



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

Τμήμα Μηχανικών Η/Υ & Πληροφορικής

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

**Μοντελοποίηση του κοινωνικού πλαισίου για τη
διαχείριση προσωπικής πληροφορίας σε συσκευές
κινητού υπολογισμού**

Αθανάσιος Πλέσσας

Επιβλέπων: Ιωάννης Γαροφαλάκης, Καθηγητής

Πάτρα, Ιανουάριος 2016

Διδακτορική Διατριβή

Αθανάσιος Πλέσσας

A.M. 441

*Μοντελοποίηση του κοινωνικού πλαισίου για τη διαχείριση προσωπικής
πληροφορίας σε συσκευές κινητού υπολογισμού*

Επταμελής Εξεταστική Επιτροπή

Γαροφαλάκης Ιωάννης, Καθηγητής, Τμήμα Μηχανικών Η/Υ & Πληροφορικής,
Πανεπιστήμιο Πατρών (Επιβλέπων Καθηγητής)

Μακρής Χρήστος, Επίκουρος Καθηγητής, Τμήμα Μηχανικών Η/Υ & Πληροφορικής,
Πανεπιστήμιο Πατρών (μέλος Τριμελούς Συμβουλευτικής Επιτροπής)

Χατζηλυγερούδης Ιωάννης, Αναπληρωτής Καθηγητής, Τμήμα Μηχανικών Η/Υ &
Πληροφορικής, Πανεπιστήμιο Πατρών (μέλος Τριμελούς Συμβουλευτικής
Επιτροπής)

Μεγαλοοικονόμου Βασίλειος, Καθηγητής, Τμήμα Μηχανικών Η/Υ &
Πληροφορικής, Πανεπιστήμιο Πατρών (μέλος Εξεταστικής Επιτροπής)

Νικολετσέας Σωτήριος, Αναπληρωτής Καθηγητής, Τμήμα Μηχανικών Η/Υ &
Πληροφορικής, Πανεπιστήμιο Πατρών (μέλος Εξεταστικής Επιτροπής)

Σιούτας Σπυρίδων, Αναπληρωτής Καθηγητής, Τμήμα Πληροφορικής, Ιόνιο
Πανεπιστήμιο (μέλος Εξεταστικής Επιτροπής)

Τσακαλίδης Αθανάσιος, Καθηγητής, Τμήμα Μηχανικών Η/Υ & Πληροφορικής,
Πανεπιστήμιο Πατρών (μέλος Εξεταστικής Επιτροπής)

*Αφιερωμένο στην Ειρήνη και
στους μικρούς Παύλο και Ανδρέα*

O CAPTAIN! my Captain! our fearful trip is done,
The ship has weather'd every rack, the prize we sought is won,
The port is near, the bells I hear, the people all exulting,
While follow eyes the steady keel, the vessel grim and daring;

Ευχαριστίες

Όταν κανείς ξεκινάει μια διδακτορική έρευνα, η στιγμή που αυτή ολοκληρώνεται με τη συγγραφή της διατριβής και τη δημόσια υπεράσπισή της φαίνεται εξαιρετικά μακρινή. Όταν, όμως, φθάσει αυτή η ώρα, διαπιστώνει πως η επιτυχής ολοκλήρωση της δε θα ήταν εφικτή χωρίς την καθοριστική συμβολή κάποιων ανθρώπων. Αυτούς τους ανθρώπους αισθάνομαι πλέον την ανάγκη να ευχαριστήσω δημόσια.

Κατ'αρχάς θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Ιωάννη Γαροφαλάκη για την εμπιστοσύνη που μου έχει δείξει σε όλη τη διάρκεια των ακαδημαϊκών μου σπουδών, καθώς και για τη στήριξη που μου έχει προσφέρει.

Επίσης, θα ήθελα να ευχαριστήσω τους καθηγητές μου στο Τμήμα Μηχανικών Η/Υ & Πληροφορικής κ. Χρήστο Μακρή και κ. Ιωάννη Χατζηλυγερούδη ως μέλη της τριμελούς συμβουλευτικής επιτροπής μου.

Αυτή η έρευνα δε θα είχε ολοκληρωθεί χωρίς τη συνεργασία των συναδέλφων μου από την ερευνητική μας ομάδα. Χάρη στην κοινή μας δουλειά προέκυψαν οι δημοσιεύσεις που επικυρώνουν τα αποτελέσματα αυτής της έρευνας. Πρόκειται για τους Δρ Βασίλη Στεφανή και Δρ Ανδρέα Κομνηνό, με τους οποίους έχω την τύχη να είμαστε και καλοί φίλοι.

Ένα μεγάλο ευχαριστώ στους γονείς μου και τον αδερφό μου που είναι πάντα δίπλα μου και που έβαλαν τις βάσεις για ότι έχω πετύχει στην ενήλικη ζωή μου.

Τέλος, ευχαριστώ τη σύντροφο της ζωής μου Ειρήνη για τη στήριξη και την υπομονή της και τους γιους μου Παύλο και Ανδρέα που αποτελούν πηγή έμπνευσης για μένα.

Περίληψη

Σε μια ιδανική εφαρμογή της διαχείρισης προσωπικής πληροφορίας, οι χρήστες θα έπρεπε να έχουν πάντα διαθέσιμη την κατάλληλη πληροφορία στο σωστό μέρος και στην κατάλληλη μορφή, σε εκείνη την ποιότητα και την πληρότητα που απαιτείται για να ικανοποιήσουν τις ανάγκες τους. Η ανάγκη αυτή γίνεται ακόμα πιο επιτακτική στην περίπτωση των χρηστών συσκευών κινητού υπολογισμού, οι οποίες παρουσιάζουν ιδιαιτερότητες σε σχέση με τους σταθερούς υπολογιστές που μεγεθύνουν τις προκλήσεις και δυσχεραίνουν τη διαχείριση της προσωπικής πληροφορίας.

