



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΔΙΑ ΒΙΟΥ
ΜΑΘΗΣΗΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΡΕΥΝΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ «ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ»

ΕΘΝΙΚΟ ΣΤΡΑΤΗΓΙΚΟ ΠΛΑΙΣΙΟ ΑΝΑΦΟΡΑΣ
ΕΣΠΑ 2007-2013

ΔΡΑΣΗ «ΑΡΙΣΤΕΙΑ»

Παραδοτέο 4.1 Προδιαγραφές συστήματος ΣΥΣΔΑ

ΑΚΡΩΝΥΜΙΟ/ΚΩΔΙΚΟΣ:	INCEPTION/940
ΚΥΡΙΟΣ ΕΡΕΥΝΗΤΗΣ:	ΓΕΩΡΓΙΟΣ Δ. ΣΤΑΜΟΥΛΗΣ
ΦΟΡΕΑΣ ΥΠΟΔΟΧΗΣ:	ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ (ΟΠΑ)
ΣΥΝΤΑΚΤΕΣ ΠΑΡΑΔΟΤΕΟΥ:	Β. Καλογεράκη (ΟΠΑ), Ι. Μπούτσης (ΟΠΑ), Δ. Γουνόπουλος (ΕΚΠΑ)



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Ευρωπαϊκό Κοινωνικό Ταμείο

Περιεχόμενα

1. ΕΙΣΑΓΩΓΗ	3
2. ΠΡΟΔΙΑΓΡΑΦΕΣ ΥΠΟΔΟΜΗΣ ΕΠΙΚΟΙΝΩΝΙΑΣ ΣΥΣΤΗΜΑΤΟΣ ΣΥΣΔΑ	6
2.1 ΧΡΗΣΤΕΣ	7
2.2 ΔΕΔΟΜΕΝΑ	8
2.2.1 Παραγωγή Δεδομένων	8
2.2.2 Προσθήκη δεδομένων στο MarketBoard	9
2.2.3 Δημοσίευση Αναφορών	11
2.2.4 Ανάκτηση Δεδομένων	11
3. ΑΛΓΟΡΙΘΜΟΣ ΕΠΙΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ ΑΙΣΘΗΤΗΡΩΝ	13
4. ΣΥΝΟΨΗ – ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ	16
5. ΑΝΑΦΟΡΕΣ	17

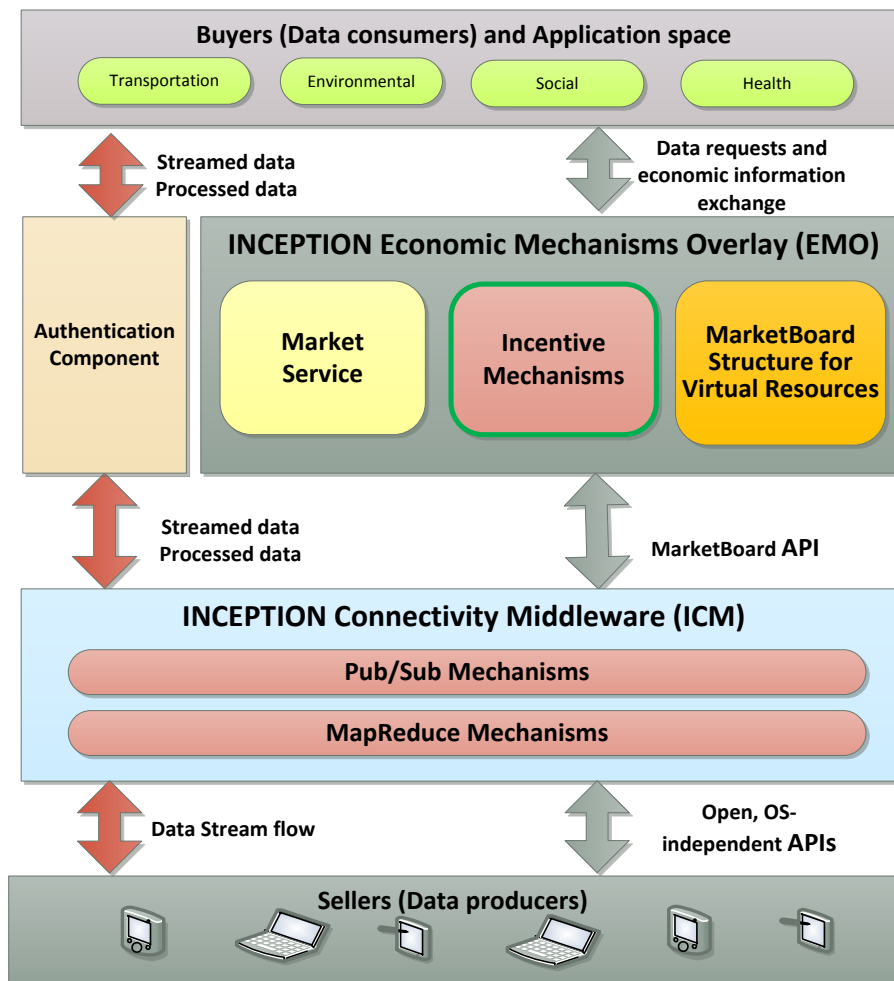
1. Εισαγωγή

Ο στόχος του έργου INCEPTION είναι να διερευνήσει τα κοινωνικοοικονομικά κίνητρα ενός περιβάλλοντος χρηστών που συμμετέχουν σε δίκτυα αισθητήρων, και αλληλεπιδρούν μεταξύ τους και με άλλες κοινωνικές ομάδες με σκοπό να επιτύχουν ανάπτυξη και οικονομική βιωσιμότητα. Στην προσέγγισή μας, οι χρήστες που διαθέτουν αισθητήρες αντιμετωπίζονται ως οικονομικές οντότητες: υφίστανται κόστος όταν συλλέγουν και μεταδίδουν δεδομένα, αλλά ωφελούνται από τη χρήση εφαρμογών που βασίζονται στην συλλογή πληροφοριών που παρέχονται σε συνδυασμό και από τα άλλα μέλη της αντίστοιχης κοινότητας.

Στόχος μας είναι να αναπτύξουμε ένα νέο σύστημα για ΣΥΣΔΑ, δηλ. για «ΣΥμμετοχική Συλλογή Δεδομένων Αισθητήρων», που να προωθεί τη συμβατότητα των κινήτρων για την ανταλλαγή πληροφοριών μεταξύ των χρηστών και των εφαρμογών. Όπως έχει ήδη εξηγηθεί λεπτομερώς στο Παραδοτέο 2.1 του έργου «Απαιτήσεις συστήματος, αρχιτεκτονική και σενάρια αξιολόγησης», οι βασικές απαιτήσεις του παρόντος συστήματος είναι οι εξής

1. Οι εφαρμογές που αξιοποιούν τα δεδομένα που δύνανται να συνεισφέρουν οι χρήστες πρέπει να βασίζονται στα κίνητρα των χρηστών, και συνεπώς, να τους παρέχουν τις κατάλληλες ανταμοιβές, όταν η συμπεριφορά τους στη συγκέντρωση δεδομένων είναι χρήσιμη στους άλλους χρήστες, αυξάνοντας έτσι την οικονομική αποδοτικότητα και τη βελτιστοποίηση των συνολικά διαθέσιμων δεδομένων στο σύστημα, και συντελώντας στην οικονομική βιωσιμότητα του. Οι χρήστες θα πρέπει να ενθαρρύνονται να παρέχουν τα δεδομένα εκείνα που είναι πιο χρήσιμα για τις εφαρμογές και τις κοινωνικές ομάδες που είναι ενεργές, αποφεύγοντας παράλληλα την παραγωγή αυτών που είναι δαπανηρά και λιγότερο χρήσιμα. Αυτή η δυναμική πτυχή της προστιθέμενης αξίας των πληροφοριών είναι ζωτικής σημασίας και μπορεί να αντιμετωπιστεί μόνο μέσα σε ένα οικονομικό πλαίσιο, όπως αυτό που προτείνει το έργο INCEPTION.
2. Σε τεχνολογικό επίπεδο πρέπει το σύστημα να επιτρέπει την αλληλεπίδραση των χρηστών μέσω κινητών συσκευών που χρησιμοποιούν λειτουργικά συστήματα ευρείας χρήσης και αποδοχής όπως το Android. Επίσης, το ενδιάμεσο λογισμικό (βλέπε κατωτέρω) θα πρέπει να παρέχει πρόσβαση σε ένα βασικό αριθμό αισθητήρων που υποστηρίζουν οι σύγχρονες έξυπνες συσκευές, όπως για παράδειγμα πρόσβαση στον αισθητήρα GPS. Επιπρόσθετα, το σύστημα θα πρέπει να εξασφαλίζει την ασφαλή και εγγυημένη ανώνυμη επικοινωνία με τον χρήστη, ώστε να προστατεύει τα προσωπικά του στοιχεία αλλά και τα δεδομένα του από τρίτους, και έτσι να αίρονται τυχόν ενδοιασμοί του χρήστη για την χρήση του συστήματος του INCEPTION.
3. Σε κοινωνικό επίπεδο, η κύρια απαίτηση είναι η προαναφερθείσα για προστασία της ιδιωτικότητας (privacy) του χρήστη, η οποία αποτελεί ένα σημαντικό ζήτημα γενικώς για τα νέα συστήματα επεξεργασίας δεδομένων, και θέτει αντίστοιχες τεχνολογικές απαιτήσεις στην υλοποίησή τους.

