθεί από προηγούμενους χρήστες. Οι τοποθεσίες που προκύπτουν γίνονται επεξεργασία σε συνδυασμό με τον αλγόριθμο εκτίμησης, σε πραγματικό χρόνο. Ο αλγόριθμος εκτίμησης αποτελείται από δύο παραμέτρους: Έναν *map-matcher* ο οποίος καθορίζει τους δρόμους που δύναται να κυκλοφορήσουν οχήματα, και έναν *travel-time* εκτιμητή ο οποίος καθορίζει τους χρόνους μετακίνησης, για τα τμήματα του δρόμου που προκύπτουν από την χαρτογράφηση [3].



ΣΧΗΜΑ 6.3: Λειτουργία του εξυπηρετητή [3].

### 6.3 Βασικοί Στόχοι του VTrack Project

#### Ανίχνευση και απεικόνιση hotspots

Το οδικό τμήμα του οποίου, ο χρόνος μετακίνησης *υπερβαίνει* τον φυσιολογικό χρόνο που χρειάζεται ένας χρήστης υπό μια σταθερή ταχύτητα να το διασχίσει, τότε το τμήμα αυτό ονομάζεται ενεργό τμήμα (*hotspot*). Τα τμήματα αυτά υπάρχουν κάθε μέρα σε ώρα αιχμής. Μέσω της εφαρμογής, μπορεί να γίνει ανίχνευση και εμφάνιση όλων των hotspot τμημάτων, για μια δεδομένη γεωγραφική τοποθεσία. Ο χρήστης παρατηρεί τον χάρτη με ένα πρόγραμμα περιήγησης και μπορεί να τα εντοπίσει και να τα αποφύγει.

#### Σχεδιασμός της διαδρομής σε πραγματικό χρόνο

Με εξαίρεση τα *ενεργά σημεία*, οι χρήστες εν γένει ενδιαφέρονται για το συνολικό χρόνο μετακίνησής τους, εν αντιθέση για το χρόνο που πρόκειται να ξοδέψουν σε ένα συγκεκριμένο οδικό τμήμα. Ο σχεδιασμός της διαδρομής, ώστε να ελαχιστοποιηθεί ο χρόνος μετακίνησης, είναι μια μέθοδος για τους χρήστες να εντοπίσουν μια σύντομη διαδρομή προς έναν συγκεκριμένο προορισμό. Ο σχεδιασμός της διαδρομής σε πραγματικό χρόνο είναι ιδιαίτερα χρήσιμος, διότι επιτρέπει στους χρήστες να κάνουν σε πραγματικό χρόνο πιθανή αλλαγή της πορείας τους [3].

## 6.4 Αλγόριθμοι Υλοποίησης

Η μέθοδος του *map matching*, είναι η διαδικασία της αντιστοίχισης μιας πιθανής *θρομβώδης* ακολουθίας δειγμάτων, ή αραιών δειγμάτων τοποθεσίας, σε μια συνεχή αλληλουχία τμημάτων του οδικού δικτύου. Είναι σημαντική μέθοδος για την εκτίμηση του χρόνου μετακίνησης, καθώς οι χρήστες ενδιαφέρονται για το χρόνο που χρειάζονται για να διασχίσουν ένα συγκεκριμένο τμήμα του οδικού τμήματος.

Επειδή υπάρχει περίπτωση να μην αντιστοιχίζονται όλα τα παρατηρηθέντα σημεία σωστά στον χάρτη, γίνεται χρήση του Hidden Markov Model και εφαρμόζεται ο αλγόριθμος του Viterbi, ώστε να εντοπιστεί η βέλτιστη διαδρομή.

### 6.4.1 Hidden Markov Model και Μονοπάτι Viterbi

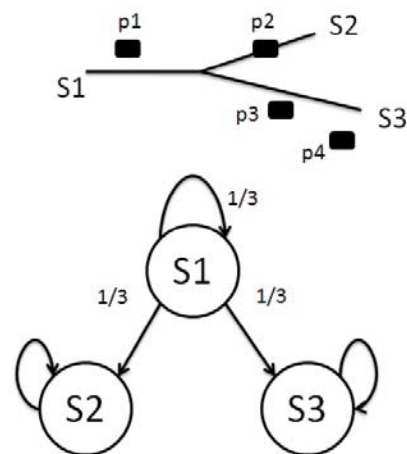
Ένα HMM αποτελείται από ένα σύνολο με κρυμμένες καταστάσεις (*hidden states*) και παρατηρήσιμα χαρακτηριστικά (*observables*). Κάθε κατάσταση εκπέμπει ένα παρατηρήσιμο χαρακτηριστικό με μια συγκεκριμένη δεσμευμένη κατανομή πιθανότητας, η οποία ονομάζεται κατανομή πιθανότητας εκπομπής. Δεδομένης κάποιας εισόδου, ένα HMM διασχίζει τις καταστάσεις και παράγει την έξοδο, η οποία αποτελείται από μια λίστα των παρατηρήσιμων χαρακτηριστικών. Τα παρατηρήσιμα χαρακτηριστικά (*observables*) εκπέμπονται σε κάθε κατάσταση. Η ακολουθία των καταστάσεων δεν είναι γνωστή, σε αντίθεση με την έξοδο. Το πρόβλημα που πρέπει να αντιμετωπιστεί, είναι να καθοριστεί η πιθανή ακολουθία των καταστάσεων που παράγαν την έξοδο [3].

Οι μεταβάσεις μεταξύ των κρυμμένων καταστάσεων, διέπονται από ένα σύνολο πιθανοτήτων που ονομάζονται *πιθανοτικές μεταβάσεις* (*transition probabilities*). Η πιθανότητα μετάβασης ορίζεται ως η πιθανότητα μετάβασης από το ένα τμήμα στο επόμενο. Δεδομένης μιας ακολουθίας θέσεων, σκοπός είναι ο προσδιορισμός της πιο πιθανής ακολουθίας των οδικών τμημάτων, που αντιστοιχεί σε αυτές τις θέσεις. Η *πιθανότητα εκπομπής* για ένα δεδομένο ζεύγος (τμήματος, θέσης), αντιπροσωπεύει την πιθανότητα να εντοπιστεί στην εν λόγω θέση το όχημα που βρίσκεται στο εν λόγω τμήμα.

Το *μονοπάτι Viterbi* είναι μια τεχνική δυναμικού προγραμματισμού για την εύρεση της μέγιστης πιθανοτικής ακολουθίας των κρυμμένων καταστάσεων δοθέντος ενός συνόλου από παρατηρήσιμα χαρακτηριστικά, καθώς και της κατανομής πιθανότητας εκπομπών και των πιθανοτήτων μετάβασης. Οι κρυφές καταστάσεις αντιστοιχούν στη διαδρομή, και τα παρατηρήσιμα χαρακτηριστικά στα δείγματα θέσεως [3].

Στο ΣΧΗΜΑ: 6.4 παρουσιάζεται ένα παράδειγμα, όπου οι καταστάσεις  $S_1$ ,  $S_2$  και  $S_3$  είναι οδικά τμήματα, και τα  $p_1$ ,  $p_2$ ,  $p_3$ , και  $p_4$  αποτελούν δείγματα τοποθεσίας.

Η πιθανότητα μετάβασης από την κατάσταση  $S1$  στον εαυτό της, στην  $S2$  και στην  $S3$  ισούται με  $\frac{1}{3}$ . Λόγω της συνάρτησης της πυκνότητας πιθανότητας, υποθέτοντας τη μετάβαση όπως φαίνεται στο διάγραμμα καταστάσεων, η μέγιστη πιθανότητα ακολουθίας των τμημάτων για τη δεδομένη ακολουθία των δειγμάτων θέσης είναι  $S1, S3, S3$  και  $S3$ . Παρά το γεγονός ότι το  $p2$  σημείο είναι πιο κοντά στο τμήμα  $S2$ , η πιθανή κρυφή κατάσταση του σημείου είναι η  $S3$ .

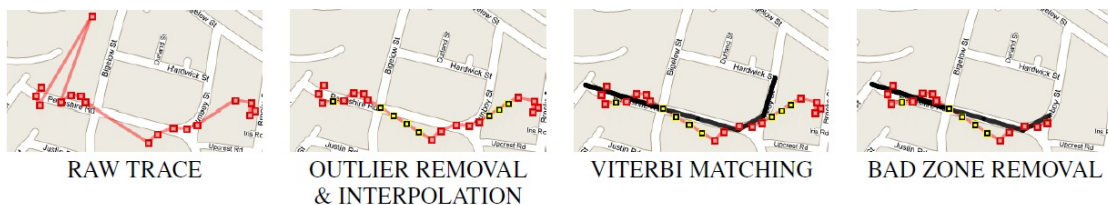


ΣΧΗΜΑ 6.4: Παράδειγμα Hidden Markov Model [3].

Είναι σαφές, από το ανώτερο παράδειγμα, γιατί ένα HMM είναι χρήσιμο για την χαρτογράφηση. Είναι *εύρωστο* σε δείγματα θέσεων που βρίσκονται πιο κοντά σε ένα τμήμα τοποθεσίας, σε αντίθεση με το σημείο στο οποίο είχαν παρατηρηθεί, και έχει τη δυνατότητα να καταγράψει την έννοια μιας συνεχούς διαδρομής, αντί για μια ακολουθία τμημάτων [3].

#### 6.4.2 Χαρτογράφηση Διαδρομών

Στο ΣΧΗΜΑ: 6.5 απεικονίζεται ένα παράδειγμα της χαρτογράφησης.



ΣΧΗΜΑ 6.5: Η διαδικασία χαρτογράφησης διαδρομών[3].

Πριν από την χρήση ενός HMM, είναι σημαντικό να έχει γίνει προ-επεξεργασία των δεδομένων, για την εξάλειψη των πιθανών *ακραίων τιμών* ή *εσφαλμένων αποτελεσμάτων (outliers)*, ώστε να αντιμετωπιστούν πιθανές *ασυνέχειες* της διαδρομής. Ένα δείγμα  $p$  δηλώνεται ως *ακραία τιμή*, εάν παραβιάζει έναν περιορισμό ταχύτητας. Δηλαδή όταν για κάποια κατώτατη οριακή ταχύτητα  $S_{outlier}$ , ο χρήστης είχε μετακινηθεί γρηγορότερα από την τιμή  $S_{outlier}$  να ταξιδέψει από ένα προηγούμενο δείγμα στο δείγμα  $p$ , του παρατηρούμενου χρονικού διαστήματος.

Η τιμή της  $S_{outlier}$  ορίζεται να είναι περίπου διπλάσια της αναμενόμενης μέγιστης ταχύτητας σε έναν αυτοκινητόδρομο. Το κατώτατο αυτό όριο είναι κατάλληλο για την κατηγοριοποίηση δύο διαδοχικών θορυβωδών δειγμάτων GPS ή WiFi.