Παρά τις τεχνολογικές εξελίξεις όσον αφορά την υπολογιστική ισχύ, τη χωρητικότητα και τη συνδεσιμότητα, οι οθόνες των συσκευών κινητού υπολογισμού παραμένουν σημαντικά μικρότερες από αυτές των σταθερών υπολογιστών, περιορίζοντας την ποσότητα πληροφορίας που μπορούν να παρουσιάσουν ταυτόχρονα, ενώ η ανάγνωση μεγάλης ποσότητας πληροφορίας σε αυτές τις συσκευές απαιτεί συχνή οριζόντια κύλιση και εστιασμένη συγκέντρωση. Επιπλέον, ο εξοπλισμός και οι εφαρμογές που ενσωματώνονται σε αυτές τις συσκευές επιτρέπουν την εύκολη δημιουργία πολύ μεγάλης ποσότητας προσωπικής πληροφορίας, όπως βίντεο, φωτογραφίες, σημειώσεις κ.α. Επίσης, το πλαίσιο χρήσης αυτών των συσκευών είναι συχνά μεταβαλλόμενο (π.χ. καθώς οι χρήστες κινούνται αλλάζει η τοποθεσία τους, οι συσκευές και οι άνθρωποι με τους οποίους αλληλεπιδρούν, καθώς και οι στόχοι και οι ανάγκες τους) και πιο απαιτητικό σε σχέση με τη χρήση των σταθερών υπολογιστών. Όλα τα παραπάνω χαρακτηριστικά σε συνδυασμό με την απροθυμία των χρηστών να αφιερώσουν χρόνο και προσπάθεια για τη διαγραφή προσωπικών δεδομένων, έχει ως αποτέλεσμα να δυσχεραίνεται η ανάκτηση και επομένως η αξιοποίηση της πληροφορίας.

Από την άλλη, οι συσκευές κινητού υπολογισμού είναι εξοπλισμένες με ένα πλήθος αισθητήρων που καταγράφουν το κινητό πλαίσιο του χρήστη (mobile context): τοποθεσία, δραστηριότητα (π.χ. περπάτημα, τρέξιμο, σε κλήση), επίπεδα θορύβου, περιβαλλοντικές συνθήκες κ.α. Επιπλέον, καθώς οι χρήστες τις θεωρούν έμπιστες συσκευές και τις έχουν συχνά μαζί τους σε λειτουργία σε 24ωρη βάση, περιέχουν μεγάλες ποσότητες προσωπικής πληροφορίας που σχετίζεται με το κοινωνικό περιβάλλον των χρηστών. Η πληροφορία αυτή συχνά δημιουργείται αυτόματα από τις ίδιες τις συσκευές. Για παράδειγμα, για κάθε κλήση που πραγματοποιείται σε ένα κινητό τηλέφωνο αποθηκεύεται στη λίστα κλήσεων μια σχετική εγγραφή, η οποία

περιλαμβάνει την ημερομηνία και την ώρα, την επαφή, τη διάρκεια. Επίσης, μέρος της πληροφορίας δημιουργείται από τους ίδιους τους χρήστες: SMS και MMS, φωτογραφίες, βίντεο και αρχεία ήχου, γεγονότα ημερολογίου, επαφές κ.α. Συνεπώς, οι συσκευές αυτές εκτός από το κινητό πλαίσιο των χρηστών, είναι ενήμερες και για το κοινωνικό περιβάλλον τους (κοινωνικό πλαίσιο – social context). Ο συνδυασμός του κινητού και του κοινωνικού περιβάλλοντος έχει ως αποτέλεσμα ένα δυναμικά ορισμένο κοινωνικό πλαίσιο, γνωστό ως κινητό κοινωνικό πλαίσιο (mobile social context).

Η προσωπική πληροφορία σχετίζεται σε μεγάλο βαθμό με το πλαίσιο, κάτι που ισχύει και για τις δραστηριότητες διαχείρισής της. Καθώς οι πληροφοριακές ανάγκες των χρηστών εξαρτώνται όχι μόνο από το κινητό πλαίσιο αλλά και από τις κοινωνικές τους αλληλεπιδράσεις, στόχος της παρούσας διδακτορικής έρευνας ήταν η δημιουργία ενός μοντέλου που αξιοποιεί την πληροφορία πλαισίου που συλλέγουν οι συσκευές κινητού υπολογισμού για να βελτιώσει τη διαχείριση της προσωπικής πληροφορίας από τους χρήστες. Το στάδιο εκείνο του κύκλου ζωής της διαχείρισης προσωπικής πληροφορίας στο οποίο δόθηκε έμφαση ήταν η αξιοποίηση. Μελετήθηκε η περίπτωση των επαφών ως αντιπροσωπευτικό πρόβλημα τομέα, επειδή το πλήθος τους αυξάνει συνεχώς με τη πάροδο των χρόνων, οι χρήστες είναι απρόθυμοι να αφιερώσουν χρόνο και προσπάθεια για την οργάνωση ή διαγραφή τους και καθώς οι συσκευές αυτές στην πιο διαδεδομένη μορφή τους (Smartphones) έχουν ως κύρια αποστολή τη διευκόλυνση της επικοινωνίας. Σκοπός ήταν να διερευνηθεί πως το πλαίσιο μπορεί να βελτιώσει την ανάκτηση των επαφών για την πραγματοποίηση επικοινωνίας και συγκεκριμένα εξερχόμενων κλήσεων. Εντούτοις, η προσέγγιση που ακολουθήθηκε και παρουσιάζεται μπορεί να επεκταθεί και να εφαρμοστεί σε οποιοδήποτε άλλο στοιχείο προσωπικής πληροφορίας (π.χ. μηνύματα, emails, σημειώσεις, φωτογραφίες, βίντεο κ.λπ.).

Η διάρθρωση της παρούσας διατριβής είναι ως ακολούθως:

Αρχικά, στο κεφάλαιο 1, γίνεται μια εισαγωγή στο πρόβλημα της διαχείρισης προσωπικής πληροφορίας και αναλύονται οι λόγοι που το υπό μελέτη πρόβλημα παρουσιάζει αυξημένο ενδιαφέρον για την περίπτωση των συσκευών κινητού υπολογισμού. Μετά τη διασύνδεση του προβλήματος με το κινητό κοινωνικό πλαίσιο, παρουσιάζεται το πλαίσιο εκπόνησης της διδακτορικής έρευνας: το κίνητρο, οι στόχοι και τα ερευνητικά ερωτήματα.

Στο κεφάλαιο 2 παρουσιάζεται μια εκτεταμένη επισκόπηση της ερευνητικής δραστηριότητας γενικά για τη διαχείριση προσωπικής πληροφορίας σε συσκευές

κινητού υπολογισμού και την αξιοποίηση του πλαισίου, αλλά και ειδικότερα για την περίπτωση των επαφών.