Για να επιτευχθούν τα παραπάνω η ερευνητική ομάδα του έργου έχει ήδη προτείνει την υλοποίηση μας αρχιτεκτονικής για το σύστημα ΣΥΣΔΑ η οποία περιγράφηκε στο Παραδοτέο 2.1 «Απαιτήσεις συστήματος, αρχιτεκτονική και σενάρια αξιολόγησης» και παρουσιάζεται στο Σχήμα 1. Σε αυτή παρουσιάζονται επίσης οι διεπαφές του συστήματος και οι αλληλεπιδράσεις μεταξύ των υποσυστημάτων, τα οποία είναι τα εξής:



Σχήμα 1. Η αρχιτεκτονική του INCEPTION

- Υποσύστημα Οικονομικών Μηχανισμών (Economic Mechanisms Overlay), που περιλαμβάνει και εκτελεί τους μηχανισμούς αγοράς και διαχειρίζεται τις περιγραφές των δεδομένων που διατίθενται από χρήστες και τα οποία εκχωρούνται μέσω των μηχανισμών αυτών.
- Ενδιάμεσο Λογισμικό (Connectivity Middleware), το οποίο διαχειρίζεται τα δεδομένα τα διατιθέμενα από χρήστες και αποθηκεύει την διαθεσιμότητά τους.

Σημειωτέον ότι σε μια υλοποίηση της αρχιτεκτονικής αυτής για πρακτική εφαρμογή, τα ανωτέρω υποσυστήματα θα συμπληρώνονταν από το Λογισμικό Αυθεντικοποίησης (authentication component) για πιστοποίηση της αυθεντικότητας του χρήστη (ταυτότητα, δικαιώματα χρήσης κτλ.) σε κάθε συναλλαγή του με το σύστημα, το οποίο όπως εκφεύγει από τα ερευνητικά πλαίσια του έργου.

Συνοπτικά, ένας χρήστης του INCEPTION αλληλεπιδρά με την αρχιτεκτονική σε τρία επίπεδα:

1. Εγγραφή: Ο χρήστης-πάροχος δεδομένων συνδέεται με το σύστημα χρησιμοποιώντας Open APIs (Application Programming Interfaces) μέσω μιας έξυπνης συσκευής. Κατά τη διαδικασία εγγραφής στο σύστημα, ο χρήστης εγκαθιστά το κατάλληλο πρόγραμμα που θα

επιτρέψει τη χρήση του συστήματος.

2. Αλληλεπίδραση με την Υπηρεσία Αγοράς: Ο χρήστης καταχωρεί και περιγράφει τη διαθεσιμότητα των δεδομένων του μέσω του συγκεκριμένου API. Οι περιγραφές στη συνέχεια φορτώνονται στην δομή της Αγοράς (Marketboard), και είναι διαθέσιμες για όλους τους χρήστες του συστήματος.

Οι εγγραφές αυτές μπορούν να ανανεωθούν από τους χρήστες ανάλογα με την διαθεσιμότητα των δεδομένων και τις ανάγκες συναλλαγών των χρηστών. Ο χρήστης μπορεί ο ίδιος να ορίσει μία τιμή για τα δεδομένα, ή να του προτείνει το σύστημα μία με βάση τις άλλες προσφορές. Ένας χρήστης-αγοραστής δεδομένων από την άλλη χρησιμοποιεί το API ως μηχανισμό αναζήτησης για να μάθει τί διαθέσιμα δεδομένα υπάρχουν που ταιριάζουν με τα κριτήρια που έθεσε και ενδεχομένως να τα ζητήσει.

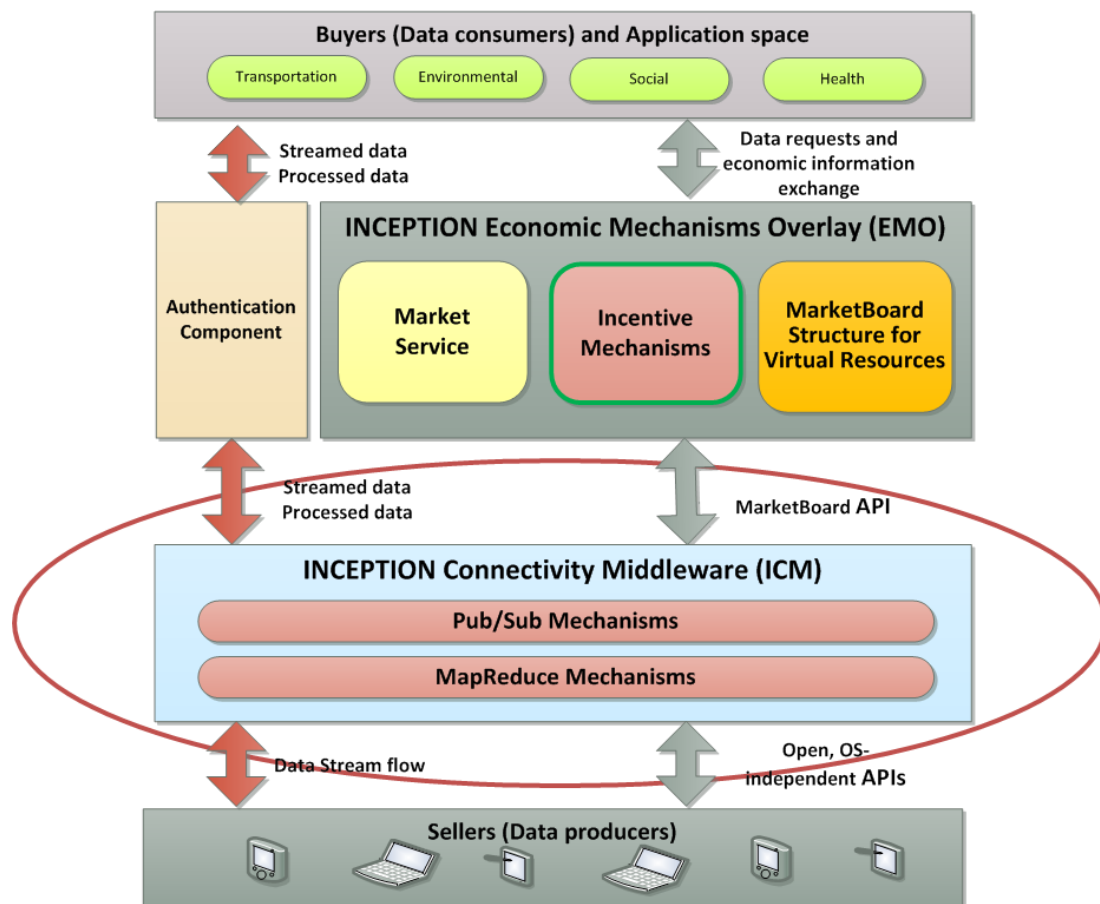
3. Η ανταλλαγή δεδομένων: Όταν μια συναλλαγή αποφασίζεται από την υπηρεσία αγοράς, η πραγματική ανταλλαγή δεδομένων πραγματοποιείται μέσω ενός ενδιάμεσου λογισμικού που υλοποιεί τις συναλλαγές που αποφασίστηκαν.