Αφού γίνει αφαίρεση των ακραίων τιμών, στην συνέχεια γίνεται επεξεργασία των *ασυνεχειών*. Ο αλγόριθμος που χρησιμοποιείται διαθέτει ένα απλό και αποτελεσματικό σύστημα,

το οποίο εισάγει *παρεμβαλλόμενα σημεία* σε περιοχές που τα δεδομένα τοποθεσίας έχουν ασυνέχεια. Ο αλγόριθμος παράγει παρεμβαλλόμενα δείγματα με χρονικά διάστημα της τάξης του 1 δευτερολέπτου, κατά μήκος του ευθύγραμμου τμήματος που συνδέει το τελευταίο παρατηρούμενο σημείο πριν την ασυνέχεια με το πρώτο μετά την ασυνέχεια, υποθέτοντας ότι η ταχύτητα κατά μήκος αυτής της διαδρομής παραμένει σταθερή. Αυτά τα παρεμβαλλόμενα σημεία στην συνέχεια τροφοδοτούνται στο HMM, μαζί με τα άλλα δειγματοληπτούμενα σημεία. Η απομάκρυνση των ακραίων τιμών και τα βήματα παρεμβολής παρουσιάζονται στο δεύτερη εικόνα του σχήματος 6.5. Τα παρεμβαλλόμενα σημεία εμφανίζονται με κίτρινο χρώμα.

Αφού οι ακραίες τιμές έχουν αφαιρεθεί και εφαρμοστεί η παρεμβολή, εφαρμόζεται ο αλγόριθμος *Viterbi*. Οι κρυμμένες καταστάσεις στο μοντέλο του Markov είναι κατευθυνόμενα τμήματα οδικών οδών, και οι παρατηρήσεις αποτελούν δείγματα θέσεων. Ο αλγόριθμος *Viterbi* χρησιμοποιείται για να προβλέψει την πιο πιθανή ακολουθία των οδικών τμημάτων, που αντιστοιχούν στα παρατηρηθέντα και παρεμβαλλόμενα σημεία. Η παρατηρηθείσα ακολουθία καθορίζει τη διαδρομή που λήφθηκε. Στην τρίτη εικόνα του σχήματος 6.5 αναπαριστάται η πιο πιθανή διαδρομή σε μαύρο χρώμα. Να σημειωθεί ότι η έξοδος είναι μια ακολουθία ζευγών (σημείου, τμήματος οδικού οδού), που διαδοχικά σημεία μπορούν να βρίσκονται στο ίδιο οδικό τμήμα.

Τέλος, εφαρμόζεται *απομάκρυνση της κακής ζώνης* (τέταρτη εικόνα του σχήματος 6.5), για την απομάκρυνση αντιστοιχίσεων *Viterbi* “χαμηλής αξιοπιστίας” όπως είναι για παράδειγμα θορυβώδη δεδομένα [3].

*Στην συνέχεια γίνεται ανάλυση των εννοιών της πιθανότητας μετάβασης που διέπει την πιθανή ακολουθία των οδικών τμημάτων, καθώς και της πιθανότητας εκπομπής η οποία διέπει τα πιο πιθανά οδικά τμήματα για ένα συγκεκριμένο σημείο, το οποίο μπορεί να έχει παρατηρηθεί.*

## Πιθανότητα Μετάβασης

Οι πιθανότητες μετάβασης χαρακτηρίζονται από τις ακόλουθες τρεις έννοιες:

1. Για ένα δεδομένο τμήμα μιας οδού, υπάρχει μια πιθανότητα ότι στο επόμενο σημείο ο χρήστης θα εξακολουθεί να είναι σε αυτό το τμήμα της οδού.
2. Ο χρήστης μπορεί να ταξιδεύει μόνο από το τέλος ενός τμήματος του δρόμου στην αρχή του επομένου, εάν χρησιμοποιεί την ίδια διασταύρωση. Έτσι εξασφαλίζεται ότι οι έξοδοι των οδικών τμημάτων έχουν συνέχεια.
3. Ένας χρήστης δεν μπορεί να ταξιδεύει χωρίς αιτία, πολύ γρήγορα σε οποιοδήποτε τμήμα.



Η πιθανότητα μετάβασης  $p$  από ένα τμήμα  $i$  στο σημείο  $t - 1$ , σε ένα τμήμα  $j$  στο σημείο  $t$  ορίζεται ως εξής:

1. Εάν  $i = j$  τότε  $p = \epsilon$
2. Εάν το τμήμα  $j$  δεν αρχίζει εκεί που τελειώνει το  $i$ , τότε  $p = 0$
3. Εάν το τμήμα  $j$  αρχίζει εκεί που τελειώνει το  $i$ , τότε  $p = \epsilon$  ή  $p = 0$

Το  $\epsilon$  ορίζεται να είναι μικρότερο της τιμής  $\frac{1}{(d_{max}+1)}$ , όπου  $d_{max}$  είναι ο μέγιστος αριθμός των τμημάτων που ξεκινούν από το τέλος του ίδιου τμήματος.

Όλες οι μη μηδενικές πιθανότητες μετάβασης παραμένουν σταθερές σε όλα τα τμήματα, προκειμένου να αποφευχθεί η προτίμηση σε διαδρομές με χαμηλόβαθμα τμήματα (δηλαδή διαδρομές με διασταυρώσεις που δεν έχουν πολλές εξερχόμενες οδικές οδούς). Ταυτόχρονα όμως, πρέπει το άθροισμα των μη μηδενικών πιθανοτήτων μετάβασης να ισούται με 1 για ένα δεδομένο τμήμα. Για να διασφαλιστεί αυτή η προϋπόθεση, χρησιμοποιείται ένα εικονικό τμήμα που οδηγεί σε αδιέξοδο και συμβολίζεται ως  $\emptyset$ , και ένα εικονικό παρατηρήσιμο  $\perp$ . Η πιθανότητα μετάβασης από το τμήμα  $\emptyset$  στο σημείο  $t - 1$ , προς το τμήμα  $\emptyset$  στο σημείο  $t$  ισούται με 1. Ουσιαστικά γίνεται εκχώρηση μηδενικής πιθανότητας σε όλες τις διαδρομές που έχουν μετάβαση στο τμήμα  $\emptyset$ .

Το 3ο στοιχείο στον ορισμό της πιθανότητας μετάβασης, υποδηλώνει την απαγόρευση των χρηστών να ταξιδεύουν αδικαιολόγητα γρήγορα σε οποιοδήποτε τμήμα. Για μια μετάβαση από το τμήμα  $i$  στο τμήμα  $j$ , ο αλγόριθμος υπολογίζει το χρόνο που θα έχει κάνει ο χρήστης για να ταξιδέψει από το  $i$  στο  $j$ , με βάση τους χρόνους, στους οποίους οι θέσεις παρατηρήθηκαν. Αν ο χρόνος αυτός δηλώνει ότι ο χρήστης ταξιδεύει με υψηλότερη ταχύτητα από ό, τι την ταχύτητα κατωφλιού ( $S_{outlier}$ ), η μετάβαση αυτή είναι αδύνατη και ανατίθεται με πιθανότητα ίση με 0. Σε αντίθετη περίπτωση, η μετάβαση είναι δυνατή και ανατίθεται με πιθανότητα ίση με  $\epsilon$  [3].

### Πιθανότητα Εκπομπής

Οι πιθανότητες εκπομπής αντικατοπτρίζουν την πιθανότητα ότι ένα συγκεκριμένο σημείο να παρατηρήθηκε σε ένα κοντινό οδικό τμήμα, αλλά όχι απαραίτητα το πλησιέστερο. Συγκεκριμένα, η εκπομπή της πυκνότητας πιθανότητας του τμήματος  $i$  στην θέση  $l$  είναι  $N(dist(i, l))$  όπου  $N$  είναι μια *Gaussian* συνάρτηση με μηδενική μέση τιμή, και όπου  $dist(i, l)$  είναι η Ευκλείδεια απόσταση μεταξύ του  $i$  και του  $l$ . Η διακύμανση του  $N$  εξαρτάται από τον αισθητήρα. Γίνεται χρήση διαφορετικών διακυμάνσεων για τον αισθητήρα του GPS, και άλλων για τον αισθητήρα του WiFi, επειδή έχουν διαφορετική κατανομή λάθους [3].

## 6.5 Εκτίμηση της Χρονικής Διάρκειας της Διαδρομής

Ο χρόνος  $T(S)$  που απαιτείται για να διασχίσει ένας χρήστης ένα οποιοδήποτε τμήμα ισούται με

$$T(S) = T_{left}(S) + T_{matched}(S) + T_{right}(S) \quad (6.1)$$

Ο  $T_{left}(S)$  είναι ο χρόνος μεταξύ του *απαραίτητου* σημείου εισόδου για το τμήμα  $S$ , και του πρώτου παρατηρούμενου σημείου. Ο  $T_{matched}(S)$  είναι ο χρόνος μεταξύ των πρώτων και των τελευταίων σημείων του τμήματος  $S$ . Ο  $T_{right}(S)$  είναι ο χρόνος μεταξύ του τελευταίου σημείου του  $S$ , και του *απαραίτητου* σημείου εξόδου.

Όπως έχει αναφερθεί στην ενότητα: 6.4.2, στην χαρτογράφηση προστίθενται παρεμβολόμενα σημεία, ώστε να διασφαλιστεί ότι τα σημεία στην έξοδο είναι υψηλής συχνότητας. Τυπικά διαχωρίζονται κατά ένα δευτερόλεπτο, και εγγυημένα διαχωρίζονται, με το πολύ, δύο δευτερόλεπτα. Ως εκ τούτου, εάν κατά την χαρτογράφηση εξάγεται μια συνεχή αλληλουχία τμημάτων, και οι χρόνοι  $T_{left}(S)$  και  $T_{right}(S)$  είναι άνω φραγμένοι κατά 1 δευτερόλεπτο για όλα τα τμήματα ανεξαρτήτου μεγέθους, ο χρόνος  $T_{matched}(S)$ , είναι ο καθοριστικός παράγοντας της καθυστέρησης.

Η εκτίμηση του χρόνου γίνεται, αφού πρώτα γίνει ανάθεση του χρόνου  $T_{matched}(S)$  στο τμήμα  $S$ . Μετά γίνεται ο υπολογισμός του χρονικού διαστήματος, μεταξύ του πρώτου σημείου του  $S$  και του τελευταίου σημείου που αντιστοιχίζεται με το  $S_{prev}$  του τμήματος που προηγείται του  $S$ . Κατόπιν γίνεται διαίρεση ισομερώς, μεταξύ των  $T_{right}(S_{prev})$  και  $T_{left}(S_{prev})$ . Η ίδια διαδικασία συνεχίζεται και για το  $S_{next}$  τμήμα, που ακολουθεί το  $S$ , και η διαδικασία επαναλαμβάνεται [3].

## 6.6 Αξιολόγηση του VTrack project

Για την σωστή αξιολόγηση του συστήματος VTrack, γίνεται χρήση ενός μεγάλου συνόλου δεδομένων που έχει ληφθεί μέσω των αισθητήρων του GPS και του WiFi των τηλεφώνων των χρηστών. Γίνεται αναφορά στο σύνολο δεδομένων, καθώς και στη διαδικασία που χρησιμοποιείται, για τον **καθαρισμό και την επιλογή των μη θορυβωδών δειγμάτων** έτσι ώστε να προκύψουν λογικές τιμές αλήθειας για τις διαδρομές. Στη συνέχεια, γίνεται αξιολόγηση της ποιότητας των καταστάσεων πρόβλεψης, και του **χρόνου της χαρτογράφησης της συντομότερης διαδρομής**, καθώς επίσης και του **χρόνου ανίχνευσης των ενεργών σημείων**, για διαφορετικούς συνδυασμούς των δεδομένων του GPS και του WiFi.