Στο κεφάλαιο 3 προτείνεται μια μοντελοποίηση του πλαισίου με τη μορφή διανύσματος. Αν και πρόκειται για μια σχετικά απλή αναπαράσταση, επιτρέπει την εύκολη διαχείριση του πλαισίου και δίνει τη δυνατότητα μελλοντικής ενσωμάτωσης νέων διαστάσεων. Επιπλέον, επιτρέπει την εφαρμογή αλγορίθμων μείωσης διαστάσεων (dimensionality reduction). Μια αρχική ιδέα που διερευνήθηκε ήταν η μείωση των διαστάσεων του πλαισίου σε 2 ή 3, πρακτική που θα επέτρεπε την αναπαράσταση των αντικειμένων προσωπικής πληροφορίας στον δισδιάστατο ή τρισδιάστατο χώρο και ενδεχομένως θα βελτιώνει την ανάκτηση. Τελικά, η ιδέα εγκαταλείφθηκε, καθώς το πλήθος των διαστάσεων που χρησιμοποιήθηκαν δεν ήταν τέτοιο ώστε να δικαιολογεί την ανάγκη για εφαρμογή τέτοιων τεχνικών. Για τον εντοπισμό εκείνων των διαστάσεων του πλαισίου που παίζουν σημαντικό ρόλο στην ανάκτηση των επαφών κατά την έναρξη μιας κλήσης, πραγματοποιήθηκε και παρουσιάζεται πείραμα που περιελάμβανε την ανάλυση πραγματικών δεδομένων από συσκευές χρηστών. Η ανάλυση είχε ως αποτέλεσμα να εντοπιστούν δύο διαστάσεις ως σημαντικές (συχνότητα και προσφατότητα επικοινωνίας) και να απορριφθούν άλλες που χρησιμοποιούνται συχνά στη βιβλιογραφία, χωρίς όμως επαρκή αιτιολόγηση. Επιπλέον, αποκάλυψε σημαντικά στοιχεία που αφορούν τα πρότυπα επικοινωνίας των χρηστών, όπως για παράδειγμα μια συμπεριφορά σχετικά με το πλήθος των κλήσεων προς επαφές αρκετά κοντά στην αρχή Pareto.

Στο κεφάλαιο 4 παρουσιάζεται ο αλγόριθμος που αναπτύχθηκε για την κατάταξη και επιλογή επαφών με τη χρήση των διαστάσεων της συχνότητας και προσφατότητας της επικοινωνίας. Ο αλγόριθμος, που λειτουργεί ως αλγόριθμος πρόβλεψης των N επαφών με τη μεγαλύτερη πιθανότητα να καλέσει ο χρήστης, αξιολογήθηκε με τη μέθοδο της προσομοίωσης σε διαθέσιμα σύνολα δεδομένων. Τα αποτελέσματα είναι περισσότερο ή εξίσου επιτυχή σε σχέση με αυτά που αναφέρουν αντίστοιχες ερευνητικές προσπάθειες. Επιπλέον, υλοποιήθηκαν οι σημαντικότεροι σχετικοί αλγόριθμοι που περιγράφονται στη βιβλιογραφία και έγινε απευθείας σύγκριση των αποτελεσμάτων με χρήση των ίδιων δεδομένων, όπου τα αποτελέσματα του προτεινόμενου αλγορίθμου ήταν βελτιωμένα σε σχέση με αυτά των άλλων εργασιών. Ακόμα, εντοπίστηκαν τρεις διαφορετικές κατηγορίες χρηστών σε σχέση με την «κοινωνικότητά» τους. Για τους χρήστες που είναι λίγο ή μέτρια «κοινωνικοί» η απόδοση του αλγορίθμου είναι εξαιρετικά υψηλή, ενώ για τους πολύ «κοινωνικούς» χρήστες υπάρχουν περιθώρια βελτίωσης. Επιπροσθέτως, παρατίθενται τα αποτελέσματα πειράματος πεδίου που πραγματοποιήθηκε για την αξιολόγηση του

αλγορίθμου σε πραγματικές συνθήκες, από το οποίο επιβεβαιώθηκαν τα συμπεράσματα που προέκυψαν από τη διαδικασία της προσομοίωσης.

Στο κεφάλαιο 5 παρουσιάζεται η μελέτη της επίπτωσης της τοποθεσίας στην ανάκτηση επαφών. Η τοποθεσία θεωρείται ως μια πολύ σημαντική διάσταση πλαισίου και προτείνεται πολύ συχνά στη βιβλιογραφία που αφορά την πρόβλεψη κλήσεων. Σε πρώτη φάση αναλύθηκαν τα δεδομένα φυσικής τοποθεσίας (physical location), χωρίς να προκύψει κάποια στατιστικά σημαντική βελτίωση στα αποτελέσματα πρόβλεψης από τη χρήση της τοποθεσίας. Στη συνέχεια, αναλύθηκαν δεδομένα σημασιολογικής τοποθεσίας (semantic location), απ' όπου προέκυψαν παρόμοια συμπεράσματα. Αντίστοιχα συμπεράσματα προκύπτουν και από την ενσωμάτωση στον αλγόριθμο πρόβλεψης διαστάσεων που σχετίζονται έμμεσα με τη σημασιολογική τοποθεσία, όπως είναι το χρονικό πλαίσιο (ώρα της ημέρας). Αν και πρέπει να είμαστε επιφυλακτικοί στη γενίκευση των συμπερασμάτων, φαίνεται ότι η τοποθεσία δεν είναι τόσο χρήσιμος παράγοντας πλαισίου για αρκετούς χρήστες όσον αφορά το συγκεκριμένο πρόβλημα, όσο είναι διαδεδομένο ως αντίληψη στην επιστημονική κοινότητα.

Τέλος, το κείμενο της διατριβής ολοκληρώνεται με τον επίλογο, όπου συνοψίζονται τα συμπεράσματα, απαντώνται τα ερευνητικά ερωτήματα και δίνεται έμφαση στη συνεισφορά της διδακτορικής έρευνας. Παράλληλα, παρουσιάζεται μια γενίκευση του προτεινόμενου μοντέλου που μπορεί να εφαρμοστεί για οποιοδήποτε τύπο προσωπικής πληροφορίας ή ακόμα και για συνδυασμό διαφορετικών τύπων.

Executive Summary

In an ideal application of personal information management, users should always have available the appropriate information to the right place and in the right format, at that quality and completeness that is necessary to meet their needs. This requirement becomes even more urgent in the case of users with mobile computing devices, which face limitations compared to desktop computers that increase the challenges and complicate personal information management.

Despite the technological advancements regarding computing power, storage capacity and connectivity, the screens of mobile computing devices remain significantly smaller than those of desktop computers, reducing the amount of information that can be presented simultaneously, while reading a large amount of information on these devices requires frequent horizontal scrolling and focused concentration. Moreover, the equipment and applications that are integrated in these devices allow the easy creation of big volumes of personal information data, such as videos, photos, notes, etc. Also, the context of use of these devices is often changing (e.g. while users are on the move their location changes, as well as the devices and people with whom they interact, their objectives and their needs) and it is more demanding than in the case of desktop use. The above characteristics, combined with the reluctance of users to devote time and effort to delete personal data result in information retrieval difficulties.