Σε μια τέτοια δομή αγοράς ωστόσο μπορεί τελικά να συμμετέχει ένα ευρύ φάσμα χρηστών οι οποίοι παράγουν μεγάλες ποσότητες από παροδικές ροές δεδομένων. Μάλιστα αυτή η ευρεία συμμετοχή είναι επιθυμητή, γιατί έτσι το σύστημα παράγει μεγαλύτερη αξία για την κοινωνία των χρηστών. Στις επιμέρους ενότητες που ακολουθούν, θα αναλυθούν σε μεγαλύτερη λεπτομέρεια τα παραπάνω στάδια κατά την περιγραφή των υποσυστημάτων επικοινωνίας της αρχιτεκτονικής του έργου και των προδιαγραφών της, λαμβάνοντας υπ' όψιν και την ανωτέρω απαίτηση για κλιμακωσιμότητα του συστήματος ΣΥΣΔΑ.

Στο παρόν παραδοτέο αναλύουμε την Προδιαγραφές του συστήματος ΣΥΣΔΑ για την Υποδομή Επικοινωνίας του έργου Inception. Περιγράφουμε αναλυτικά τον τρόπο συλλογής, αποθήκευσης και ανάκτησης δεδομένων μαζί με τους μηχανισμούς που χρησιμοποιούνται για την αποδοτική λειτουργία ενός τέτοιου συστήματος. Επίσης, αναφέρουμε την αλληλεπίδραση αυτών των μηχανισμών με τους μηχανισμούς αγοράς για την δημοπράτηση των δεδομένων των χρηστών, οι οποίοι περιγράφονται λεπτομερώς στο Παραδοτέο 3.1 του έργου. Τέλος, παρουσιάζουμε ενδεικτικά έναν αλγόριθμο επιλογής των δεδομένων από τους αισθητήρες που προσφέρει το μεγαλύτερο όφελος στους χρήστες με βάση την ποιότητα των δεδομένων τους, προκειμένου να καταδείξουμε το πώς μπορούν να αξιοποιηθούν οι δομές και τα δεδομένα του συστήματος.

2. Προδιαγραφές Υποδομής Επικοινωνίας Συστήματος ΣΥΣΔΑ

Η ενότητα αυτή περιγράφει και προδιαγράφει τους μηχανισμούς επικοινωνίας μεταξύ των αισθητήρων και των μηχανισμών αγοράς/πώλησης δεδομένων στην αρχιτεκτονική του συστήματος Συμμετοχικής Συλλογή Δεδομένων Αισθητήρων (ΣΥΣΔΑ) του έργου Inception. Το μέρος αυτό της αρχιτεκτονικής σημειώνεται σε κόκκινο περίγραμμα στο σχήμα 2.



Σχήμα 2. Μηχανισμοί επικοινωνίας μεταξύ αισθητήρων και μηχανισμών

Στο σύστημα μας κάθε χρήστης μπορεί να έχει δύο διακριτούς ρόλους:

- Να προσφέρει πληροφορίες για την διαθεσιμότητα των δεδομένων.
- Να κάνει επερωτήσεις για δεδομένα.

Ένας χρήστης μπορεί να χρησιμοποιεί και τους δύο ρόλους ώστε να παράγει αλλά και να ζητάει δεδομένα από το σύστημα. Χρησιμοποιούμε δύο διαφορετικούς μηχανισμούς για αυτούς τους δύο ρόλους ώστε να εκτελούνται οι απαραίτητες διαδικασίες αποδοτικά.

Ουσιαστικά η παραγωγή των δεδομένων και η δημοσίευση των αναφορών στην δομή Αγοράς MarketBoard αναφέρεται στην λειτουργία Publish ενώ η ανάκτησή τους όταν υπάρχει ζήτηση από κάποιους χρήστες αναφέρεται στην λειτουργία Subscribe μίας Pub/Sub (Publish/

Subscribe) αρχιτεκτονικής [4][5]. Σε ένα τέτοιο σύστημα μπορεί να παραχθεί ένας μεγάλος όγκος δεδομένων. Για να μπορούμε να διαχειριστούμε αποδοτικά τον μεγάλο όγκο των δεδομένων χρησιμοποιούμε την Map Reduce αρχιτεκτονική [1][2] τόσο για την αποθήκευση όσο και για την ανάκτηση των δεδομένων. Σημειωτέον ότι τόσο μια περιληπτική εισαγωγή σε αρχιτεκτονικές Pub/Sub και Map Reduce έχει ήδη γίνει στο Παραδοτέο 2.1 του έργου.

2.1 Χρήστες

Θεωρούμε ότι το σύστημα αποτελείται από N χρήστες που συμμετέχουν μέσω των συσκευών τους (κινητά, tablets κτλ) στο μηχανισμό παραγωγής δημοπρασίας δεδομένων. Οι συσκευές αυτές θα πρέπει να έχουν εγκατεστημένο το λειτουργικό σύστημα Android, στο οποίο θα υλοποιηθούν οι τεχνικές μας. Ο αριθμός των χρηστών παίζει σημαντικό ρόλο στην αποδοτική λειτουργία του συστήματος γιατί είναι άρρηκτα συνδεδεμένος με τον αριθμό των αισθητήρων που προσφέρονται στο MarketBoard.

Μια τέτοια δομή αγοράς ωστόσο, μπορεί (και είναι επιθυμητό) να προκαλέσει να συμμετέχουν ένα ευρύ φάσμα χρηστών που μπορεί να παράγουν μεγάλες ποσότητες από παροδικές ροές δεδομένων. Σε ένα τέτοιο σύστημα όμως δεν είναι όλα τα δεδομένα εξ' ίσου σημαντικά. Η πραγματική αξία των δεδομένων του χρήστη εξαρτάται από διάφορους παράγοντες, συμπεριλαμβανομένων των εξής: τα χαρακτηριστικά των δεδομένων τηλεπισκόπησης, το περιβάλλον του χρήστη (π.χ., γεωγραφική θέση), τους διαθέσιμους πόρους (π.χ., επικοινωνία), την πυκνότητα της περιοχής όπου εξάγονται δεδομένα, το χρόνο που μεσολαβεί από τις προηγούμενες μετρήσεις, κλπ.

Μια σημαντική ερώτηση που αφορά το σύστημα αυτό είναι πώς να επιλέξουμε ένα κατάλληλο αντιπροσωπευτικό δείγμα από τα δεδομένα, το οποίο να διατηρεί τα επιθυμητά χαρακτηριστικά των δεδομένων, αλλά με παράλληλη εξοικονόμηση του κόστους επεξεργασίας και επικοινωνίας κατά τη λήψη της δεδομένων και να ικανοποιήσει οποιοσδήποτε περιορισμούς εφαρμογής.

Το πρόβλημα αυτό είναι δύσκολο εξ' αιτίας των ακόλουθων πτυχών που θα πρέπει να ληφθούν υπ' όψιν:

- Αποτελεσματικότητα: χρειαζόμαστε μια μέθοδο για να εκπροσωπή αποτελεσματικά την ροή δεδομένων που διατηρεί τις ιδιότητες των χωροχρονικών πληροφοριών με ελάχιστα λάθη.
- Επεξεργασία σε πραγματικό χρόνο: η τεχνική πρέπει να είναι σε θέση να επεξεργάζεται το κινητό ρεύμα δεδομένων στο πλαίσιο πραγματικού χρόνου και των περιορισμένων πόρων.
- Πολυπλοκότητα: η τεχνική πρέπει να είναι σε θέση να αντιμετωπίσει τα ρεύματα πολύ μεγάλου όγκου δεδομένων που δημιουργούνται από τα κινητά τηλέφωνα.