### 6.6.1 Λογικές Τιμές Δεδομένων και Πληροφοριών

Για ένα σύστημα αξιολόγησης και μέτρησης του χρονικού διαστήματος μιας διαδρομής, η απόκτηση λογικών τιμών αλήθειας αποτελεί θεμελιώδη παράμετρο. Φυσικά, λόγω της κατασκευής των τηλεφώνων αλλά και των συνθηκών που επικρατούν στο περιβάλλον, είναι αδύνατο να ληφθούν ιδανικές τιμές αλήθειας.

Στο VTrack project, πραγματοποιείται χρήση μιας τεχνικής που βασίζεται στον απευθείας “καθαρισμό των δεδομένων” για την εξαγωγή λογικών και αληθινών τιμών, για κάθε υποσύνολο της διαδρομής. Η τεχνική ακολουθεί τα εξής βήματα:

1. Για κάθε σημείο  $g$  του δείγματος από τον αισθητήρα GPS, το σύνολο των τμημάτων ορίζεται ως  $S_g$ , σε ακτίνα 15 μέτρων από το σημείο  $g$ . Η επιλογή των 15 μέτρων γίνεται, διότι τόσο είναι το πλάτος ενός τυπικού τμήματος του δρόμου.
2. Αφού διαπιστωθεί ο κενός χώρος που δημιουργούν τα τμήματα  $S_g$ , η ακολουθία των σημείων  $g$  ενώνεται, με στόχο την δημιουργία μιας συνεχής αλληλουχίας  $X_g$ , όπου ισχύει ότι κάθε  $X_g \in S_g$ . Επομένως, κάθε σημείο αντιστοιχίζεται με ένα από τα γειτονικά σημεία του. Επίσης γίνεται αφαίρεση όλων των  $g$  σημείων, τα οποία δεν μπορούν να αντιστοιχηθούν. Για παράδειγμα, εάν δεν υπάρχουν τμήματα εντός μιας ακτίνας των 15 μέτρων.
3. Μετά την δημιουργία της συνεχής αλληλουχίας  $X_g$ , γίνεται έλεγχος για πιθανές ασυνέχειες σε περισσότερα χρονικά διαστήματα.
4. Αφού ολοκληρωθεί η διαδικασία, γίνεται προβολή κάθε σημείου  $g$ , στο πλησιέστερο σημείο της αλληλουχίας  $X_g$  ώστε να ληφθεί το αντίστοιχο αληθινό σημείο  $g'$ .

Η νέα αλληλουχία των σημείων  $g'$  που παράγεται, συνθέτει την διαδρομή χωρίς θορυβώδη δεδομένα. Η σύγκληση των σημείων αυτών, ικανοποιεί τρεις περιορισμούς:

- δεν υφίσταται κενό ανάμεσα στα σημεία,
- κάθε σημείο από τον αισθητήρα του GPS, αντιστοιχίζεται με ένα τμήμα που απέχει το πολύ 15 μέτρα από αυτό,
- τα τμήματα που προκύπτουν αποτελούν μια αδιάσπαστη διαδρομή.

Η τεχνική εύρεσης των αληθινών σημείων, δεν είναι χρονοβόρα. Ωστόσο για τα σημεία που προκύπτουν, υπάρχει πιθανότητα να μην προσδιορίζουν την πιο συμφέρουσα διαδρομή που μπορεί να επιτευχθεί. Αυτό συμβαίνει διότι χάνονται σημεία, κατά την διάρκεια εκτέλεσης του αλγορίθμου.

### 6.6.2 Χαρτογράφηση των Διαδρομών

Κατά τον σχεδιασμό της διαδρομής, στόχος είναι να βρεθούν οι καλύτερες και οι συντομότερες διαδρομές μεταξύ των ζευγών (πηγής - προορισμού) του οδικού δικτύου. Για τον λόγο αυτό, αφού γίνει υπολογισμός του χρόνου εκτίμησης, που παράγεται για διαφορετικούς συνδυασμούς των GPS και WiFi δειγμάτων, βασικό μέλημα είναι η βελτιστοποίηση και η ελαχιστοποίηση του αναμενόμενου χρόνου της διάρκειας της διαδρομής.

Για να γίνει σωστός σχεδιασμός, λαμβάνεται ως είσοδος ένα σύνολο, από τις παραγόμενες διαδρομές χωρίς θορυβώδη δεδομένα, που ονομάζονται  $D_{gt}$ , και ένα αντίστοιχο σύνολο από θορυβώδεις διαδρομές,  $D_{noisy}$ . Κατόπιν γίνεται εκτίμηση του χρόνου μετακίνησης για κάθε τμήμα του συνόλου  $D_{noisy}$ . Στόχος είναι να εντοπιστούν οι λάθος εκτιμήσεις, και πώς επηρεάζουν την ποιότητα της πρόβλεψης της συντομότερης διαδρομής. Δηλαδή, πόσο χειρότερες είναι οι διαδρομές που βρίσκει ο σχεδιαστής διαδρομής, κάνοντας χρήση των ανακριβών χρόνων μετακίνησης, σε σύγκριση με τις διαδρομές που μπορεί να βρει από τα σωστά παραγόμενα δεδομένα [3].

Το σύνολο  $S_{gt}$ , ορίζεται ως το σύνολο των οδικών τμημάτων, για το οποίο υπάρχουν σωστές τιμές εκτίμησης του χρόνου μετακίνησης, από τουλάχιστον μια διαδρομή του συνόλου  $D_{gt}$ . Στην συνέχεια σχηματίζεται το επαγόμενο γράφημα  $G_{gt}$  ολόκληρου του οδικού δικτύου, του συνόλου  $S_{gt}$ . Το γράφημα  $G_{gt}$  υπάρχει περίπτωση να μην παραμείνει σταθερό. Αν συμβεί αυτό, το διαιρούμε σε συνεκτικές συνιστώσες. Αφού επιλεγεί μια συνεκτική συνιστώσα τυχαία, προσομοιώνεται ένα δρομολόγιο ανάμεσα σε ένα τυχαίο ζευγάρι (πηγής - προορισμού), εντός της εν λόγω συνιστώσας. Για κάθε τμήμα  $S$  του γραφήματος  $G_{gt}$ , επιλέγονται δύο σύνολα βάρους:

- Μια τιμή βάρους για τις σωστές τιμές του  $S$ , από κάποια διαδρομή του  $D_{gt}$ .
- Ένα εκτιμώμενο βάρος επιλέγεται, για την εκτίμηση του χρόνου μετακίνησης για το τμήμα  $S$ , από τη θορυβώδη διαδρομή. Εάν η τροχιά για το τμήμα του  $D_{noisy}$  παραλείψει το τμήμα  $S$ , λαμβάνεται υπ' όψιν η εκτίμηση του χρόνου μετακίνησης, κάνοντας χρήση του κλιμακούμενου ορίου ταχύτητας.

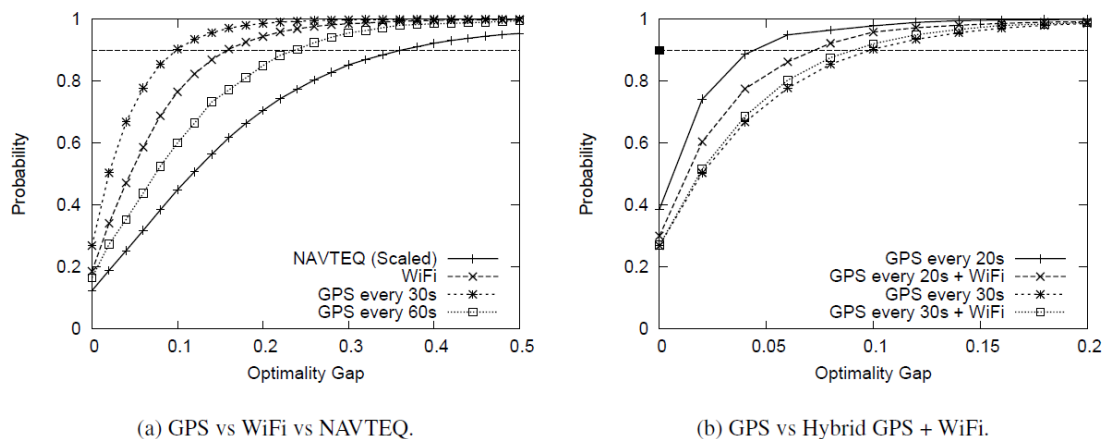
Κατόπιν εφαρμόζεται ο αλγόριθμος **Dijkstra** στο γράφημα  $G_{gt}$  με τα δύο διαφορετικά σύνολα βάρους, ώστε να βρεθούν δύο μονοπάτια το  $P_{gt}$  και το  $P_{noisy}$ . Για την αξιολόγηση της ποιότητας του σχεδιασμού της διαδρομής, συγκρίνονται οι χρόνοι των σωστών τιμών του χρόνου μετακίνησης για αυτές τις δύο διαδρομές. Ο τύπος 6.2, υπολογίζει την βέλτιστη διαφορά για κάθε ζευγάρι (πηγής - προορισμού):

$$\frac{Time(P_{noisy}) - Time(P_{gt})}{Time(P_{gt})} \quad (6.2)$$

Το ΣΧΗΜΑ: 6.6a παρουσιάζει τις *αθροιστικές συναρτήσεις κατανομής* (CDF's) της βέλτιστης διαφοράς σε δείγμα 10,000 τυχαία επιλεγμένων ζευγών (πηγής - προορισμού).

Το ΣΧΗΜΑ: 6.6b παρουσιάζει την ίδια CDF, αλλά συγκρίνοντας τη στρατηγική της GPS δειγματοληψίας κάθε  $k$  δευτερόλεπτα, παρεμβάλλοντας ανάμεσα εκτιμήσεις από τον αισθητήρα WiFi.

Όπως φαίνεται από τις μετρήσεις, οι χρόνοι μετακίνησης για τα δείγματα από τον αισθητήρα WiFi, είναι από μόνοι τους αρκετά καλοί για τον σχεδιασμό της διαδρομής. Ομοίως, οι χρόνοι μετακίνησης από την GPS δειγματοληψία κάθε 30 δευτερόλεπτα ή λίγο περισσότερο, είναι αρκετά καλοί. Και οι δύο παραπάνω στρατηγικές έχουν σημαντικά υψηλές επιδόσεις,



(a) GPS vs WiFi vs NAVTEQ.

(b) GPS vs Hybrid GPS + WiFi.

ΣΧΗΜΑ 6.6: Αθροιστικές Συναρτήσεις Κατανομής [3].

κάνοντας χρήση κλιμακωτών ορίων ταχύτητας. Επιπρόσθετα, μια υβριδική στρατηγική χρησιμοποιώντας τόσο δείγματα του GPS κάθε 30 δευτερόλεπτα καθώς και δείγματα του αισθητήρα WiFi, βελτιώνει την απόδοση σε σχέση με απλή GPS δειγματοληψία ή απλά χρησιμοποιώντας τον αισθητήρα του WiFi.

### 6.6.3 Ανίχνευση Ενεργών Σημείων

Ένα τμήμα του δρόμου παρουσιάζει “υψηλή καθυστέρηση”, αν ο παρατηρούμενος χρόνος μετακίνησης σε αυτό το τμήμα διαφέρει από το χρόνο μετακίνησης που εκτιμάται, με βάση τα κλιμακωτά όρια ταχύτητας. Τα ενεργά σημεία για μια συγκεκριμένη οριακή τιμή, είναι το σύνολο των οδικών τμημάτων που έχουν μεγάλη καθυστέρηση, με βάση τα σωστές τιμές κίνησης.