On the other hand, mobile computing devices are equipped with a multitude of sensors that record the context of the user (mobile context): location, activity (e.g. walking, jogging, on a call), noise levels, environmental conditions etc. Furthermore, as users consider them as trusted devices and often leave them always on, they contain large amounts of personal information related to the users' social environment. This information is often generated automatically. For example, for each call made or received on a mobile phone a relevant record is saved in the call list, which includes the date and time, the contact and the duration. Also, part of this information is generated by the users themselves: SMS and MMS, photos, videos and audio files, calendar events, contacts etc. Therefore, these devices except from the users' mobile context are also informed about their social environment (social context). The combination of mobile and social environment results in a dynamically defined context, known as mobile social context.

Personal information is highly related to the context of use, which is also true for its management activities. As the information needs of the users depend not only on mobile context but also on their social interactions, the aim of this PhD research was to create a model that uses the contextual information collected by mobile computing devices in order to improve the management of personal information. The stage of personal information management lifecycle on which this research had focused was exploitation. The case of mobile contacts was studied as a representative problem domain, because their number is constantly increasing, while users are reluctant to devote time and effort to organize or delete them and as the main mission of these devices in their most common form (Smartphones) is to facilitate communication. The goal of the research was to investigate in which way context can improve contact retrieval when starting a communication session, especially for the case of outgoing calls. However, the presented approach can be extended and applied to any other personal information element (e.g. messages, emails, notes, photos, video, etc.).

The structure of this thesis is as follows:

Initially, Chapter 1 introduces the reader to the problem of personal information management and analyzes the reasons why the studied problem shows an increased interest for the case of mobile computing devices. Next to linking the problem with mobile social context, the framework of the PhD research is presented: motivation, objectives and research questions.

Chapter 2 presents an extensive literature review of related research activities for personal information management in mobile computing devices and use of context in general, and particularly for the case of mobile contacts.

Chapter 3 proposes a model of context representation in the form of vector. Although this is a relatively simple representation, it allows for the easy context management and enables future integration of new dimensions. Moreover, it allows for the application of dimensionality reduction algorithms. A concept that was investigated was the reduction of the context dimensions to 2 or 3, a practice which would allow representation of personal information objects in two-dimensional or three-dimensional space and could possibly improve retrieval. Eventually, the idea was abandoned, as the number of adopted dimensions were not such as to justify the need for application of such techniques. In order to identify the context dimensions that are relevant to the problem of contact retrieval when starting a new call session, an experiment that was conducted and included analysis of real data from users' devices is presented. The analysis identified two dimensions as important (frequency and

recency of communication) and rejected others that are often used in literature, nevertheless without sufficient justification. Moreover, it revealed interesting communication patterns of the users, such as a form of Pareto Principle regarding the number of calls to contacts.

Chapter 4 presents the algorithm that was developed for the retrieval, based on the ranking and selection of contacts using frequency and recency of communication. The algorithm, which operates as a prediction algorithm of the N contacts that are most likely for the user to call, was assessed by simulation on the available datasets. The results are more or equally successful than those that are reported in relevant research papers. In addition, some of the available algorithms in literature were implemented and were directly compared to the proposed algorithm, using the same dataset. This comparison resulted in much better performance of the proposed solution. Furthermore, three different types of users were identified according to their perceived “socialness”. For low or medium “social” users the performance of the algorithm is extremely high, while for more “social” users there is still room for improvement. In addition, in this chapter the results of a field experiment that was conducted to evaluate the algorithm under real conditions are presented, confirming the conclusions of the simulation procedure.

Chapter 5 presents a study on the impact of location on contact retrieval. Location is considered as a very important context dimension and is very often suggested as a candidate dimension for call prediction in literature. Initially, physical location was analyzed, without finding a significant improvement in the prediction results when incorporating this dimension. Then, semantic location was also considered in the prediction algorithm, leading to similar conclusions. Similar conclusions resulted from the integration to the algorithm of contextual dimensions that are indirectly related to location, such as temporal context (time of day). Although we must be cautious in generalizing these findings, it seems that location is not such a useful context factor for many users regarding contact retrieval, as perceived in the scientific community.

Finally, this thesis concludes with an epilogue where the conclusions are summarized, the research questions are answered and the contribution of the research is emphasized. At the same time, a generalization of the proposed model that can be applied to any type of personal information items or even to a combination of different types is presented.

Περιεχόμενα

Περίληψη	i
Executive Summary	v
Ευρετήριο Εικόνων	xiii
Ευρετήριο Γραφημάτων	xv
Ευρετήριο Πινάκων	xix
1. Εισαγωγή	1
1.1 Ορισμοί	1
1.2 Προβλήματα διαχείρισης προσωπικής πληροφορίας και συσκευές κινητού υπολογισμού	4
1.2.1 Προβλήματα στη διαχείριση προσωπικής πληροφορίας	4
1.2.2 Συσκευές κινητού υπολογισμού και προκλήσεις	6
1.3 Αξιοποίηση του κινητού κοινωνικού πλαισίου	9
1.4 Πλαίσιο εκπόνησης διατριβής	12
1.4.1 Κίνητρο - στόχος	12
1.4.2 Οι επαφές ως πρόβλημα τομέα	13
1.4.3 Ερευνητικά ερωτήματα και αντιμετώπισή τους	16
2. Επισκόπηση Σχετικής Ερευνητικής Δραστηριότητας	19
2.1 Διαχείριση προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού	19
2.1.1 Ανάγκες των κινητών χρηστών για πληροφορία	19
2.1.2 Εργαλεία διαχείρισης προσωπικής πληροφορίας και συσκευές κινητού υπολογισμού	22
2.2 Κινητό πλαίσιο και μοντελοποίηση	25
2.3 Μελέτες για το πρόβλημα της ανάκτησης των επαφών	29
2.3.1 Μελέτες ευχρηστίας της λίστας επαφών	29
2.3.2 Σύνθεση του αποθετηρίου επαφών και πρότυπα χρήσης επαφών και επικοινωνίας	31
2.3.3 Εναλλακτικά πρωτότυπα για τη λίστα επαφών	41
2.3.4 Άλλες τεχνικές για τη βελτίωση της ανάκτησης των επαφών	44