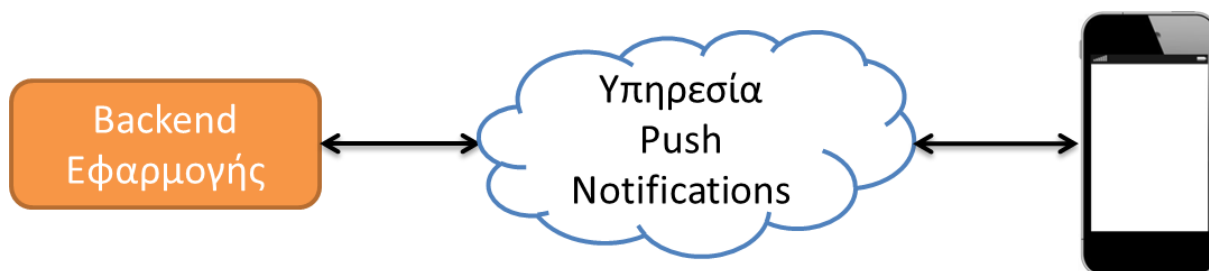
Επιπλέον, οι χρήστες επιθυμούν να λάβουν κάποια αποζημίωση για τη συμμετοχή τους σε ένα τέτοιο σύστημα, και γι' αυτό το λόγο οι χρήστες πρέπει να ανταμειφθούν για τα δεδομένα που παρέχουν. Το ζήτημα αυτό καλύπτεται λεπτομερώς στο Παραδοτέο 3.1 «Παραδοτέο 3.1 Προδιαγραφή υπηρεσίας αγοράς και μηχανισμού κινήτρων».

2.2 Δεδομένα

Τα δεδομένα παράγονται από τους αισθητήρες των συσκευών των χρηστών του συστήματος και αποθηκεύονται τοπικά σε κάθε συσκευή χωρίς να δημοσιεύονται στο MarketBoard. Έτσι το MarketBoard γνωρίζει μόνο τις αναφορές των δεδομένων που μπορεί κάποιος να δημοπρατήσει αλλά όχι τα ίδια τα δεδομένα. Εφόσον οι χρήστες δεν είναι συνδεδεμένοι συνεχώς στο σύστημα, είτε λόγω της διακοπτόμενης συνδεσιμότητας των φορητών συσκευών είτε επειδή μπορεί να μην επιθυμούν να συμμετέχουν στο σύστημα για κάποιο χρονικό διάστημα, αντίστοιχα ούτε τα δεδομένα τους θα προσφέρονται συνεχώς στο MarketBoard.

2.2.1 Παραγωγή Δεδομένων

Η παραγωγή των δεδομένων γίνεται από χρήστες που είναι εγγεγραμμένοι στην υπηρεσία. Παράλληλα οι χρήστες εγγράφονται και σε μία υπηρεσία Push Notifications ώστε να μπορούν να εντοπιστούν και να λάβουν αιτήσεις (για παράδειγμα, για ανάκτηση των δεδομένων τους) χωρίς να χρειάζεται το σύστημα να τους «παρακολουθεί» συνεχώς διατηρώντας ανοιχτές συνδέσεις. Αυτές οι υπηρεσίες έχουν αναπτυχθεί σημαντικά τα τελευταία χρόνια και υποστηρίζονται από όλα τα γνωστά λειτουργικά συστήματα κινητών συσκευών (πχ. Android, iOS, Windows Phone). Οι υπηρεσίες αυτές παρέχουν τη δυνατότητα στο σύστημα να στέλνει μηνύματα σε κάθε εγγεγραμμένο χρήστη, γνωρίζοντας μόνο το id του (δηλ. τον συνδυασμό χαρακτήρων με τον οποίο ταυτοποιείται ο χρήστης στο επίπεδο εφαρμογής – βλέπε κατωτέρω), χωρίς να χρειάζεται να γνωρίζει την διεύθυνση IP του, τον τρόπο σύνδεσης του (WiFi ή 3g) και χωρίς να είναι το σύστημα συνεχώς συνδεδεμένο με το χρήστη μέσω κάποιου socket, όπως έχει ήδη εξηγηθεί. Αυτός ο τρόπος επικοινωνίας είναι πολύ σημαντικός γιατί επιτρέπει την δημιουργία κλιμακώσιμων συστημάτων. Οπότε, όταν το σύστημα χρειάζεται να επικοινωνήσει με τον χρήστη αρκεί να στείλει το μήνυμα στην υπηρεσία Push Notifications μαζί με το id του χρήστη και η υπηρεσία είναι υπεύθυνη να του το προωθήσει. Η επικοινωνία αυτή που διενεργείται από το MarketBoard για να ειδοποιήσει τους χρήστες παρουσιάζεται στο Σχήμα 3.



Σχήμα 3. Push Notifications

Οι χρήστες που συμμετέχουν προσφέροντας δεδομένα στο σύστημα θα πρέπει να έχουν εγκαταστήσει στην εφαρμογή τους το αντίστοιχο λογισμικό το οποίο θα τους δίνει τη δυνατότητα αφότου συνδεθούν να διαλέξουν το σύνολο των αισθητήρων τους οποίους θα χρησιμοποιήσουν για την παραγωγή δεδομένων, καθώς και το χρονικό διάστημα στο οποίο θα γίνεται μία νέα δειγματοληψία. Η επιλογή αυτή θα μπορεί να αλλάξει αργότερα εφόσον ο χρήστης το θελήσει. Επίσης οι χρήστες θα μπορούν να παράγουν και δεδομένα, όποτε το

θελήσουν, για αισθητήρες στους οποίους χρειάζεται ανθρώπινη παρέμβαση (πχ. κάμερα).

Κάθε συσκευή (κινητό τηλέφωνο, tablet, κτλ) παράγει πλειάδες δεδομένων για το σύνολο των αισθητήρων που έχουν επιλεγεί από το χρήστη στο δεδομένο περιοδικό διάστημα. Κάθε πλειάδα δεδομένων που παράγεται από τον χρήστη i για μία χρονική στιγμή t αποθηκεύεται τοπικά στη συσκευή του χρήστη που τις παράγει.

Η μορφή των δεδομένων εξαρτάται από τον τύπο του κάθε αισθητήρα ξεχωριστά και είναι: $\langle ID, id_χρήστη, latitude, longitude, χρονοσφραγίδα, τύπος\ αισθητήρα, δεδομένα \rangle$ όπου:

- Το ID αναφέρεται στο μοναδικό κωδικό της πλειάδας που δημιουργείται από τον κατακερματισμό (hashing) της θέσης και της χρονοσφραγίδας. Θεωρούμε ότι η ακρίβεια της θέσης είναι τέτοια ώστε να μην μπορούν να υπάρχουν 2 χρήστες στο ίδιο σημείο ταυτόχρονα.
- Το $id_χρήστη$ είναι το id του χρήστη στο σύστημα το οποίο είναι αλφαριθμητικής μορφής.
- Τα $latitude, longitude$ προσδιορίζουν τη γεωγραφική θέση της πλειάδας ως γεωγραφικό μήκος και γεωγραφικό πλάτος.
- Η χρονοσφραγίδα αναφέρεται στο χρόνο παραγωγής των δεδομένων.
- Ο τύπος αισθητήρα προσδιορίζεται με το αντίστοιχο αναγνωριστικό.

Τα δεδομένα αναφέρονται στα δεδομένα που παρήγαγε η συσκευή και έχουν τη μορφή που έχει οριστεί για το συγκεκριμένο τύπο αισθητήρα.

Για παράδειγμα μία πλειάδα θα μπορούσε να περιέχει τις ακόλουθες τιμές:

```
<322, aueb_123, 37.9794500, 23.7162200, 09/16/2013@18:0:0EST,
ΑΙΣΘΗΤΗΡΑΣ_ΥΓΡΑΣΙΑΣ, 64%>
```

Το μέγεθος των πλειάδων μπορεί να διαφέρει ανάλογα με τον τύπο του αισθητήρα.

2.2.2 Προσθήκη δεδομένων στο MarketBoard

Για να αποθηκευτούν τα δεδομένα από τους χρήστες στο σύστημα μας, χρησιμοποιούμε το MISCO [3][8][9] το οποίο είναι ένα πλαίσιο αρχιτεκτονικής (framework) για εφαρμογή του MapReduce [1][6][7] εξειδικευμένο σε κινητές συσκευές, το οποίο και μας επιτρέπει την παράλληλη επεξεργασία μεγάλου όγκου δεδομένων.