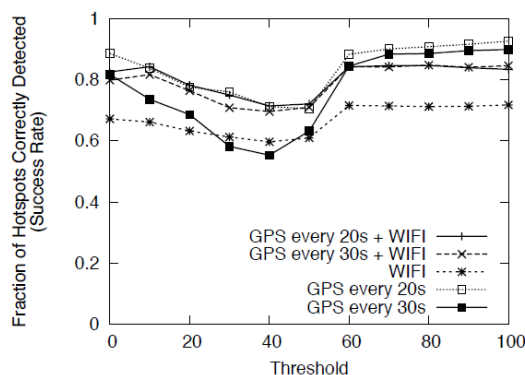
Ένας εναλλακτικός ορισμός ενός ενεργού σημείου είναι ένα οδικό τμήμα στο οποίο ο παρατηρούμενος χρόνος μετακίνησης είναι μεγαλύτερος από το κατώτατο όριο του χρόνου μετακίνησης που εκτιμάται με τα κλιμακωτά όρια ταχύτητας. Ο ορισμός όμως αυτός δεν είναι αντιπροσωπευτικός, λόγω της εξάρτησής του από το μήκος του τμήματος. Μικρά τμήματα τα οποία έχουν χαμηλό εκτιμώμενο χρόνο μετακίνησης, θα ήταν πιο πιθανό να επισημαίνονται ως ενεργά σημεία, σε σύγκριση με τα μεγάλα τμήματα τα οποία έχουν υψηλό εκτιμώμενο χρόνο μετακίνησης. Έτσι, χρησιμοποιείται ο πρώτος ορισμός, καθώς αντικατοπτρίζει με μεγαλύτερη ακρίβεια τα τμήματα οδών ως ενεργά σημεία.

Για να γίνει η ανίχνευση, χρησιμοποιούνται τα δεδομένα από τους αισθητήρες, και γίνεται εντοπισμός των τμημάτων που παρουσιάζουν υψηλή καθυστέρηση στο χρόνο. Στην συνέχεια, γίνεται ταξινόμηση του καθενός από αυτά τα τμήματα ως ένα ενεργό σημείο, καθώς και τα δύο προσκείμενα τμήματα σε αυτά.

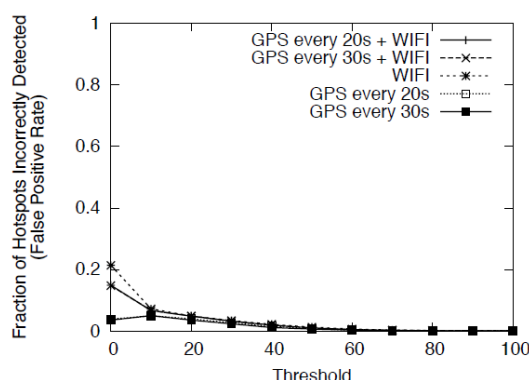
Για την αποδοτικότητα της μεθόδου, χρησιμοποιούνται δύο μετρικές: το ποσοστό επιτυχίας, καθώς επίσης και το ποσοστό των λανθασμένων αποτελεσμάτων. Το ποσοστό επιτυχίας αντιπροσωπεύει πόσα ενεργά σημεία βρέθηκαν με επιτυχία. Το ποσοστό των λανθασμένων

αποτελεσμάτων ισούται με τα ενεργά σημεία που “ανιχνεύονται”, αλλά δεν ήταν στην πραγματικότητα. Δηλαδή υποδηλώνεται στον χρήστη να αποφύγει ένα τμήμα, που δεν παρουσιάζει στην πραγματικότητα κυκλοφοριακή συμφόρηση [3].

Το ΣΧΗΜΑ: 6.7 απεικονίζει τα ποσοστά επιτυχούς ανίχνευσης ενεργών σημείων, ενώ το ΣΧΗΜΑ: 6.8 δείχνει τα ποσοστά των λανθασμένων αποτελεσμάτων, για κάθε στρατηγική.



ΣΧΗΜΑ 6.7: Ποσοστά επιτυχούς ανίχνευσης Hotspot [3].



ΣΧΗΜΑ 6.8: Ποσοστά λανθασμένων αποτελεσμάτων ανίχνευσης Hotspot [3].

Όπως φαίνεται και από τις ανωτέρω γραφικές παραστάσεις, όσον αφορά τις στρατηγικές που περιλαμβάνουν δεδομένα από τον αισθητήρα GPS, το ποσοστό επιτυχούς ανίχνευσης είναι σταθερά πάνω από 0.8, και πολλές φορές γύρω στο 0.9. Αυτό συνεπάγεται ότι οι στρατηγικές αυτές μπορούν να ανιχνεύσουν με επιτυχία πιθανά ενεργά σημεία μεταξύ 80% και 90%. Το ποσοστό επιτυχίας για τα WiFi δεδομένα είναι πολύ χειρότερο, γύρω στο 0.65. Το γεγονός αυτό προκύπτει, διότι υπάρχουν σημαντικές ασυνέχειες στα WiFi δεδομένα, και ο αλγόριθμος ανίχνευσης ενεργών σημείων δεν μπορεί να τα εντοπίσει, εάν δεν διαθέτει τα απαραίτητα δεδομένα.

Με εξαίρεση τα ληφθέντα δεδομένα, κάθε 30 δευτερόλεπτα, του αισθητήρα GPS, το ποσοστό επιτυχίας κάθε στρατηγικής παραμένει σχετικά ομοιόμορφο, υποδεικνύοντας ότι ο αλγόριθμος είναι εύρωστος, και απεικονίζονται ορθά τα ενεργά σημεία.

Επιπρόσθετα, το ποσοστό των λανθασμένων αποτελεσμάτων παραμένει σε χαμηλά επίπεδα. Αυτό υποδεικνύει ότι ο αλγόριθμος ανίχνευσης δεν επισημαίνει όλα τα τμήματα, τα οποία δεν είναι ενεργά σημεία. Αυτή είναι μια επιθυμητή ιδιότητα, καθώς δεν είναι επιθυμητό οι χρήστες να αποφύγουν οδικά τμήματα τα οποία στην πραγματικότητα δεν παρουσιάζουν κυκλοφοριακή συμφόρηση.

Το γεγονός ότι το ποσοστό των λανθασμένων αποτελεσμάτων αυξάνει μόνο για μικρά όρια, συμβαίνει διότι για ένα μικρό όριο είναι πιθανό να επισημανθούν πολλές ομάδες τμημάτων ως ενεργά σημεία. Οι ομάδες όμως αυτές, μπορεί να περιλαμβάνουν ορισμένα τμήματα που δεν έχουν υψηλή καθυστέρηση, αλλά συμπεριλήφθηκαν επειδή ήταν κοντά σε ένα τμήμα με υψηλή καθυστέρηση.

Στα κατώτατα όρια τα οποία είναι χαμηλά, έχουν επισημανθεί πολλά τμήματα ως ενεργά σημεία. Αυτό συνεπάγεται να υπάρχει υψηλό ποσοστό επιτυχίας. Καθώς η τιμή του κατώτατου ορίου αρχίζει να αυξάνεται, αγνοούνται ορισμένα τμήματα ως ενεργά σημεία, και τα λανθασμένα ποσοστά αποτελεσμάτων μειώνονται. Αυτό εξηγείται στο μέρος του γραφήματος, που υπάρχει πτώση στην γραφική παράσταση για όλες τις στρατηγικές.

Στο δεύτερο τμήμα του γραφήματος καθώς αυξάνει το κατώτατο όριο, ο αριθμός του συνολικού αριθμού των ενεργών σημείων μειώνεται. Για περίπου 40 δευτερόλεπτα, το ποσοστό της μείωσης επιβραδύνεται, και από τα 60 δευτερόλεπτα και έπειτα ο αριθμός των ενεργών σημείων παραμένει αρκετά σταθερός. Αυτό συνεπάγεται ότι πολλά από τα οδικά τμήματα που είναι ενεργά σημεία με το όριο των 60 δευτερολέπτων, είναι επίσης ενεργά σημεία και με κατώτατο όριο στα 100 δευτερόλεπτα [3].

## 6.7 Επίλογος

Στο παρόν κεφάλαιο παρουσιάστηκε η λειτουργία του VTrack συστήματος. Το πλήθος, αφού πραγματοποιήσει την εγκατάσταση της εφαρμογής στο τηλέφωνό του, συλλέγει δεδομένα μέσω των αισθητήρων του GPS και του WiFi. Τα δεδομένα αυτά, αφού υποστούν επεξεργασία, εξάγουν χρήσιμα συμπεράσματα όπως είναι η εκτίμηση των χρόνων μετακίνησης στο οδικό δίκτυο που κινείται ο κάθε χρήστης. Επίσης δύναται να εντοπιστούν πιθανές περιοχές που εμφανίζουν αυξημένη κυκλοφορία. Ο συνδυασμός των δεδομένων, μπορεί να προσδιορίσει την συντομότερη διαδρομή, εάν υπάρχει. Οι δυνατότητες αυτές μπορούν να επιτευχθούν, χρησιμοποιώντας μια αλληλουχία δειγμάτων τοποθεσίας, και κάνοντας τον σχεδιασμό της διαδρομής.

## Κεφάλαιο 7

# Mobile Millennium Project

### 7.1 Εισαγωγή

Το **Mobile Millennium project** [1, 37] είναι μια εφαρμογή που βασίζεται στην συμμετοχή του πλήθους. Το πλήθος, έχοντας στην κατοχή του συσκευές που υποστηρίζουν την εφαρμογή, μπορεί να συλλέξει δεδομένα και πληροφορίες. Η σπουδαιότητά του για την ενημέρωση των συνθηκών που επικρατούν στα αστικά κέντρα, παρουσιάζει έντονο ενδιαφέρον.

Το σύστημα της εφαρμογής Mobile Millennium παρουσιάζει τις συνθήκες της κυκλοφορίας για πολυπληθέστερες περιοχές. Είναι ένα σύστημα πρόβλεψης και εκτίμησης των κυκλοφοριακών συνθηκών, κάνοντας χρήση των μετρήσεων από τον αισθητήρα GPS των συσκευών που διαθέτουν οι χρήστες [21]. Οι συνθήκες κυκλοφορίας στη συνέχεια επιστρέφονται πίσω στις συσκευές των χρηστών, παρέχοντας την δυνατότητα να ακολουθήσουν πιο σύντομες διαδρομές [17].

Εξαιτίας του γεγονότος ότι η χρήση του επεκτάθηκε, και τα αποτελέσματα έπρεπε να είναι πάντα επίκαιρα, γίνεται χρήση της τεχνολογίας cloud computing. Τα πλεονεκτήματα από αυτήν την ενέργεια είναι οι *επίκαιρες προβλέψεις*, η *δυνατότητα επέκτασης σε μεγαλύτερα οδικά δίκτυα* και η *δυνατότητα να χρησιμοποιηθούν πιο ακριβή υπολογιστικά μοντέλα της κυκλοφορίας*, γεγονός όμως πιο δαπανηρό σε κόστος.

Ο βασικός αλγόριθμος του συστήματος είναι ο **Expectation Maximization (EM) αλγόριθμος**, βάση του οποίου γίνεται εκτίμηση των συνθηκών της κυκλοφορίας που επικρατεί στους οδικούς άξονες [21].