2.4 Προσαρμοζόμενες λίστες επαφών και πρόβλεψη κλήσεων.....	45
2.5 Η τοποθεσία ως παράγοντας πλαισίου για την πρόβλεψη κλήσεων	59
2.6 Σύνοψη.....	61
3. Μοντελοποίηση Πλαισίου Συσκευών Κινητού Υπολογισμού.....	63
3.1 Αναπαράσταση πλαισίου ως διάνυσμα	63
3.1.1 Μείωση διαστάσεων	66
3.1.2 Απόδοση του αλγορίθμου Fastmap σε συσκευές κινητού υπολογισμού	68
3.2 Πείραμα μελέτης αποθετηρίου επαφών και προτύπων επικοινωνίας.....	71
3.2.1 Σχεδιασμός πειράματος και δείγμα συμμετεχόντων.....	72
3.2.2 Αποτελέσματα.....	74
3.2.3 Συμπεράσματα	89
3.3 Σύνοψη.....	92
4. Αξιοποίηση Συχνότητας και Προσφατότητας ως Παραγόντων Κοινωνικού Πλαισίου Συσκευών Κινητού Υπολογισμού	95
4.1 Αλγόριθμος αξιοποίησης πλαισίου για την πρόβλεψη εξερχόμενων κλήσεων..	95
4.2 Προκαταρκτικό πείραμα αξιολόγησης αλγορίθμου	98
4.2.1 Σύνολο δεδομένων	100
4.2.2 Πειραματικά αποτελέσματα.....	104
4.2.2.1 Πειράματα βάσης σύγκρισης	104
4.2.2.2 Κυρίως πειράματα	105
4.3 Συζήτηση αποτελεσμάτων.....	113
4.4 Μεγάλης κλίμακας πείραμα αξιολόγησης αλγορίθμου.....	115
4.4.1 Σύνολο δεδομένων	115
4.4.2 Πειραματικά αποτελέσματα.....	118
4.4.2.1 Πειράματα βάσης σύγκρισης	118
4.4.2.2 Κυρίως πειράματα	120
4.4.3 Πειραματισμός με διαφορετικά μήκη παραθύρων εκπαίδευσης	123
4.4.4 Απευθείας σύγκριση με άλλους αλγορίθμους	126
4.4.5 Συζήτηση αποτελεσμάτων.....	127

4.5 Πείραμα πεδίου για την αξιολόγηση του αλγορίθμου	129
4.5.1 Η εφαρμογή Calchas	129
4.5.2 Σύνολο δεδομένων	131
4.5.3 Πειραματικά αποτελέσματα.....	132
4.5.3.1 Ποσοτικά δεδομένα	132
4.5.3.2 Ποιοτικά δεδομένα	137
4.5.4 Συζήτηση αποτελεσμάτων.....	138
4.6 Σύνοψη.....	140
5. Οι Επιπτώσεις της Τοποθεσίας στη Διαχείριση Προσωπικής Πληροφορίας Συσκευών Κινητού Υπολογισμού.....	143
5.1 Προκαταρκτική διερεύνηση του χρονικού πλαισίου.....	143
5.2 Φυσική vs σημασιολογική τοποθεσία.....	145
5.3 Μελέτη της επίπτωσης της φυσικής τοποθεσίας	146
5.3.1 Σύνολο δεδομένων	147
5.3.2 Συσταδοποίηση (clustering) δεδομένων τοποθεσίας	150
5.3.3 Τύποι χρηστών	156
5.3.4 Αποτελέσματα πρόβλεψης	159
5.4 Μελέτη της επίπτωσης της σημασιολογικής τοποθεσίας	165
5.4.1 Σύνολο δεδομένων	165
5.4.2 Αποτελέσματα πρόβλεψης	168
5.5. Συμπεράσματα.....	170
5.6 Σύνοψη.....	173
6. Επίλογος	175
Βιβλιογραφικές Αναφορές.....	181

Ευρετήριο Εικόνων

Εικόνα 1. Κατηγορίες πλαισίου στο έργο AmbieSense (Mikalsen & Kofod-Petersen, 2004)	10
Εικόνα 2. Από αριστερά προς τα δεξιά και από πάνω προς τα κάτω: αποθετήριο επαφών, αναζήτηση στο αποθετήριο, λίστα πιο συχνά χρησιμοποιούμενων επαφών, λίστα πρόσφατων κλήσεων, λίστα χαμένων κλήσεων, λίστα γρήγορης επανάκλησης (speed dial)	14
Εικόνα 3. Ένα αφαιρετικό εργαλείο διαχείρισης προσωπικής πληροφορίας (Boardman, 2004)	14
Εικόνα 4. ContextML σχήμα αναπαράστασης πλαισίου (Knappmeyer, Kiani, Frà, Moltchanov & Baker, 2010)	27
Εικόνα 5. Μοντελοποίηση πλαισίου με UML για το σύστημα HEP (Chihani, Bertin, Jeanne & Crespi, 2011).....	27
Εικόνα 6. Η διαδικασία ανάκτησης με ένα τυπικό NOKIA UI (αριστερά) και το πρωτότυπο των Bergman, Komninos, Liarokapis και Clarke (2012) (δεξιά)	33
Εικόνα 7. Κλάσματα κλήσεων με επαφές από τις πέντε ομάδες κατά τη διάρκεια της ημέρας (Le, Henze & Schmidt, 2013).....	37
Εικόνα 8. Το ποσοστό των κλήσεων (αριστερά) και SMS (δεξιά) για τις N πιο συχνά χρησιμοποιούμενες επαφές (Bentley & Chen, 2015)	39
Εικόνα 9. Ποσοστό των N πιο συχνά χρησιμοποιούμενων επαφών που περιλαμβάνονται στις N πιο πρόσφατες επαφές (Bentley & Chen, 2015).....	40
Εικόνα 10. Η λειτουργία των top επαφών από το (Jung, Anttila & Blom, 2008). Από αριστερά προς τα δεξιά: αυτόματα επιλεγμένες top επαφές, επιλεγμένες από τον χρήστη top επαφές, ρυθμίσεις.	42
Εικόνα 11. Έξυπνες ομάδες επαφών που δείχνουν τις επαφές που ταιριάζουν σε συγκεκριμένες συνθήκες (Jung, Anttila & Blom, 2008)	42
Εικόνα 12. Συγκρίσεις των μέσων ποσοστών επιτυχίας από την προσομοίωση των Lee, Seo & Lee (2010)	48
Εικόνα 13. Απόδοση μίξης χαρακτηριστικών ως προς την επιτυχία πρόβλεψης. MFU=συχνά χρησιμοποιούμενες επαφές, MRU=πιο πρόσφατα χρησιμοποιούμενες επαφές, SF=μίξη με βάση το σκορ, RF=μίξη με βάση την κατάταξη	52
Εικόνα 14. Απόδοση του συστήματος πρόβλεψης εξερχόμενων κλήσεων (CPL) με και χωρίς τους για πρώτη φορά κληθέντες (Phithakkitnukoon & Dantu, 2011)	53
Εικόνα 15. Ακρίβεια του συστήματος πρόβλεψης εξερχόμενων κλήσεων (CPL) για διαφορετικά μήκη της λίστας προτάσεων και διαφορετικές κοινωνικές ομάδες (Phithakkitnukoon & Dantu, 2011)	54