Η συνάρτηση Map εκτελείται σε κάθε χρήστη και χρησιμοποιείται:

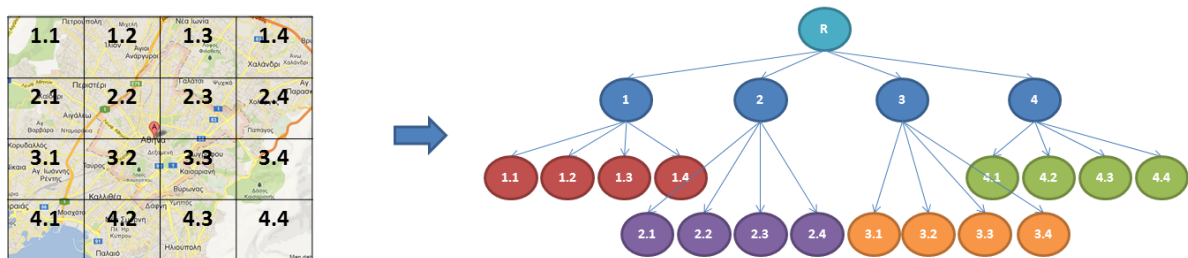
- για την παραγωγή των μονάδων δεδομένων (Application Data Units - ADU) όπως περιγράψαμε παραπάνω (χρησιμοποιώντας τους ενσωματωμένους αισθητήρες της συσκευής)
- για την αποστολή τους στον κεντρικό κόμβο.

Μόλις ολοκληρωθεί η χρονική περίοδος που έχει οριστεί για την παραγωγή δεδομένων η συσκευή χρησιμοποιεί το API του λειτουργικού για να αποκτήσει πρόσβαση στους αισθητήρες που έχουν επιλεγεί από το χρήστη και να κάνει δειγματοληψία. Αυτό έχει ως αποτέλεσμα την παραγωγή μίας νέας πλειάδας για κάθε αισθητήρα. Η πλειάδα αποθηκεύεται τοπικά στη συσκευή είτε σαν σειριακό αρχείο είτε σε κάποια ενσωματωμένη βάση δεδομένων του λειτουργικού συστήματος της συσκευής (πχ. SQLite), ώστε να μπορεί να ανακτηθεί όταν

ζητηθεί από το MarketBoard.

Μόλις παραχθούν τα δεδομένα, δημιουργείται μία δεύτερη πλειάδα με το ID της πλειάδας και του χρήστη, τη γεωγραφική περιοχή των δεδομένων, τη χρονοσφραγίδα και τον τύπο του αισθητήρα ως εξής: $\langle ID, id_χρήστη, latitude, longitude, χρονοσφραγίδα, τύπος\ αισθητήρα \rangle$. Αυτή η πλειάδα προσδιορίζει την αναφορά των δεδομένων και δεν περιέχει τα δεδομένα του αισθητήρα. Η πλειάδα αυτή μεταφέρεται σε ένα κεντρικό κόμβο σαν έξοδος της Map συνάρτησης και σαν είσοδος της Reduce συνάρτησης (που παράγει ενδιάμεσα αποτελέσματα). Η επικοινωνία μεταξύ της Map και της Reduce συνάρτησης βασίζεται σε TCP Sockets με βάση το πρωτόκολλο που έχει οριστεί στο MISCO [3]. Ο σκοπός της χρήσης TCP sockets υποστηρίζεται από την ανάγκη να μην χαθούν πακέτα μηνυμάτων, ειδικά σε ένα ασταθές περιβάλλον όπως είναι αυτό των κινητών συσκευών.

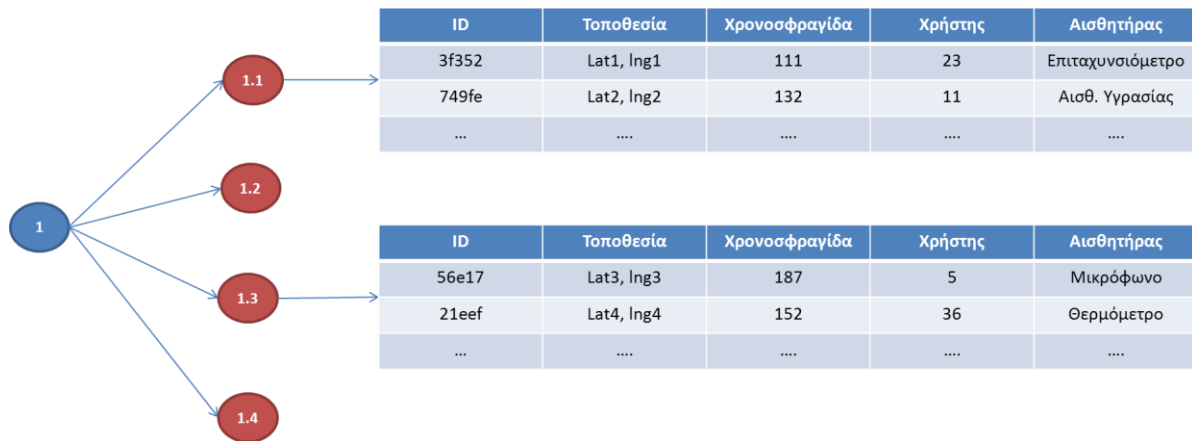
Οι Reduce συναρτήσεις παίρνουν ως είσοδο τις αναφορές πλειάδων των χρηστών (ενδιάμεσα αποτελέσματα) που παρήχθησαν για μια προκαθορισμένη τη χρονική περίοδο. Ο σκοπός της κάθε Reduce συνάρτησης είναι να πάρει τη λίστα των αναφορών που αναφέρονται σε μία συγκεκριμένη γεωγραφική περιοχή και αποθηκεύει τις αναφορές αυτές σε μια δενδρική δομή. Η δενδρική δομή αυτή χρησιμοποιείται για να διαιρέσει κάθε περιοχή σε ένα σύνολο υποπεριοχών σε κάθε επίπεδο του δένδρου (Σχήμα 4). Έτσι στα φύλλα του δένδρου υπάρχουν όλες οι υποπεριοχές και για κάθε μία από αυτές χρησιμοποιείται μία ξεχωριστή reduce συνάρτηση για να αποθηκεύσει τις αναφορές.



Σχήμα 4. Δενδρική Δομή

Για την αποθήκευση των αναφορών σε κάθε φύλλο χρησιμοποιούμε μια δομή κατακερματισμού όπου αποθηκεύεται το σύνολο των αναφορών που ανήκουν σε αυτή τη περιοχή (Σχήμα 5). Οι αναφορές που αποθηκεύονται μέσω της Reduce και η αποθήκευση γίνεται σε μία βάση δεδομένων βασισμένη σε SQL, παρόλο που θα μπορούσε να υλοποιηθεί και σε άλλο τύπο βάσης δεδομένων.

Θεωρούμε ότι μας ενδιαφέρουν μόνο τα τρέχοντα δεδομένα και όχι τα ιστορικά. Συνεπώς, ο κατακερματισμός λαμβάνει υπ' όψιν το id του χρήστη και τον τύπο του αισθητήρα. Έτσι κάθε φορά που ένας χρήστης παράγει νέα δεδομένα για τον ίδιο αισθητήρα, και στην ίδια περιοχή, αυτά θα έχουν το ίδιο αποτέλεσμα στη συνάρτηση κατακερματισμού και έτσι θα αποθηκεύεται μόνο η πιο πρόσφατη αναφορά. Η διαδικασία αυτή μειώνει το κόστος αποθήκευσης και την ταχύτητα ανάκτησης, ενώ παράλληλα αυξάνεται η κλιμακωσιμότητα του συστήματος. Φυσικά, με κατάλληλες επεκτάσεις του συστήματος, θα μπορούσε να γίνει αποθήκευση και των ιστορικών δεδομένων.