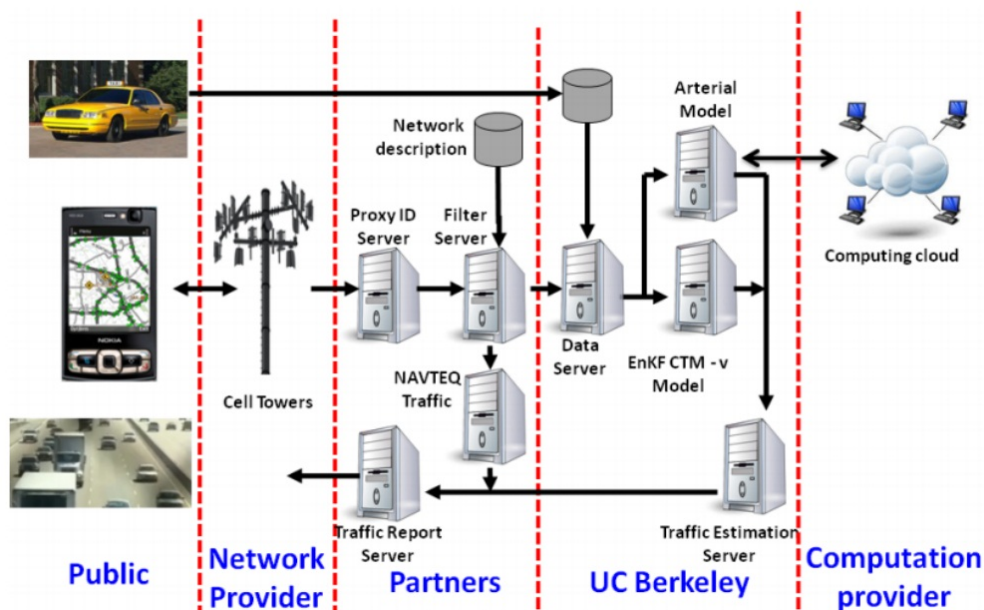
### 7.2 Το σύστημα του Mobile Millennium Project

Το mobile millennium project αποτελεί έναν πλήρη σύστημα για τη λήψη δεδομένων από τους αισθητήρες, που αφού φιλτραριστούνε διανέμονται σε εξυπηρετητές. Κατόπιν, τα επεξεργασμένα αποτελέσματα εμφανίζονται σε πραγματικό χρόνο στους χρήστες.



### 7.2.1 Δομή της Αρχιτεκτονικής

Το σύστημα χρησιμοποιεί ως είσοδο τα δεδομένα, τα οποία προέρχονται από τα τηλέφωνα των χρηστών που διαθέτουν αισθητήρα GPS. Επίσης, αποτελείται και από δύο παρόχους πληροφοριών. Μια εταιρεία κινητής τηλεφωνίας, που είναι ο παροχέας του δικτύου, ώστε να γίνεται η συγκέντρωση των δεδομένων από τα κινητά τηλέφωνα, και της κύριας χαρτογραφικής εταιρίας στις Η.Π.Α. για την εκτίμηση της κυκλοφορίας (Berkeley / Navteq). Η διαδικασία που ακολουθείται, απεικονίζεται στο ΣΧΗΜΑ: 7.1 [21].



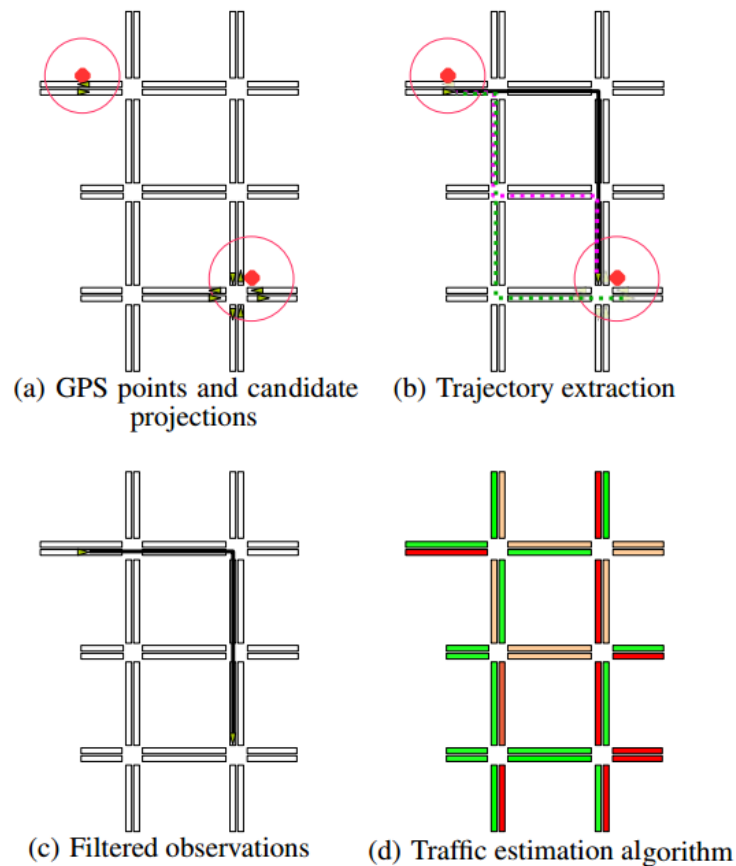
ΣΧΗΜΑ 7.1: Αρχιτεκτονική του mobile millennium συστήματος [21].

Σε κάθε κινητή συσκευή που συμμετέχει, η εφαρμογή εκτελείται και είναι υπεύθυνη για τη συλλογή των δεδομένων κίνησης. Η τεχνική που εφαρμόζεται για την συλλογή των χωρικών δεδομένων βασίζεται στην τεχνική των *Virtual Trip Lines* [18]. Η τεχνική αυτή είναι σημαντική, διότι έχει επίγνωση της χωρικής δειγματοληψίας. Η εφαρμογή κατόπιν εμφανίζει τα αποτελέσματα των εκτιμήσεων της κυκλοφορίας, τα οποία παράγονται από τα συνολικά δεδομένα όλων των συμμετεχόντων [17]. Να σημειωθεί ότι το λογισμικό που χρησιμοποιεί η εν λόγω εργασία, είναι γραμμένο σε κώδικα Java.

### 7.2.2 Λειτουργία του Συστήματος

Η λειτουργίες του συστήματος παρουσιάζεται στο ΣΧΗΜΑ: 7.2

- Γίνεται χαρτογράφηση κάθε σημείου των ακατέργαστων, και ίσως θορυβωδών, δεδομένων που συλλέγει ο GPS αισθητήρας, κοντά σε μια συλλογή από κοντινές, υποψήφιες προβολές επί του οδικού δικτύου (ΣΧΗΜΑ: 7.2a).



ΣΧΗΜΑ 7.2: Διαδικασία που ακολουθείται για την εκτίμηση της κυκλοφορίας [21].

- Για κάθε όχημα, κατασκευάζεται η πιο πιθανή τροχιά που μπορεί να ακολουθήσει (ΣΧΗΜΑ: 7.2b).
- Κάθε τμήμα της τροχιάς μεταξύ δύο σημείων του GPS, αναφέρεται ως μια παρατήρηση (ΣΧΗΜΑ: 7.2c). Μια παρατήρηση περιλαμβάνει ένα χρόνο έναρξης, ένα χρόνο λήξης και μια διαδρομή επί του οδικού δικτύου.
- Οι παρατηρήσεις ομαδοποιούνται με βάση τα χρονικά διαστήματα και αποστέλλονται σε μια μηχανή εκτίμησης της κυκλοφορίας, η οποία κάνει χρήση του [αλγόριθμου εκτίμησης της κυκλοφορίας](#). Κατόπιν επιστρέφει τις κατανομές των χρόνων μετακίνησης για κάθε σύνδεσμο. (ΣΧΗΜΑ: 7.2d).
- Οι κατανομές του χρόνου της διαδρομής στη συνέχεια αποθηκεύονται και μεταδίδονται στους χρήστες, μέσω μιας διαδικτυακής διεπαφής του παγκόσμιου ιστού, όπως φαίνεται στο ΣΧΗΜΑ: 7.3 ή στα τηλέφωνα τους.



ΣΧΗΜΑ 7.3: Απεικόνιση της εξόδου του προσομοιωτή της εκτίμησης της κυκλοφορίας [21].

### 7.2.3 Virtual Trip Line (VTL)

Οι Virtual Trip Lines (VTL's) είναι εικονικές διαδρομές που δημιουργούνται στις οδικές αρτηρίες, και αποτελούν βασική λειτουργία του συστήματος. Η λειτουργία τους βασίζεται στο γεγονός ότι το τηλέφωνο εντοπίζει και ανανεώνει τη τοποθεσία που του υποδεικνύει ο αισθητήρας GPS, κάθε λίγα δευτερόλεπτα. Έτσι το τηλέφωνο πραγματοποιεί λήψη μιας λίστας από VTLs στην ευρύτερη περιοχή όπου βρίσκεται. Όταν ο χρήστης διασχίσει μια από τις VTLs, το τηλέφωνο στέλνει μια ενημερωμένη έκδοση των δεδομένων του, στον κεντρικό VTL server αναφέροντας την ταχύτητα, το χρόνο διέλευσης, καθώς επίσης και το χρόνο της διαδρομής από το προηγούμενο VTL που είχε βρεθεί. Η ακρίβεια των παραγομένων δεδομένων, τα οποία παράγονται με συχνή GPS δειγματοληψία, ποικίλλει σε μεγάλο βαθμό. Αυτό οφείλεται στον τύπο του αισθητήρα του GPS του τηλεφώνου [1]. Μια πειραματική ανάπτυξη των VTL's απεικονίζεται στο ΣΧΗΜΑ: 7.4.

Είναι γενικά αρκετά ακριβής για την εκτίμηση της κυκλοφορίας σε κεντρικές οδικές αρτηρίες, όταν φιλτραριστούν σωστά τα δεδομένα. Ακόμα είναι ικανές να παρέχουν υψηλής ποιότητας δεδομένα κίνησης.