Εικόνα 16. Παράδειγμα προτύπων που συμπίπτουν (δεξιά) και υπόμνημα για τις τιμές εμπιστοσύνης (αριστερά) (Srinivasan, Moghaddam, Mukherji, Rachuri, Xu & Tapia, 2014)	56
Εικόνα 17. Απόδοση πρόβλεψης του MobileMiner για τους συνδυασμούς ανάκλησης και ακρίβειας και για 1,3 και 5 θέσεις πρόβλεψης (Srinivasan, Moghaddam, Mukherji, Rachuri, Xu & Tapia, 2014).....	56
Εικόνα 18. Υβριδική, μονοδιάστατη αντιστοίχιση της σημαντικότητας	67
Εικόνα 19. Διεπαφές ανάκτησης που χρησιμοποιούν 2D (αριστερά) και 3D (δεξιά) προβολές της σημαντικότητας των αντικειμένων (επαφές, μηνύματα, emails) σε συνδυασμό με διατηρημένες μη προβαλλόμενες διαστάσεις (χρόνος, απόσταση από την τρέχουσα τοποθεσία)	67
Εικόνα 20. Σχήμα ΒΔ για την αποθήκευση των δεδομένων από τις συσκευές των χρηστών	72
Εικόνα 21. Δυνατότητα χαρακτηρισμού επαφής ως αγαπημένη σε συσκευές Android	73
Εικόνα 22. Screenshots από τις οθόνες εγκατάστασης και εκτέλεσης της εφαρμογής.....	73
Εικόνα 23. Η λίστα με τις αγαπημένες επαφές στο Android, όπου προωθούνται στην κορυφή πριν από τις συχνά χρησιμοποιούμενες για γρήγορη επικοινωνία.....	76
Εικόνα 24. Σχηματική αναπαράσταση της διαδικασίας εκτέλεσης της προσομοίωσης εφαρμογής του αλγορίθμου πρόβλεψης.....	99
Εικόνα 25. Σχήμα ΒΔ για την αποθήκευση των αρχείων κλήσεων των χρηστών	117
Εικόνα 26. Η Android εφαρμογή Calchas.....	130
Εικόνα 27. Σχήμα της ΒΔ για την καταγραφή των δεδομένων.....	131
Εικόνα 28. Παράδειγμα εγγραφών από τα δεδομένα τοποθεσίας GPS	147
Εικόνα 29. Παράδειγμα εγγραφών από τα δεδομένα τοποθεσίας WLAN	147
Εικόνα 30. Παράδειγμα εγγραφών από τα δεδομένα τοποθεσίας GSM (αριστερά) και εγγραφών με λεπτομέρειες που αφορούν τις GSM κεραίες (δεξιά).....	148
Εικόνα 31. Προσαρμογή του πίνακα της ΒΔ για την αποθήκευση των πληροφοριών τοποθεσίας για κάθε κλήση.	150
Εικόνα 32. Συσταδοποίηση με $\varepsilon=10\chi\lambda\mu$ (αριστερά) και $\varepsilon=3\chi\lambda\mu$ (δεξιά)	154
Εικόνα 33. Χάρτης με τις συσταδοποιημένες κλήσεις (4 συστάδες) ενός χρήστη. Πινέζες με το σύμβολο «N» αναπαριστούν θόρυβο.....	154
Εικόνα 34. Χάρτης με τις συσταδοποιημένες κλήσεις ενός χρήστη με υψηλό ποσοστό θορύβου.....	154
Εικόνα 35. Παράδειγμα εγγραφών από τον πίνακα κλήσεων, όπου έχουν ενσωματωθεί οι πιθανότητες του ταξινομητή για τις σημασιολογικές τοποθεσίες και έχει αποδοθεί η μεγαλύτερη	166

Ευρετήριο Γραφημάτων

Γράφημα 1. Εξέλιξη του μέσου πλήθους επαφών στο βιβλίο επαφών σε διαφορετικές ερευνητικές εργασίες.....	16
Γράφημα 2. Μέσος χρόνος εκτέλεσης πειράματος για $k=2$	71
Γράφημα 3. Μέσος χρόνος εκτέλεσης πειράματος για $k=3$	71
Γράφημα 4. Η εξέλιξη του μέσου πλήθους επαφών στο αποθετήριο μετά το πείραμα.....	74
Γράφημα 5. Μέσο μέγεθος των βιβλίων επαφών των χρηστών	75
Γράφημα 6. Διασπορά αγαπημένων επαφών σε σχέση με τις συνολικές επαφές	76
Γράφημα 7. Συσχέτιση αγαπημένων και συνολικών επαφών	77
Γράφημα 8. Πλήθος διακριτών επαφών στο ιστορικό κλήσεων (τελευταίες 500 κλήσεις)	77
Γράφημα 9. Ποσοστό κλήσεων προς τις top N επαφές σε συχνότητα επικοινωνίας ..	79
Γράφημα 10. Ποσοστό κλήσεων προς τις top N επαφές σε συχνότητα επικοινωνίας, χωρίς να υπολογίζονται οι κλήσεις προς και από νούμερα που δεν ανήκουν στις επαφές	80
Γράφημα 11. Σχέση εξερχόμενης επικοινωνίας και μεγέθους λίστας επαφών	81
Γράφημα 12. Διασπορά εξερχόμενης επικοινωνίας σε σχέση με το μέγεθος της λίστας επαφών	82
Γράφημα 13. Ποσοστό εξερχόμενων κλήσεων σε σχέση με το πλήθος των αγαπημένων επαφών.....	83
Γράφημα 14. Αναλογία επαφών που έλαβαν 80% κλήσεων προς την αναλογία κλήσεων στις πιο συχνά καλούμενες επαφές	85
Γράφημα 15. Κατανομή αθροιστικής πιθανότητας για τα ζεύγη του αριθμού των εξερχόμενων κλήσεων και του πλήθους των επαφών	86
Γράφημα 16. Κατανομή αθροιστικής πιθανότητας για τα ζεύγη του αριθμού των εξερχόμενων κλήσεων και του πλήθους των επαφών, ανά ομάδα χρηστών	88
Γράφημα 17. Ποσοστό επαφών στην υψηλή, απροσδιόριστη και χαμηλή περιοχή κατανομής πιθανότητας ανά ομάδα χρηστών	88
Γράφημα 18. Ποσοστό κλήσεων για τις οποίες υπήρξε επικοινωνία με την επαφή στις προηγούμενες N κλήσεις	89
Γράφημα 19. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της προσφατότητας ($w_f=0$, $w_r=1$, ισοδυναμεί με τη λίστα πρόσφατων κλήσεων).....	105
Γράφημα 20. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της συχνότητας ($w_f=1$, $w_r=0$, ισοδυναμεί με τη λίστα συχνών επαφών).....	105
Γράφημα 21. Επιτυχία αλγορίθμου για λίστα πρόβλεψης μιας θέσης. Ο μέσος όρος παρουσιάζεται με διακεκομμένη μαύρη γραμμή.....	106