Σχήμα 5. Δομή Κατακερματισμού

Τέλος, τα δεδομένα παραμένουν τοπικά στη συσκευή που τα παράγει και μπορούν να ανακτηθούν από αυτή όταν ζητηθούν από κάποιο χρήστη στο MarketBoard, όπως αναφέρεται στην επόμενη παράγραφο.

2.2.3 Δημοσίευση Αναφορών

Η δημοσίευση των αισθητήρων που είναι διαθέσιμοι προς δημοπρασία γίνεται από κάποιο σύστημα δημόσιας πρόσβασης, όπως μια ιστοσελίδα, και είναι ανεξάρτητο από την εφαρμογή που τρέχουν οι χρήστες που παράγουν τα δεδομένα.

Στην ιστοσελίδα μπορεί κάποιος να διαλέξει ένα σύνολο από γεωγραφικές υποπεριοχές που τον ενδιαφέρουν και να του επιστραφούν οι αισθητήρες για τους οποίους υπάρχουν δεδομένα προς πώληση. Αυτή η διαδικασία γίνεται με την ανάκτηση των αποθηκευμένων αναφορών για τη συγκεκριμένη περιοχή, οι οποίες βρίσκονται μέσα σε ένα συγκεκριμένο χρονικό παράθυρο από τη παρούσα χρονική στιγμή. Το χρονικό παράθυρο χρησιμοποιείται για να μην δημοπρατηθούν δεδομένα που έχουν «λήξει», όταν ένας χρήστης έχει καθυστερήσει πολύ να προσφέρει νέα δεδομένα στο σύστημα. Σε αυτό το σημείο ο χρήστης πλέον μπορεί να επιλέξει για αγορά τα δεδομένα των αισθητήρων που τον ενδιαφέρουν με βάση τις τιμές αγοράς/ανάκτησης τους που προκύπτουν από τους μηχανισμούς δημοπρασίας, όπως περιγράφονται στο Παραδοτέο 3.1 «Προδιαγραφή υπηρεσίας αγοράς και μηχανισμού κινήτρων».

2.2.4 Ανάκτηση Δεδομένων

Κάθε χρήστης μπορεί να κάνει επερωτήσεις μέσω της ιστοσελίδας στο σύστημα για να πάρει τα δεδομένα που έχουν παραχθεί από τους χρήστες. Έτσι, κάθε επερωτήση θα πρέπει να περιέχει ένα προσδιορισμό του συνόλου των πλειάδων που θα επιστραφούν. Για παράδειγμα, οι υπερρωτήσεις μπορεί να έχουν τη μορφή: «Δώσε μου δεδομένα από 4 αισθητήρες υγρασίας που αναφέρονται σε μια συγκεκριμένη γεωγραφική περιοχή» ή (όπως προαναφέρθηκε) θα μπορεί να επιλέξει συγκεκριμένους αισθητήρες που προσφέρονται στο σύστημα για να ανακτήσει τα δεδομένα που έχουν συλλέξει.

Η ανάκτηση των δεδομένων βασίζεται και αυτή στην αρχιτεκτονική MapReduce και ξεκινάει

μόλις ένας χρήστης επιλέξει τους αισθητήρες μέσω της διεπαφής του συστήματος. Η αναζήτηση των δεδομένων που θα ανακτηθούν γίνεται στους κεντρικούς κόμβους του συστήματος (όχι στα κινητά) με βάση τις αναφορές των δεδομένων.

Για την ανάκτηση χρησιμοποιούμε το σύστημα Misco [8][9], το οποίο παίρνει σαν είσοδο για τη συγκεκριμένη εργασία, την επερώτηση του χρήστη, η οποία προσδιορίζει το σύνολο των πλειάδων για τα οποία θέλει να λάβει δεδομένα.

Μόλις το σύστημα λάβει ένα τέτοιο επερώτημα από κάποιον χρήστη, μέσω της ιστοσελίδας όπου δημοσιεύονται οι αναφορές, ξεκινάει τη Map συνάρτηση η οποία χρησιμοποιεί ένα σύνολο από workers («εργάτες») για να αναγνωρίσουν το σύνολο των δεδομένων που ανήκουν στο εύρος του επερωτήματος. Ένα τέτοιο επερώτημα μπορεί για παράδειγμα να είναι: «Δώσε μου δεδομένα από 2 αισθητήρες θερμοκρασίας που βρίσκονται σε ακτίνα 500 μέτρων από το γεωγραφικό σημείο 37.98371, 23.72930». Έτσι οι workers αναζητούν στα υποσύνολα του δένδρου (φύλλα) που ορίσαμε προηγουμένως τις αναφορές που αφορούν τη γεωγραφική περιοχή που αναζητεί ο χρήστης (πχ υποσύνολα 1.1, 1.2, 2.1 στο Σχήμα 4) για να βρουν τις αναφορές που προσδιορίζονται από το ερώτημα.

Στη συνέχεια προδιαγράφονται λεπτομερώς τα επιμέρους βήματα που πρέπει να επιτελεστούν σε κάθε τέτοια αναζήτηση.

- Κάθε worker εξετάζει ένα ή περισσότερα υποσύνολα του δένδρου (φύλλα), χωρίς όμως να υπάρχουν επικαλύψεις με άλλους εργάτες, με βάση τις περιοχές που έχει διαλέξει ο χρήστης
- Σε αυτά αναζητεί είτε τις αναφορές που ταιριάζουν στο επερώτημα είτε τον συγκεκριμένο αισθητήρα που έχει επιλεγεί από το χρήστη.
- Για κάθε τέτοια αναφορά που ταιριάζει στα κριτήρια του επερωτήματος επικοινωνεί με τη συσκευή του χρήστη που έχει δημοσιεύσει την αναφορά και εφόσον η συσκευή είναι ενεργή, ζητάει τις αντίστοιχες πλειάδες.
- Για να γίνει αυτή η επικοινωνία η συσκευή λαμβάνει μία ειδοποίηση με τη χρήση της υπηρεσίας Push Notifications (γνωρίζοντας το id του χρήστη που δημοσίευσε την αναφορά) η οποία είναι υπεύθυνη να ξεκινήσει την επικοινωνία ανάκτησης των δεδομένων.
- Κάθε χρήστης που λαμβάνει μία τέτοια αίτηση, προωθεί τα δεδομένα του στον κεντρικό κόμβο, ως ενδιάμεσα αποτελέσματα της Map Reduce αρχιτεκτονικής, μέσω του αντίστοιχου TCP Socket.
- Τα δεδομένα των διαφόρων χρηστών έρχονται σαν είσοδος στην Reduce συνάρτηση που είναι υπεύθυνη ώστε να λάβει αυτή το σύνολο των δεδομένων και στη συνέχεια να τα ενώσει. Στην συνάρτηση αυτή επιβάλλονται και οι περιορισμοί του επερωτήματος. Έτσι, για το παράδειγμα που ορίσαμε, εφόσον έχουν επιστραφεί δεδομένα για περισσότερους από 2 αισθητήρες, θα πρέπει να τα φιλτράρουμε ώστε να επιστρέψουμε στο χρήστη μόνο εκείνα για τα οποία έχει αγοράσει μέσω της δημοπρασίας. Στη συνέχεια τα αποτελέσματα επιστρέφουν στον χρήστη που έκανε την επερώτηση.

Τέλος οι παραγωγοί των δεδομένων εξοφλούνται από τους καταναλωτές για την πώληση αυτών.

3. Αλγόριθμος επιλογής δεδομένων αισθητήρων

Σε αυτήν την ενότητα περιγράφουμε έναν αλγόριθμο επιλογής δεδομένων αισθητήρων που παράγονται από τους χρήστες που έχουμε ήδη αναπτύξει [9], έτσι ώστε να γίνει κατανοητό το πώς μπορούν να αξιοποιηθούν οι δομές και τα δεδομένα του συστήματος ΣΥΣΔΑ που περιγράφηκαν στην ενότητα 2 για την βέλτιστη επιλογή δεδομένων. Η λεπτομερής ανάπτυξη των μηχανισμών δημοπρασίας/αγοράς του έργου περιγράφεται στο Παραδοτέο 3.1.