## 7.3 Αλγόριθμος Εκτίμησης της Κυκλοφορίας

Η εκτίμηση της ποιότητας της κυκλοφορίας αποτελεί απαραίτητη λειτουργία του συστήματος, ώστε να συναχθεί το συμπέρασμα πόση συμφόρηση παρουσιάζουν οι συνδέσεις σε ένα οδικό δίκτυο. Το οδικό δίκτυο μοντελοποιείται ως ένα γράφημα  $(V, E)$  όπου  $V$  είναι οι κορυφές (δηλαδή οι διασταυρώσεις) και  $E$  είναι οι ακμές (δηλαδή οι οδικές αρτηρίες). Για κάθε ακμή  $e \in E$ , εξάγεται ο χρόνος που χρειάζεται για να διαπεραστεί αυτή η ακμή ως κατανομή πιθανότητας. Για να υπάρχει η δυνατότητα εύκολης χρήσης των εξαγόμενων αποτελεσμάτων, γίνεται μοντελοποίηση της σύνδεσης, τόσες φορές όσες γίνει η διαπέραση για κάθε ακμή  $e$ ,

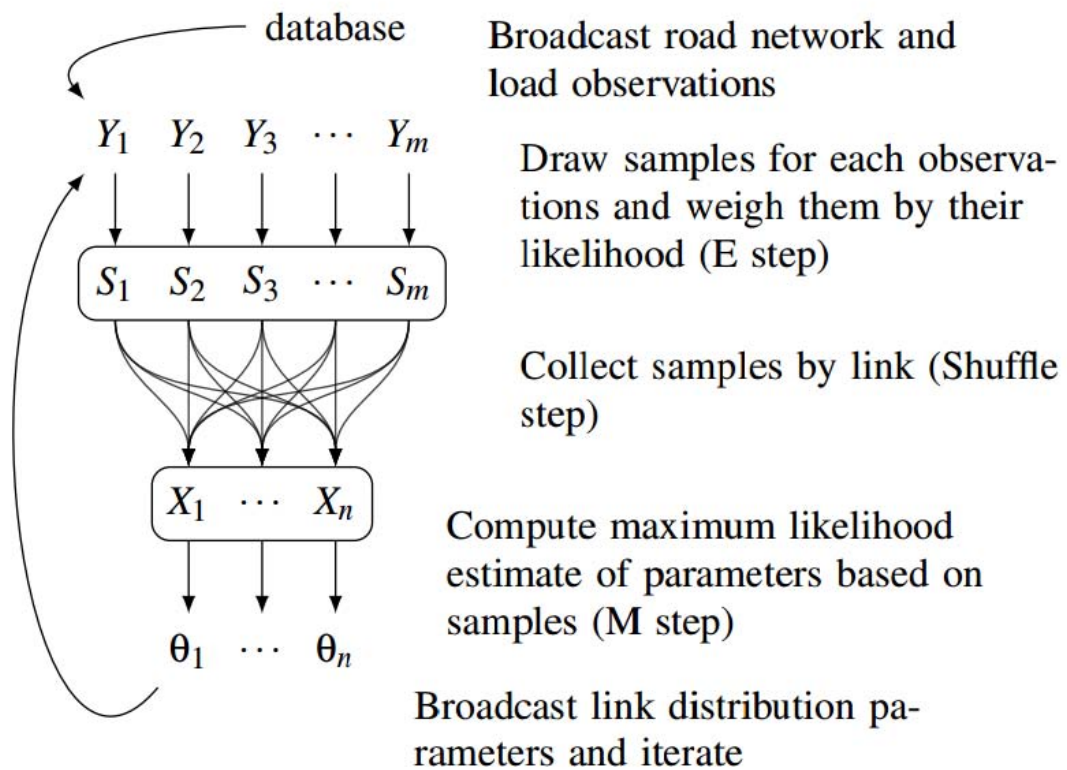


ΣΧΗΜΑ 7.4: Απεικόνιση των Virtual Trip Lines ως μέρος του mobile millennium συστήματος [1].

ως ανεξάρτητη κατανομή Gamma με παραμέτρους  $\theta_e$ . Τα  $\theta_e$  αντιπροσωπεύουν τις δύο τιμές που παραμετροποιούν μια κατανομή Gamma.

Ο αλγόριθμος που χρησιμοποιείται είναι ο **Expectation Maximization (EM)** [11, 32]. Ο αλγόριθμος μοντελοποιεί το πρόβλημα της εκτίμησης, έχοντας ως είσοδο το γράφημα  $(V, E)$  καθώς και τις παρατηρηθείσες διαδρομές  $Y$  που παράγονται από τα δείγματα του αισθητήρα GPS. Κάθε παρατήρηση  $Y_i$  περιγράφει το χρόνο μετακίνησης, και τη διαδρομή που ακολουθεί η  $i$  τροχιά, η οποία αποτελείται από μια ή περισσότερες οδικές συνδέσεις. Επίσης προσδιορίζεται και ο συνολικός χρόνος μιας διαδρομής. Τα όρια ταχύτητας και τα μήκη σύνδεσης των οδικών αρτηριών, λαμβάνονται επίσης υπ' όψιν.

Η κεντρική ιδέα του αλγορίθμου είναι να εφαρμοστεί τυχαία κατάτμηση του συνολικού χρόνου της διαδρομής μεταξύ των ακμών, για κάθε παρατήρηση. Για να επιτευχθεί ο στόχος, γίνεται αντιστοίχιση των κατατμήσεων με τις πιθανότητες τους, σύμφωνα με την τρέχουσα, εκείνη την στιγμή, εκτίμηση των κατανομών του χρόνου μετακίνησης. Κατόπιν, με βάση τα αντιστοιχισμένα δείγματα του χρόνου μετακίνησης που παράγονται για κάθε ακμή, ενημερώνονται οι παράμετροι κατανομής του χρόνου διαδρομής, ώστε να μεγιστοποιηθεί η πιθανότητα αυτών των δειγμάτων. Με την διαδοχικά επαναλαμβανόμενη επανάληψη αυτής της διαδικασίας, ο αλγόριθμος συγκλίνει σε ένα σύνολο παραμέτρων κατανομής του χρόνου διαδρομής, που ταιριάζει επιτυχώς τα δεδομένα. Το στάδιο παραγωγής των δειγμάτων καλείται *Expectation (E)* βήμα, ενώ το στάδιο της ενημέρωσης των παραμέτρων είναι το *Maximization (M)* βήμα [1].



ΣΧΗΜΑ 7.5: Διάγραμμα ροής της λειτουργίας του EM αλγορίθμου [21].

Στο ΣΧΗΜΑ: 7.5 απεικονίζεται η ροή των δεδομένων, με βάση τον EM αλγόριθμο. Στο  $E$ -βήμα παράγονται για κάθε ακμή δείγματα του χρόνου της διαδρομής, από το σύνολο όλων των διαδρομών.

Ειδικότερα, για κάθε παρατήρηση, παράγεται ένα σύνολο δειγμάτων  $S_i = \{(s_{e_i}, w_{e_i})\}_{e_i \in Y_i}$  με τυχαίο τρόπο, διαιρώντας τους παρατηρηθέντες χρόνους διαδρομής των  $Y_i$  με τις πιθανές ακμές. Γίνεται έτσι η παραγωγή ενός χρόνου διαδρομής  $s_{e_i}$  και ανατίθεται ένα βάρος  $w_{e_i}$ , ως η πιθανότητα του χρόνου διαδρομής  $s_{e_i}$ , σύμφωνα με την  $e$  κατανομή του χρόνου διαδρομής  $\theta_e$ .

Στο στάδιο shuffle, ομαδοποιούνται τα δείγματα του  $S_i$  συνόλου κατά ακμές, έτσι ώστε κάθε  $e$  ακμή να έχει  $X_e = \{(s_{e_i}, w_{e_i})\}$  δείγματα, από όλες τις τροχιές.

Στο  $M$ -βήμα υπολογίζουμε ξανά τις παραμέτρους  $\theta_e$  ώστε να επιτευχθεί ταίριασμα των  $e$  ακμών, με τις κατανομές του χρόνου διαδρομής των δειγμάτων  $X_e$ .

Ο αλγόριθμος χρειάζεται να γνωρίζει τον αριθμό των δειγμάτων, ώστε να λειτουργήσει σωστά.

## 7.4 Επίλογος

Στο παρόν κεφάλαιο έγινε αναφορά στο mobile millennium σύστημα. Τα δεδομένα που συλλέγονται μέσω του αισθητήρα GPS, χρησιμοποιούνται για να γίνει πρόβλεψη της κυκλοφορίας. Ο συνδυασμός του cloud computing και του EM αλγορίθμου, μειώνει σημαντικά τον χρόνο υπολογισμού. Έτσι, το εξαγόμενο αποτέλεσμα που προκύπτει, πλησιάζει τον αληθινή κίνηση που επικρατεί στα οδικά τμήματα.

## Κεφάλαιο 8

# Crowdsourcing εφαρμογές των τεσσάρων κύριων τομέων

### 8.1 Εισαγωγή

Όπως έχει αναφερθεί και στο [κεφάλαιο 4](#), οι crowdsourcing εφαρμογές δύναται να επεξεργαστούν δεδομένα από όλους τους τύπους των αισθητήρων. Στο παρόν κεφάλαιο παρουσιάζονται εφαρμογές που έχουν αναπτυχθεί, και κατηγοριοποιούνται στις ακόλουθες ενότητες.

### 8.2 Εφαρμογές Παρακολούθησης της Υγείας

Για την προσωπική παρακολούθηση της υγείας, τα τηλέφωνα με τους εγκατεστημένους αισθητήρες, χρησιμοποιούνται για την παρακολούθηση της φυσικής κατάστασης και της υγείας των ασθενών. Η εφαρμογή *DietSense* βοηθάει τους συμμετέχοντες που θέλουν να χάσουν βάρος, με την καταγραφή των διατροφικών τους επιλογών μέσα από εικόνες και ηχητικά δείγματα. Τα τηλέφωνα κάνουν λήψη εικόνων των γευμάτων. Με αυτόν τον τρόπο, γίνεται υπολογισμός των θερμίδων που θα λάβει ο οργανισμός των συμμετεχόντων. Επίσης γίνεται καταγραφή της μέρας και της ώρας του φαγητού, καθώς επίσης και της τοποθεσίας και πιθανών ηχητικών δειγμάτων. Η τοποθεσία και τα ηχητικά δείγματα είναι ιδιαίτερα σημαντικά, διότι προσδιορίζεται εάν το γεύμα λαμβάνεται σε κάποιο εστιατόριο ή στο σπίτι των συμμετεχόντων. Όλα τα δεδομένα που συλλέγονται, στέλνονται σε έναν προσωπικό χώρο αποθήκευσης, και με αυτόν τον τρόπο ο προσωπικός ιατρός ελέγχει την διατροφή των συμμετεχόντων.

Το σύστημα *Balance* συνδυάζει, μια διαδραστική φόρμα εισαγωγής δεδομένων των θερμίδων των γευμάτων, με έναν αισθητήρα, στο πρότυπο του επιταχυνσιομέτρου, ο οποίος βρίσκεται στην περιοχή του σώματος και καταγράφει την δραστηριότητα του χρήστη. Το εν λόγω σύστημα βασίζεται στην ανάλυση που πραγματοποιεί ο αισθητήρας, για να διαπιστωθεί η δραστηριότητα του χρήστη. Για παράδειγμα, εάν κινείται με ποδήλατο ή εάν κάνει βόδινη. Συσχετίζοντας τον τύπο των δραστηριοτήτων με την αντίστοιχη διάρκεια, γίνεται εκτίμηση της

κατανάλωσης των θερμίδων των χρηστών. Το σύστημα *Jog Falls* είναι και αυτό ένα σύστημα παρακολούθησης της διατροφής των χρηστών. Τα συλλεγόμενα δεδομένα των θερμίδων των γευμάτων και της λειτουργίας του οργανισμού, συνδυάζονται ως μια εφαρμογή στο κινητό τηλέφωνο. Έχοντας ως δεδομένα την αρτηριακή πίεση και την μέτρηση του βάρους, οι διατολόγοι ενημερώνονται για τα συνολικά επιτεύγματα, όσον αφορά τη διατροφή και την στοχευόμενη απώλεια βάρους.