Γράφημα 22. Επιτυχία αλγορίθμου για λίστα πρόβλεψης 3 θέσεων.....	106
Γράφημα 23. Επιτυχία αλγορίθμου για λίστα πρόβλεψης 5 θέσεων.....	107
Γράφημα 24. Επιτυχία αλγορίθμου για λίστα πρόβλεψης 8 θέσεων.....	107
Γράφημα 25. Μέση απόδοση πρόβλεψης για όλο τον πληθυσμό ανά μήκος λίστας προτάσεων	108
Γράφημα 26. Μέση απόδοση πρόβλεψης για όλα τα μήκη της λίστας προτάσεων ανά ομάδα χρηστών	108
Γράφημα 27. Απόδοση πρόβλεψης για κάθε ζεύγος ομάδας και μήκους λίστας προτάσεων	109
Γράφημα 28. Επίπτωση των επαφών με τις οποίες δεν υπάρχει προηγούμενη επικοινωνία στην επιτυχία των προβλέψεων.....	110
Γράφημα 29. Μέση απόδοση σκορ πρόβλεψης για όλο τον πληθυσμό των χρηστών	111
Γράφημα 30. Μέση απόδοση σκορ πρόβλεψης για όλα τα μήκη λιστών. Με μωβ το θεωρητικά βέλτιστο σκορ (5,33).....	112
Γράφημα 31. Απόδοση σκορ πρόβλεψης για όλα τα ζεύγη ομάδων και μήκους λίστας	112
Γράφημα 32. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της προσφατότητας ($w_f=0$, $w_r=1$, ισοδυναμεί με τη λίστα πρόσφατων κλήσεων).....	118
Γράφημα 33. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της συχνότητας ($w_f=1$, $w_r=0$, ισοδυναμεί με τη λίστα συχνών επαφών).....	119
Γράφημα 34. Μέση απόδοση πρόβλεψης για όλο τον πληθυσμό ανά μήκος λίστας προτάσεων. Με συνεχή γραμμή οι παραστάσεις από το προκαταρκτικό πείραμα και με διακεκομμένη από το πείραμα με το σύνολο της Nokia.....	120
Γράφημα 35. Μέση απόδοση πρόβλεψης για όλα τα μήκη της λίστας προτάσεων ανά ομάδα χρηστών	121
Γράφημα 36. Απόδοση πρόβλεψης για κάθε ζεύγος ομάδας και μήκους λίστας προτάσεων	121
Γράφημα 37. Μέση απόδοση σκορ πρόβλεψης για όλο τον πληθυσμό των χρηστών	122
Γράφημα 38. Μέση απόδοση σκορ πρόβλεψης για όλα τα μήκη λιστών. Με μωβ το θεωρητικά βέλτιστο σκορ (5,33).....	122
Γράφημα 39. Απόδοση σκορ πρόβλεψης για όλα τα ζεύγη ομάδων και μήκους λίστας	123
Γράφημα 40. Μέση απόδοση πρόβλεψης για το σύνολο των χρηστών για διαφορετικά μήκη παραθύρου εκπαίδευσης	124

Γράφημα 41. Μέση απόδοση σκορ πρόβλεψης για το σύνολο των χρηστών για διαφορετικά μήκη παραθύρου εκπαίδευσης.....	125
Γράφημα 42. Επίπτωση του παραθύρου εκπαίδευσης για την Ομάδα 3 και μήκος λίστας πρόβλεψης 8 θέσεων	125
Γράφημα 43. Σύγκριση προτεινόμενης προσέγγισης με άλλες ερευνητικές εργασίες στα δεδομένα του συνόλου της Nokia.....	126
Γράφημα 44. Απόδοση πρόβλεψης του αλγορίθμου σε σύγκριση με το (Stefanis, Plessas, Komninos & Garofalakis, 2013)	136
Γράφημα 45. Απόδοση πρόβλεψης για κάθε διάσταση ξεχωριστά για την ομάδα με τους χρήστες μεγαλύτερης κοινωνικότητας	144
Γράφημα 46. Σύγκριση της απόδοσης για τις τρεις συνδυασμένες διαστάσεις πλαισίου (συχνότητα, προσφατότητα, ώρα της ημέρας) για την ομάδα με τους χρήστες μεγαλύτερης κοινωνικότητας	145
Γράφημα 47. Ποσοστό γεω-χαρακτηρισμένων κλήσεων ανά χρήστη σε φθίνουσα κατάταξη	150
Γράφημα 48. Ποσοστό εξερχόμενων κλήσεων σε επαφές ανά συστάδα για τον χρήστη με ID 6168.....	157
Γράφημα 49. Ποσοστό εξερχόμενων κλήσεων σε επαφές ανά συστάδα για τον χρήστη με ID 5947.....	158
Γράφημα 50. Κατανομή συχνότητας εξερχόμενης επικοινωνίας για τον χρήστη με ID 5947.....	163
Γράφημα 51. Κατανομή συχνότητας εξερχόμενης επικοινωνίας για τον χρήστη με ID 5976.....	163
Γράφημα 52. Ποσοστό γεω-χαρακτηρισμένων κλήσεων ανά χρήστη σε φθίνουσα σειρά	166

Ευρετήριο Πινάκων

Πίνακας 1. Έρευνα για τις μεθόδους πραγματοποίησης κλήσεων από τους Lee, Seo και Lee (2010)	36
Πίνακας 2. Πιθανές τιμές για τις διαστάσεις που χρησιμοποιούνται για την πρόβλεψη στον αλγόριθμο των Lee, Seo και Lee (2010)	48
Πίνακας 3. Στατιστικά του συνόλου δεδομένων που χρησιμοποιήθηκε στο (Sun, Wang, Zheng & Hsu, 2013) για τη λίστα επαφών	52
Πίνακας 4. Σύνολα δεδομένων που χρησιμοποιήθηκαν στα πειράματα	69
Πίνακας 5. Χαρακτηριστικά συσκευών που χρησιμοποιήθηκαν στα πειράματα	69
Πίνακας 6. Μέσος χρόνος εκτέλεσης πειράματος ανά συσκευή και σύνολο δεδομένων σε msec για $k=2$	70
Πίνακας 7. Μέσος χρόνος εκτέλεσης πειράματος ανά συσκευή και σύνολο δεδομένων σε msec για $k=3$	70
Πίνακας 8. Επαφές με τις οποίες υπήρξε τηλεφωνική επικοινωνία ανά χρήστη για τη διάρκεια που καλύπτει το αρχείο κλήσεων	78
Πίνακας 9. Ομάδες επαφών (contact groups)	84
Πίνακας 10. Ομάδες χρηστών ανάλογα με τη χρονική διάρκεια του ιστορικού κλήσεων	87
Πίνακας 11. Αλγόριθμος αξιοποίησης πλαισίου για την πρόβλεψη εξερχόμενων κλήσεων	98
Πίνακας 12. Χαρακτηριστικά των χρηστών του συνόλου δεδομένων	101
Πίνακας 13. Κατανομή χρηστών σε ομάδες με βάση την τιμή της μετρικής της «κοινωνικότητας»	103
Πίνακας 14. Συγκεντρωτικά αποτελέσματα ανάλυσης ANOVA	113
Πίνακας 15. Σύγκριση αποτελεσμάτων πρόβλεψης με άλλες εργασίες	114
Πίνακας 16. Τύποι και ποσότητες δεδομένων που συλλέχθηκαν από την εκστρατεία της Nokia	116
Πίνακας 17. Σύγκριση των χαρακτηριστικών «κοινωνικότητας» των δύο συνόλων δεδομένων	118
Πίνακας 18. Σύνοψη πειραματικών μετρήσεων	133
Πίνακας 19. Ανάλυση των μεταβλητών ανά ομάδα «κοινωνικότητας» (2 = μέτρια «κοινωνικοί», 3 = περισσότερο «κοινωνικοί»)	135
Πίνακας 20. Συχνότητα αξιοποίησης παραδοσιακών μεθόδων ανάκτησης επαφών για έναρξη εξερχόμενων κλήσεων	137