Θεωρούμε τις περιοχές R_1, R_2, \dots, R_k μιας δεδομένης γεωγραφικής περιοχής όπου βρίσκονται M κινητοί κόμβοι. Κάθε κινητός κόμβος i (δηλ. χρήστης με κινητό τηλέφωνο) παράγει μονάδες δεδομένων (ADUs) συναρτήσει του χρόνου t με μια συνάρτηση ποιότητας των πληροφοριών $QoI_i(t)$ και ένα κόστος $Cost_i(t)$. Η συνάρτηση ποιότητας αντανακλά τη σχετική χρησιμότητα (σημασία) των δεδομένων που παρέχεται από το κινητό τηλέφωνο i στο σύστημα. Το σύστημα επιλέγει πολλαπλές ροές δεδομένων βάσει ενός προϋπολογισμού (Budget) έτσι ώστε να μεγιστοποιηθεί η ποιότητα της πληροφορίας. Η επιλογή των δεδομένων θα πρέπει επίσης να εξετάσει τους περιορισμούς που επιβάλλονται από τους πόρους του συστήματος, ώστε το σύστημα να μπορεί να ανταποκριθεί στο ρυθμό επεξεργασίας $Rate_q$ της εφαρμογής q σε ένα χρονικό διάστημα $Deadline_q$, όπως αυτό καθορίζεται από το μηχανισμό κατανομής των ροών δεδομένων.

Έτσι, ο στόχος μας είναι να καθορίσουμε αυτά τα ADUs, τα οποία οδηγούν σε μεγιστοποίηση του οφέλους των πληροφοριών, αλλά ικανοποιώντας και τους προαναφερθέντες περιορισμούς. Το όφελος των πληροφοριών $Information_i(t)$ που μας δίνει κάθε κινητός κόμβος μπορεί να είναι διαφορετικό από την ποιότητα των πληροφοριών $QoI_i(t)$ και μπορεί να εκφραστεί με διάφορες μετρικές [9]. Το υπό εξέταση πρόβλημα βελτιστοποίησης μπορεί να εκφραστεί ως εξής:

$$\begin{aligned}
 & \text{Maximize } \sum_{i \in M} Information_i(t) \\
 & \text{subject to } \sum_{i \in M} cost_i(t) \leq Budget \\
 & \sum_{i \in M} ADU_i(t) / Deadline_q \leq Rate_q
 \end{aligned}$$

Ο σκοπός μας είναι να λύσουμε αυτό το πρόβλημα, στη διάρκεια της εκτέλεσης του συστήματος, προκειμένου να μεγιστοποιηθεί η πληροφορία που διατίθενται στους χρήστες επιλέγοντας τις πιο σημαντικές ροές δεδομένων. Ωστόσο, η ακριβής επίλυση ενός τέτοιου προβλήματος μεγιστοποίησης είναι υπολογιστικά ιδιαίτερα περίπλοκη (βλέπε κατωτέρω), και συνεπώς προσφέρει την ευκαιρία για εφαρμογή διαφορετικών ευριστικών λύσεων. Ως εκ τούτου, θα πρέπει να μελετήσουμε διαφορετικές ευριστικές στρατηγικές για την επίλυση του προβλήματος μεγιστοποίησης και να συγκρίνουμε τις επιδόσεις του.

Η βασική ιδέα της προσέγγισής μας είναι να καταμηθεί η τροχιά κάθε κινητής συσκευής σε τμήματα με γραμμικά μέρη, με βάση τα χωρικά χαρακτηριστικά των δεδομένων αισθητήρων. Αυτή η τμηματοποίηση είναι στατική και μπορεί να προσδιοριστεί με βάση τις υπο-περιοχές που ορίζονται στη γεωγραφική περιοχή. Η βασική ιδέα είναι ότι όταν υπάρχουν πολλαπλά τμήματα εντός μιας συγκεκριμένης γεωγραφικής περιοχής, το σύστημα μπορεί να εκχωρήσει διαφορετικό κόστος ανά τομέα με βάση τη διαθεσιμότητα των τμημάτων. Με αυτόν

τον τρόπο μπορούμε να εκφράσουμε το πρόβλημα ως ένα πρόβλημα βελτιστοποίησης που ορίζεται από την ποιότητα των δεδομένων, το κόστος των πόρων και την πολυπλοκότητα του χρόνου αναζήτησης. Μια άλλη εύλογη αναπαράσταση θα ήταν η τμηματοποίηση με βάση τη χρονική διάσταση. Ωστόσο, μια τέτοια τμηματοποίηση θα μπορούσε να οδηγήσει σε μεγαλύτερη πολυπλοκότητα, αφού κάθε τροχιά θα μπορεί να συμμετέχει σε διαφορετικό αριθμό χρονικών περιόδων, ποικίλης διάρκειας. Έτσι, θα πρέπει να συγκριθούν τα διάφορα τμήματα, προκειμένου να προσδιοριστούν οι αντίστοιχες γεωγραφικές περιοχές.

Έχουμε υλοποιήσει μια τεχνική δειγματοληψίας που είναι υπεύθυνη για την απόκτηση των προσφορών που αντιστοιχούν στα δεδομένα κάθε κινητού κόμβου και αποφασίζει ποιές θα είναι οι ροές δεδομένων που θα προσφερθούν στους χρήστες. Στην τεχνική αυτή παίρνουμε υπ' όψιν μας τους περιορισμούς πόρων οι οποίες καθορίζουν ένα ανώτατο ποσό ροών δεδομένων όπου το σύστημα μπορεί να ανταποκριθεί επαρκώς, ανάλογα με τη διαθεσιμότητα των πόρων.

Ας υποθέσουμε ότι υπάρχουν m_i κινητοί κόμβοι στην περιοχή R_i . Ο στόχος της τεχνικής είναι να επιλέξουμε ένα υποσύνολο των ροών δεδομένων s_i , (όπου $s_i < m_i$) για να υποβληθούν σε επεξεργασία.

Η βασική ιδέα της τεχνικής που αναπτύξαμε είναι να επιλέξει το σύστημα τις βέλτιστες προσφορές, που παρέχουν την υψηλότερη χρησιμότητα των πληροφοριών, όταν κάθε προσφορά έχει ένα $QoI_i(t)$ και ένα αντίστοιχο $cost_i(t)$, χωρίς να γίνεται υπέρβαση του προϋπολογισμού (Budget) ή των πόρων του συστήματος. Έχει αποδειχθεί ότι ο καθορισμός των βέλτιστων προσφορών, όπου κάθε προσφορά έχει ένα όφελος και ένα αντίστοιχο κόστος και δεν μπορεί να υπερβεί το συνολικό προϋπολογισμό είναι ένα NP-hard πρόβλημα, δεδομένου ότι μπορεί να αναχθεί στο γνωστό πρόβλημα Knapsack.

Στην προσέγγισή μας παρέχουμε επιπλέον περιορισμούς, για να εξασφαλιστεί ότι το σύνολο των προσφορών δεν θα υπερβαίνει τους διαθέσιμους πόρους που μπορεί να επεξεργαστεί το σύστημα σε πραγματικό χρόνο. Έτσι, εισάγουμε τους περιορισμούς αυτούς χωρίς να επηρεάζει η πολυπλοκότητα ή τα χαρακτηριστικά της συμπεριφοράς των μηχανισμών αυτών.

Στην προσέγγισή μας, όλοι οι κόμβοι που επιθυμούν να συμμετάσχουν στη δειγματοληψία, πρέπει να υποβάλουν τις προσφορές τους $\langle QoI_i(t), Cost_i(t) \rangle$ στο σύστημα. Η τεχνική που ακολουθείται για την κατανομή ρυθμού επεξεργασίας στο σύστημα ενημερώνει την τεχνική δειγματοληψίας για το σύνολο των ADUs που το σύστημα μπορεί να επεξεργαστεί στο επόμενο χρονικό παράθυρο ($Rate_q$), για την κάλυψη του καθορισμένου χρονικού περιορισμού ($Deadline_q$). Έτσι, η τεχνική δειγματοληψίας θα είναι σε θέση να επιλέξει τους κόμβους με τις υψηλότερες προσφορές, με βάση την στρατηγική δειγματοληψίας, και θα τους ενημερώσει να μεταδώσουν τα αντίστοιχα δεδομένα (ADUs). Στο τέλος της διαδικασίας δειγματοληψίας, η μικρότερη τιμή ποιότητας QoI από το σύνολο των επιλεγμένων προσφορών (την οποία καλούμε $Threshold$), ταυτοποιείται και επιστρέφεται στους κόμβους που δεν έχουν επιλεγεί για τη δειγματοληψία, έτσι ώστε στον επόμενο γύρο δειγματοληψίας, να μπορούν να επιλέξουν να μην υποβάλουν τις προσφορές τους, αν η αξία του $QoI_i(t)$ είναι μικρότερη από το $Threshold$. Η χρήση αυτού του απλού ορίου, οδηγεί σε σημαντική βελτίωση της απόδοσης.