Μια άλλη εφαρμογή που παρέχει πληροφορίες σχετικά με την υγεία των χρηστών είναι η *MobAsthma*. Η εφαρμογή αυτή ελέγχει το επίπεδο του άσθματος των ασθενών, και την έκθεσή τους σε ρύπανση. Μέσω σύνδεσης Bluetooth είναι συνδεδεμένοι με το κινητό τηλέφωνο, ένας μετρητής και ένας αισθητήρας μέτρησης της ρύπανσης. Στόχος είναι να πραγματοποιείται τόσο μέτρηση του όγκου του αέρα που εισπνέεται, όσο και μέτρηση της συγκέντρωσης των αερομεταφερόμενων σωματιδίων. Οι μετρήσεις αυτές σε συνδυασμό με την τοποθεσία των χρηστών, τίθενται στην διάθεση έμπειρων γιατρών, ώστε να γίνουν οι κατάλληλες θεραπείες αντιμετώπισής του.

Φυσικά, εκτός από την ένδειξη του επιπέδου του άσθματος, μια άλλη εφαρμογή επιτρέπει στους χρήστες να χρησιμοποιούν το κινητό τους τηλέφωνο για να καθοριστεί η έκθεσή τους σε περιβαλλοντικούς ρύπους. Αυτή η εφαρμογή ονομάζεται *PEIR* (Personal Environmental Impact Report). Μέσω του αισθητήρα GPS καθορίζεται η τοποθεσία του χρήστη, καθώς επίσης και το μέσο μεταφοράς που χρησιμοποιείται. Κατόπιν τα δεδομένα στέλνονται σε έναν κεντρικό εξυπηρετητή. Ο εξυπηρετητής αφού επεξεργαστεί τα δεδομένα, στέλνει πίσω στους χρήστες πληροφορίες που αφορούν τις περιβαλλοντικές επιπτώσεις, όσον αφορά τα επίπεδα του διοξειδίου του άνθρακα και των εκπεμπόμενων σωματιδίων. Με αυτήν την μέθοδο οι χρήστες ενημερώνονται για τις τοποθεσίες που παρουσιάζουν υψηλά επίπεδα ρύπανσης, και έχουν την ευελιξία να τις αποφύγουν. Ο τρόπος μεταφοράς συνάγεται με ενδείξεις από το επιταχυνσιόμετρο. Επιπρόσθετες παράμετροι εισόδου και μοντέλα λαμβάνονται υπόψη για τον καθορισμό των περιβαλλοντικών παραγόντων, όπως είναι οι καιρικές συνθήκες ή μοντέλα ροής της οδικής κυκλοφορίας [10, 27].

### 8.3 Εφαρμογές Παρακολούθησης του Περιβάλλοντος

Στην εφαρμογή *Haze Watch* τα κινητά τηλέφωνα είναι συνδεδεμένα με εξωτερικούς αισθητήρες μέτρησης της ρύπανσης, προκειμένου να μετρηθεί η συγκέντρωση του μονοξειδίου του άνθρακα, το όζον, το διοξείδιο του θείου, το άζωτο, καθώς επίσης και η συγκέντρωση του διοξειδίου στον αέρα. Σε σύγκριση με μετεωρολογικούς σταθμούς, τα κινητά τηλέφωνα μπορεί να συλλέξουν λιγότερο ακριβείς μετρήσεις. Ωστόσο, η κινητικότητα που διαθέτουν, τους επιτρέπει να παρατηρούν απρόβλεπτα γεγονότα, για παράδειγμα μια τυχαία ρύπανση η οποία σπάνια μπορεί να ανιχνευτεί από στατικούς σταθμούς, και να παρέχουν μεγάλη γεωγραφική



κάλυψη. Η διαφορά της εφαρμογής *Haze Watch* με την εφαρμογή *PEIR*, είναι ότι οι συγκεντρώσεις των ρύπων ανιχνεύονται σε πραγματικό χρόνο από τους αισθητήρες, και δεν προκύπτουν με βάση τις καιρικές συνθήκες.

Η εφαρμογή *Ikarus* είναι και αυτή μια εφαρμογή που μελετά το περιβάλλον, η οποία απευθύνεται κυρίως στους αλεξιπτωτιστές. Συλλέγει πληροφορίες σχετικά με τα θερμά ρεύματα του αέρα, οι οποίες χρησιμοποιούνται από τους αλεξιπτωτιστές προκειμένου να αποκτήσουν υψόμετρο κατά την διάρκεια της πτήση τους. Η εφαρμογή δημιουργεί θερμικούς χάρτες, από τα δεδομένα των αισθητήρων μέτρησης της βαρομετρικής πίεσης και του GPS.

## 8.4 Εφαρμογές Κοινωνικής Δικτύωσης

Υπάρχουν όμως και εφαρμογές που κάνουν χρήση των δεδομένων που λαμβάνονται από τους αισθητήρες, για να εμπλουτίσουν το περιεχόμενο που μοιράζεται σε μέσα κοινωνικής δικτύωσης, όπως είναι τα blogs. Η εφαρμογή *Micro-Blog* προτείνεται για την δημιουργία μιας “εικονικής πλατφόρμας πληροφοριών”, η οποία θα παρέχει μια υψηλή ανάλυση του κόσμου, με την αξιοποίηση των κινητών τηλεφώνων. Οι χρήστες δύναται να δημιουργήσουν καταχωρήσεις με γεωγραφικές ετικέτες, και να τις ενισχύσουν με πληροφορίες πολυμέσων. Για παράδειγμα, με αρχεία ήχου, με εικόνες, ή με δεδομένα του επιταχυνσιόμετρου που συλλέγονται μέσω των κινητών τηλεφώνων. Οι καταχωρήσεις στην συνέχεια φορτώνονται σε έναν διακομιστή, ο οποίος μπορεί να τις τοποθετήσει με έναν παγκόσμιο χάρτη, με βάση την γεωγραφική ετικέτα. Ο χάρτης αυτός είναι προσπελάσιμος από το πλήθος που χρησιμοποιεί την εφαρμογή. Έτσι οι χρήστες μπορούν να αναζητήσουν πληροφορίες για συγκεκριμένα σημεία που τους ενδιαφέρουν. Σε περίπτωση που δεν εντοπιστεί η πληροφορία που αναζητούν, στέλνουν το αντίστοιχο ερώτημα στον διακομιστή, και αυτός το αναμεταδίδει στους χρήστες που εντοπίζονται, για να το εκτελέσουν.

Παρόμοια, η εφαρμογή *CenceMe* ενοποιεί εικονικές αναπαραστάσεις της κατάστασης που βρίσκονται οι χρήστες με το περιεχόμενο των κοινωνικών δικτύων. Οι πολυτροπικές πληροφορίες (όπως είναι τα ηχητικά δείγματα, οι φωτογραφίες, ή η τοποθεσία) που λαμβάνονται από τα κινητά τηλέφωνα, συνδυάζονται με τις πληροφορίες που αναφέρονται στα κοινωνικά δίκτυα, όπως για παράδειγμα η διάθεση του χρήστη ή για την δραστηριότητα που εκτελεί. Σαν αποτέλεσμα παράγεται κάποιο μήνυμα κατάστασης ή μια εικονική αναπαράσταση της διάθεσης των χρηστών.

### Εφαρμογή Ελέγχου Τιμών

Μια εφαρμογή η οποία πραγματοποιεί καταγραφή της τιμής των καυσίμων, ονομάζεται *PetrolWatch*. Το κινητό τηλέφωνο τοποθετείται σε σημείο που η κάμερα μπορεί να τραβήξει φωτογραφίες από τους πίνακες των τιμών των πρατηρίων καυσίμων, κάνοντας επίσης χρήση και του αισθητήρα GPS για προσδιορισμό της τοποθεσίας. Οι εικόνες στη συνέχεια στέλνονται

σε έναν εξυπηρετητή, ο οποίος είναι υπεύθυνος για την επεξεργασία τους, και την ανίχνευση της ζητούμενης τιμής. Καθώς κάθε πρατήριο καυσίμων ανήκει σε διαφορετική εταιρία, το σήμα της εταιρίας αφαιρείται από την εικόνα, ώστε να μειωθεί η πολυπλοκότητα της επεξεργασίας της εικόνας. Με αυτήν την τεχνική, και αφού αποκρυπτογραφηθεί η εικόνα, ο εξυπηρετητής αποθηκεύει σε μια βάση δεδομένων τις τιμές. Έτσι ο χρήστης μπορεί να αναζητήσει το φθηνότερο καύσιμο κοντά στην περιοχή που βρίσκεται [10, 27].

## 8.5 Εφαρμογές Παρακολούθησης της Οδικής Κυκλοφορίας

Όσον αφορά για εφαρμογές που παρακολουθούν την οδική κυκλοφορία, έχουν αναφερθεί στα προηγούμενα κεφάλαια η εφαρμογή VTrack και η εφαρμογή Mobile Millennium. Είναι δύο εφαρμογές που θεωρούνται ως οι καλύτερες που έχουν αναπτυχθεί. Μια επέκταση των εφαρμογών αυτών, είναι η εφαρμογή *Transit Tracking*. Η έλλειψη ενημέρωσης του πλήθους, σχετικά με την περιοχή που βρίσκονται τα μέσα μαζικής μεταφοράς, αποτελεί σημαντικό πρόβλημα. Κάνοντας χρήση των τεχνολογιών που έχουν αναπτυχθεί, η εφαρμογή εκτελείται στα κινητά τηλέφωνα των επιβατών, και έτσι τα οχήματα εντοπίζονται και απεικονίζονται σε έναν διαδραστικό χάρτη [10, 27].

**Μέρος III**

**ΕΠΙΛΟΓΟΣ**

## Κεφάλαιο 9

# Επίλογος και Μελλοντικές Έρευνες

### 9.1 Επίλογος

Στην παρούσα διατριβή παρουσιάστηκε ο όρος crowdsourcing. Ο συνδυασμός της γνώσης του πλήθους με την τεχνολογική ανάπτυξη των κινητών συσκευών, προσφέρει την δυνατότητα στους χρήστες να επικοινωνούν μέσω των εφαρμογών. Εφαρμογές όπως είναι το Foursquare, παρουσιάζουν την τοποθεσία του κάθε ατόμου στους φίλους του. Το Twitter ή το Facebook, τα οποία είναι δίκτυα κοινωνικής δικτύωσης, είναι και αυτές crowdsourcing εφαρμογές, καθώς με βάση την ανατιθέμενη IP που έχει κάθε συσκευή ή με χρήση του αισθητήρα GPS, μπορεί να εντοπιστεί η γεωγραφική τοποθεσία του χρήστη. Έτσι πραγματοποιείται διαμοιρασμών δεδομένων μεταξύ των χρηστών, και υπάρχει η δυνατότητα να εντοπιστούν λύσεις σε πιθανά προβλήματα.

Η βασική λειτουργία του crowdsourcing εντοπίζεται στην ανάπτυξη εφαρμογών για τον εντοπισμό της κίνησης των κεντρικών αρτηριών των πόλεων. Για το πώς αναπτύσσονται οι εφαρμογές τέτοιου τύπου, γίνεται παρουσίαση των project VTrack και Mobile Millennium. Πρόσφατα, οι ερευνητές απέδειξαν ότι η εκτίμηση των συνθηκών κυκλοφορίας στους αυτοκινητοδρόμους, μπορεί να γίνει μέσω μόνο των δεδομένων των GPS των οχημάτων. Μεγάλοι αστικοί δρόμοι των πόλεων που συνδέουν τα πληθυσμιακά κέντρα, παρέχουν πρόσθετες προκλήσεις για την εκτίμηση της κυκλοφορίας. Ειδικοί αισθητήρες κίνησης είναι ακριβοί, για την εγκατάσταση, την συντήρηση και την λειτουργία τους. Η έλλειψη κάλυψης αισθητήρων σε ολόκληρο το αρτηριακό δίκτυο παρακινεί έτσι τη χρήση των δεδομένων των GPS των οχημάτων για την εκτίμηση των συνθηκών κυκλοφορίας [21].

### 9.2 Μελλοντικές έρευνες

Όλο και περισσότεροι αισθητήρες γίνονται μέρος των κινητών τηλεφώνων. Τρεις σημαντικοί αισθητήρες, που δεν έχουν ακόμα εγκατασταθεί με το κινητό τηλέφωνο είναι το βαρόμετρο, ο δείκτης της θερμοκρασίας, καθώς επίσης και ο αισθητήρας για την υγρασία. Ο καθένας

από αυτούς μπορεί να συμβάλλει σημαντικά σε crowdsourcing εφαρμογές. Για παράδειγμα, ο συνδυασμός του επιταχυνσιομέτρου με το βαρόμετρο καθιστά εύκολο τον προσδιορισμό της κατεύθυνσης. Δηλαδή όταν κάποιος χρήστης κάνει βάδην, ή όταν ανεβαίνει σκάλες, μπορεί να προσδιοριστεί η κατεύθυνση που έχει. Όσο ο παράγοντας του κόστους μειώνεται, είναι πιθανό να προστεθούν περισσότεροι αισθητήρες στα τηλέφωνα [27].

Τα μελλοντικά συστήματα ανίχνευσης, θα δύνανται να εντοπίζουν τα πάντα όσων αφορά την προσωπική αίσθηση. Ταυτόχρονα, οι ερευνητές εξετάζουν κατά πόσο ο χρήστης, θα πρέπει να συμμετέχει ενεργά κατά τη διάρκεια της δραστηριότητας της ανίχνευσης (για παράδειγμα, να λαμβάνει το τηλέφωνο για να συλλέξει ένα δείγμα ήχου ή να κάνει λήψη φωτογραφιών).

Όσον αφορά το ευαίσθητο θέμα της ασφάλειας των δεδομένων των χρηστών, το πιο δύσκολο ζήτημα είναι η εφαρμογή της ανωνυμίας στην τρέχουσα αρχιτεκτονική του εξυπηρετητή-χρήστη. Για να διατηρηθεί η ανωνυμία στους μεμονωμένους χρήστες, ο εξυπηρετητής που κάνει χρήση της τοποθεσίας, πρέπει να έχει επαρκή αριθμό χρηστών σε μια γεωγραφική περιοχή. Οι αλγόριθμοι που χρησιμοποιούνται είναι αποτελεσματικοί, και θα μπορούσαν να εκτελεστούν σε μια ασύρματη συσκευή. Ωστόσο, απαιτούν πληροφορίες θέσης από διαφορετικές συσκευές στην περιοχή που βρίσκονται, προκειμένου να κριθεί η ποσότητα και η πυκνότητα των συσκευών. Επομένως, ένα “peer-to-peer” ανώνυμο γεωγραφικό σύστημα, αποτελεί μια εύρωστη λύση του προβλήματος.

Η ανάπτυξη των εφαρμογών VTrack και Mobile Millennium, αποτέλεσε το έναυσμα για την δημιουργία νέων εφαρμογών. Η εύρεση της ποιότητας του περιβάλλοντος [29], ή η βελτίωση της προσωπικής υγείας, είναι δυο τομείς όπου οι crowdsourcing εφαρμογές έχουν εξελιχθεί απαραίτητες. Ωστόσο μια σημαντική βελτίωση σε όλα τα συστήματα θα μπορούσε να ήταν η καταγραφή των πληροφοριών σε μια βάση δεδομένων, και να γίνεται πρόβλεψη των μελλοντικών τιμών. Έτσι θα μειώνονταν σημαντικά το χρονικό διάστημα επεξεργασίας, και εμφάνισης των συνθηκών που επικρατούν στην εκάστοτε περιοχή.

Εξαιτίας της αυξημένης χρήσης και παροχής του διαδικτύου σε κάθε σπίτι, υπάρχει η δυνατότητα να δημιουργηθεί ένα ασύρματο δίκτυο, σε όλες τις περιοχές. Με αυτόν τον τρόπο, θα υπάρχει σχεδόν παντού πρόσβαση σε όλες τις υπηρεσίες του διαδικτύου. Επιπρόσθετα, υλοποιώντας υπάρχοντες αλγόριθμους, υπάρχει η δυνατότητα να γίνει χαρτογράφηση των περιοχών αυτών. Έτσι το πλήθος χωρίς να αντιμετωπίζει προβλήματα συνδεσιμότητας, μπορεί να μοιράζεται δεδομένα με όλους τους συνδεδεμένους χρήστες.

## **Μέρος IV**

# **ΒΙΒΛΙΟΓΡΑΦΙΑ**

# Βιβλιογραφία

- [1] Alexander M. Bayen, Joe Butler, and Anthony D. Patire. Mobile Millennium Final Report. Technical Report Documentation Page CA11-1913, September 2011.
- [2] Ankit Sharma. Crowdsourcing Critical Success Factor Model, 2010.
- [3] Arvind Thiagarajan, Lenin Ravindranath Sivalingam, Katrina LaCurts, Sivan Toledo, Jakob Eriksson, Samuel Madden and Hari Balakrishnan. VTrack: Accurate, Energy-Aware Traffic Delay Estimation Using Mobile Phones. *In 7th ACM Conference on Embedded Networked Sensor Systems (SenSys)*, Berkeley, CA, November 2009.
- [4] Beresford A. and Stajano F. Location Privacy in Pervasive Computing. *IEEE Pervasive Computing*, 2(01):46–55, 2003.
- [5] Bozanis P. *Algorithms*. Tziola Publishing, 2003.
- [6] Calabrese F., Colonna M., Lovisolo P., Parata D. and Ratti, C. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *Intelligent Transportation Systems, IEEE Transactions on*, 12(01):141–151, New York, NY, USA, 2009. ACM.
- [7] Chatzimilioudis G., Konstantinidis A., Laoudias C., and Zeinalipour-Yazti D. Crowdsourcing with Smartphones. *Internet Computing, IEEE*, 16(5):36–44, 2012.
- [8] Costas Pontikakos, Thomas Glezakos and Theodore Tsiligiridis. Location-based services: architecture overview. 2005.
- [9] Daren C. Brabham. Crowdsourcing: A Model for Leveraging Online Communities. *UNC School of Journalism & Mass Communication*, March 2011.
- [10] Delphine Christin, Andreas Reilhardt, Salil S. Kanhere, and Matthias Hollock. A Survey on Privacy in Mobile Participatory Sensing Applications. *Systems and Software*, 84(11):1928–1946, November 2011.
- [11] Dempster A., Laird N., and Rubin D.. Maximum likelihood from incomplete data via the em algorithm. *Royal Statistical Society, Series B (Methodological)*:1–38, 1977.
- [12] Erran Carmel. The New Software Exporting Nations: Success Factors. *EJISDC*, 13(04):1–12, 2003.

- [13] Foursquare <http://www.foursquare.com/>
- [14] Gabriel Ghinita, Keliang Zhao, Dimitris Papadias, and Panos Kalnis. A Reciprocal Framework for Spatial K-Anonymity.
- [15] Ganti R. K., Pham N., Ahmadi N., Nangia S., and Abdelzaher T. F. GreenGPS: A Participatory Sensing Fuel-efficient Maps Application. *Proceedings of the 8th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '10)*, 151–164, June 2010.
- [16] Gruteser M. and Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking. *MobiSys ACM*, 2003.
- [17] Herring R., Hofleitner A, Work D., Tossavainen O.–P., and A. M. Bayen. Mobile Millennium – Participatory Traffic Estimation Using Mobile Phones.
- [18] Hoh B., Gruteser M., Herring R., Ban J., Work D., Herrera J.-C., Bayen A., Annavaram M., and Jacobson Q.. Virtual trip lines for distributed privacy-preserving traffic monitoring. *6th International Conference on Mobile Systems, Applications, and Services*, 15–28, June 2008.
- [19] Howe J. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.
- [20] Hunter T., Herring R., Bayen A., and Abbeel P. Trajectory reconstruction of noisy GPS probe vehicles in arterial traffic. *In preparation for IEEE Transactions on Intelligent Transport Systems*, 2011.
- [21] Hunter T., Moldovan T., Zaharia M., Merzgui S., Ma J., Franklin M. J., Abbeel P., and Bayen A. M. Scaling the mobile millennium system in the cloud. *Proceedings of the 2nd ACM Symposium on Cloud Computing (SOCC 11)*, October 2011.
- [22] Jacob Thebault–Spieker. Crowdsourced Participatory Sensing: applications and motivation of work. January 2012.
- [23] Kang G. Shin, Xiaoen Ju, Zhigang Chen, and Xin Hu. Privacy Protection For Users Of Location-Based Services. *Communications Wireless Magazine, IEEE*, February 2012
- [24] Kazemi L. and Shahabi C. GeoCrowd: Enabling Query Answering with Spatial Crowdsourcing. *GIS '12*, November 2012.
- [25] Kleinberg J. and Tardos E. *Algorithm Design*. Addison-Wesley Longman Publishing Co., Inc., 2005.
- [26] Krumm J., Letchner J., and Horvitz E. Map matching with travel time constraints. *SAE World Congress*, 2007.



- [27] Lane N.D., Miluzzo E., Hong Lu, Peebles D., Choudhury T. and Campbell A.T. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(09):140–150, September 2010.
- [28] Lou Y., Zhang C., Zheng Y., Xie X., Wang W. and Huang Y. Map-Matching for Low-Sampling-Rate GPS Trajectories. *GIS '09*, November 2009.
- [29] Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Peter Boda. Peir, the personal environmental impact report, as a platform for participatory sensing systems research. *Proceedings of the 7th international conference on Mobile systems, applications, and services (MobiSys '09)*, 55–68, June 2011.
- [30] Mohan P., Padmanabhan V., and Ramjee R. Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems (SenSys '08)*, 323–336, November 2008.
- [31] Mokbel Mohamed F., Chow Chi-Yin, and Aref Walid G. The New Casper: query processing for location services without compromising privacy. *Proceedings of the 32nd international conference on Very large data bases*, 763–774, 2006.
- [32] Neal R. and Hinton G. A view of the em algorithm that justifies incremental, sparse, and other variants. *Learning in graphical models*, 89:355–368, 1998.
- [33] Nguyen T. and Poole N. Gathering and Utilizing Crowdsourced Pedestrian Flow Data from Mobile Devices. *CSCI 5980: From GPS & Google Earth to Spatial Computing*
- [34] Nicole A. Ozer, Chris Conley, Hari O'Connell, Ellen Ginsburg and Tamar Gubins. Location-Based Services: Time for a Privacy Check-In. November 2010.
- [35] Rabiner, L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(02):257–286, 1989.
- [36] Sweeney L. k-anonymity: A model for protecting privacy. *Uncertainty, Fuzziness, and* 10(05):557–570, 2002.
- [37] UC Berkeley / Nokia / NAVTEQ. Mobile Millennium. <http://traffic.berkeley.edu/>
- [38] Wikipedia. <http://en.wikipedia.org/>
- [39] Yang D., Xue Guoliang X., Fang X. and Tang J. Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing.
- [40] Zaharia M., Chowdhury M., Franklin M. J., Shenker S., and Stoica I. Spark: Cluster Computing with Working Sets. *In HotCloud*, 2010.