Στρατηγική ISAM (Information-based SAMpling): Στην πρώτη στρατηγική που αναπτύξαμε, οι κινητοί κόμβοι επιλέγονται από την τεχνική δειγματοληψίας, λαμβάνοντας την ίδια,

προκαθορισμένη τιμή για την πληρωμή τους. Αυτός ο μηχανισμός είναι εμπνευσμένος από δημοπρασίες τύπου Sealed-bid. Στην Δημοπρασία First-Price Sealed-bid οι συμμετέχοντες υποβάλουν την προσφορά τους σε σφραγισμένο φάκελο και η υψηλότερη προσφορά κερδίζει τη δημοπρασία. Στην προκειμένη περίπτωση έχουμε δημοπρασία Sealed-bid που εφαρμόζεται για πολλαπλά «αγαθά» και με ενιαία τιμή. Δηλαδή, για τη συγκεκριμένη στρατηγική θεωρούμε ότι το $cost_i(t)$ είναι μια σταθερή τιμή για κάθε προσφορά, που προσδιορίζεται από το σύστημα και συμβολίζεται ως K . Δεδομένου ότι το σύστημα πληρώνει τους επιλεγμένους κόμβους με το ίδιο ποσό χρημάτων, ανεξάρτητα από το QoI τους, η στρατηγική ISAM πρόκειται να επιλέξει αυτούς τους κινητούς κόμβους που παρέχουν την υψηλότερη $\sum_{i \in M} QoI(t)$. Έτσι, το πρόβλημα μετατρέπεται στο:

$$\begin{aligned} & \text{Maximize } \sum_{i \in M} QoI_i(t) \\ & \text{subject to } \sum_{i \in M} cost_i(t) \leq Budget \\ & \sum_{i \in M} ADU_i(t) / Deadline_q \leq Rate_q \end{aligned}$$

Ο μηχανισμός δουλεύει ως εξής:

- Ταξινομούμε όλες τις προσφορές στο σύνολο A ώστε $QoI_1(t) \geq QoI_2(t) \dots \geq QoI_{|A|}(t)$
- Θέτουμε ως νικητήριο σύνολο το $S = \emptyset$ και τη μεταβλητή $k = 1$
- Θέτουμε $Cost_i(t) = K, \forall i \in A$
- While ($k \leq |A|$ && $|S|+1/Deadline_q \leq Rate_q$ && $cost_k(t) \leq Budget * (QoI_k(t) / \sum_{i \in S \cup \{k\}} QoI(t))$) {
 $S \leftarrow S \cup \{k\}; k \leftarrow k + 1; }$

Τελικά ο μηχανισμός επιστρέφει το σύνολο S ώστε να γίνει η δειγματοληψία των αντίστοιχων κινητών συσκευών. Η αξιολόγηση του μηχανισμού παρουσιάζεται στο [9]. Τα αποτελέσματα αυτά δείχνουν ότι η τεχνική αυτή είναι πρακτική, αποδοτική και έχει καλή επίδοση.

4. Σύνοψη – Συμπεράσματα – Μελλοντική Εργασία

Το παρόν παραδοτέο παρουσιάζει τις προδιαγραφές των μηχανισμών που θα χρησιμοποιηθούν στο έργο INCEPTION για την επικοινωνία μεταξύ των αισθητήρων, των έξυπνων κινητών συσκευών που διαθέτουν οι χρήστες και την αλληλεπίδραση τους με τους μηχανισμούς αγοράς/πώλησης των δεδομένων των αισθητήρων αυτών.

Στα πλαίσια αυτά, αναλύθηκαν λεπτομερώς η αρχιτεκτονική και οι μηχανισμοί που θα χρησιμοποιηθούν για την επικοινωνία αλλά και οι μηχανισμοί για την αποθήκευση και ανάκτηση των δεδομένων, και παρουσιάστηκαν οι προδιαγραφές τους. Στο σύστημα μας κάθε χρήστης μπορεί να έχει δύο διακριτούς ρόλους (ή και τους δύο ταυτόχρονα): Να προσφέρει δεδομένα ή να κάνει επερωτήσεις για δεδομένα. Το σύστημα στηρίζεται στην αρχιτεκτονική Map Reduce τόσο για την αποθήκευση όσο και για την ανάκτηση των δεδομένων, και γενικά στην προσέγγιση αρχιτεκτονικής Publish/Subscribe.

Στη συνέχεια παρουσιάστηκε ο τρόπος επικοινωνίας των μηχανισμών αυτών με τα υπόλοιπα μέρη της πλατφόρμας του Inception, καθώς και η σύνδεση με τους μηχανισμούς αγοράς των δεδομένων. Επίσης αναπτύχθηκε ένας πρώτος αλγόριθμος επιλογής δεδομένων με βάση τα ακόλουθα κριτήρια: αποτελεσματικότητα, επεξεργασία σε πραγματικό χρόνο και πολυπλοκότητα, που στοχεύει να προσδιορίσει το κατάλληλο υποσύνολο των δεδομένων ώστε να ικανοποιήσει τις απαιτήσεις των χρηστών για δεδομένα αισθητήρων καθώς και τα αντίστοιχα κόστη. Τα πρώτα αποτελέσματα αξιολόγησης δείχνουν ότι η τεχνική αυτή είναι πρακτική, αποδοτική και έχει καλή επίδοση. Η αξιολόγηση των διαφόρων αλγορίθμων επιλογής δεδομένων θα γίνει στο Παραδοτέο 3.2 «Υλοποίηση και αξιολόγηση υπηρεσίας αγοράς και μηχανισμού κινήτρων», ενώ η υλοποίηση του συστήματος ΣΥΣΔΑ θα παρουσιαστεί στο Παραδοτέο 4.2 «Ανάπτυξη συστήματος ΣΥΣΔΑ».

5. Αναφορές

- [1] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." *Communications of the ACM* 51.1 (2008): 107-113.
- [2] Marinelli, Eugene E. Hyrax: cloud computing on mobile devices using MapReduce. No. CMU-CS-09-164. CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, 2009.
- [3] A. Dou, V. Kalogeraki, D. Gunopulos, T. Mielikainen and V.H. Tuulos, "Misco: A MapReduce Framework for Mobile Systems", PETRA 2010 Samos, Greece, June 2010.
- [4] Eugster, Patrick Th, et al. "The many faces of publish/subscribe." *ACM Computing Surveys (CSUR)* 35.2 (2003): 114-131.
- [5] N. Fotiou, G. C. Polyzos and D. Trossen. "Illustrating a Publish-Subscribe Internet Architecture", *Telecommunication Systems*, Springer, February 2011, available on-line in <http://www.springerlink.com/content/t6m0k022042088t5/fulltext.pdf>
- [6] Borthakur, Dhruba. "The hadoop distributed file system: Architecture and design." (2007).
- [7] Shvachko, Konstantin, et al. "The hadoop distributed file system." *Mass Storage Systems and Technologies (MSST)*, 2010 IEEE 26th Symposium on. IEEE, 2010.
- [8] Kakantousis, Theofilos, et al. "Misco: A system for data analysis applications on networks of smartphones using mapreduce." *Mobile Data Management (MDM)*, 2012 IEEE 13th International Conference on. IEEE, 2012.
- [9] I. Boutsis, V. Kalogeraki, "Mobile Stream Sampling under Time Constraints", 14th IEEE International Conference on Mobile Data Management (MDM 2013), Milan, Italy, June 2013.