Πίνακας 21. Σύγκριση της προτίμησης υπερ της εφαρμογής σε σχέση με τις παραδοσιακές μεθόδους ανάκτησης επαφών για την έναρξη εξερχόμενων κλήσεων	138
Πίνακας 22. Συνολικό ποσοστό γεω-χαρακτηρισμένων κλήσεων για διάφορες τιμές του εύρους.....	148
Πίνακας 23. Ο αλγόριθμος DBSCAN	152
Πίνακας 24. Κατανομή των κλήσεων στις γεωγραφικές συστάδες που σχηματίστηκαν ανά χρήστη	155
Πίνακας 25. Σύνοψη αποτελεσμάτων συσταδοποίησης γεω-χαρακτηρισμένων κλήσεων	156
Πίνακας 26. Πλήθος χρηστών ανά τύπο χρήστη με βάση την απόκλιση της επικοινωνίας	159
Πίνακας 27. Επίπτωση της φυσικής τοποθεσίας στην ακρίβεια πρόβλεψης (ΔLx) στα αποτελέσματα της βασικής απόδοσης (Bx) για 1, 3 και 5 θέσεις προτάσεων.....	161
Πίνακας 28. Κατανομή κλήσεων στις σημασιολογικές τοποθεσίες.....	167
Πίνακας 29. Πλήθος χρηστών ανά τύπο χρήστη με βάση την απόκλιση της επικοινωνίας	168
Πίνακας 30. Επίπτωση της σημασιολογικής τοποθεσίας στην ακρίβεια πρόβλεψης (ΔLx) στα αποτελέσματα της βασικής απόδοσης (Bx) για 1, 3 και 5 θέσεις προτάσεων	169

1. Εισαγωγή

1.1 Ορισμοί

Για τον όρο διαχείριση προσωπικής πληροφορίας (Personal Information Management - PIM) έχουν προταθεί στην ερευνητική βιβλιογραφία πολλοί διαφορετικοί ορισμοί. Για παράδειγμα, ένας ενδεικτικός ορισμός είναι ο παρακάτω (Boardman, 2004):

«Η διαχείριση προσωπικής πληροφορίας είναι ένας όρος που χρησιμοποιείται για να περιγράψει τη συλλογή, αποθήκευση, οργάνωση και ανάκτηση ψηφιακών αντικειμένων πληροφορίας (π.χ. μηνύματα ηλεκτρονικού ταχυδρομείου, αρχεία, υπενθυμίσεις, επαφές, σελιδοδείκτες κ.α.) από ένα άτομο στο προσωπικό του υπολογιστικό περιβάλλον».

Ο Lansdale (1988) ορίζει τη διαχείριση προσωπικής πληροφορίας ως «τις μεθόδους και διαδικασίες με τις οποίες χειριζόμαστε, κατηγοριοποιούμε και ανακτούμε πληροφορία σε καθημερινή βάση».

Ένας ακόμα ορισμός που δεν περιορίζεται στον ψηφιακό κόσμο, είναι ο ακόλουθος (Jones, 2007):

«Η διαχείριση προσωπικής πληροφορίας αναφέρεται τόσο στην πρακτική όσο και στη μελέτη των δραστηριοτήτων που πραγματοποιεί ένα άτομο με σκοπό να αποκτήσει ή δημιουργήσει, αποθηκεύσει, οργανώσει, διατηρήσει, ανακτήσει, χρησιμοποιήσει και διανείμει την πληροφορία που είναι απαραίτητη για την ολοκλήρωση εργασιών και να εκπληρώσει διάφορους ρόλους και ευθύνες (για παράδειγμα ως γονιός, εργαζόμενος, φίλος ή μέλος μιας κοινότητας)».

Ο όρος προσωπική πληροφορία (personal information) μπορεί να έχει μια από τις παρακάτω έννοιες (Jones, 2007):

1. Πληροφορία που οι άνθρωποι διατηρούν για προσωπική χρήση.
2. Πληροφορία για κάποιο άτομο που βρίσκεται υπό τον έλεγχο τρίτων (π.χ. τα δεδομένα που διατηρεί ένας γιατρός για τους ασθενείς του).
3. Πληροφορία που αφορά εμπειρίες ενός ατόμου, αλλά βρίσκεται εκτός του ελέγχου του (π.χ. η καταγραφή του δανεισμού ενός βιβλίου που ο χρήστης της βιβλιοθήκης έχει επιστρέψει).

Στη συνέχεια του κειμένου ως προσωπική πληροφορία θεωρείται πληροφορία που αφορά την πρώτη από τις παραπάνω περιπτώσεις.

Με τον όρο στοιχεία πληροφορίας (information items) περιγράφονται «πακέτα» πληροφορίας όπως έντυπα έγγραφα, ηλεκτρονικά αρχεία, μηνύματα ηλεκτρονικού ταχυδρομείου, φωτογραφίες, διευθύνσεις ιστοσελίδων κ.α. (Jones, 2007).

Η διαχείριση προσωπικής πληροφορίας δίνει ιδιαίτερη έμφαση στην οργάνωση και διατήρηση συλλογών προσωπικής πληροφορίας (personal information collections), στις οποίες σύνολα από στοιχεία προσωπικής πληροφορίας (PIM items) αποθηκεύονται (συνήθως στην ίδια μορφή – π.χ. μηνύματα ηλεκτρονικού ταχυδρομείου, επαφές κ.α.) για μελλοντική και ενδεχομένως επαναλαμβανόμενη χρησιμοποίηση (Jones, 2007).

Ένα εργαλείο διαχείρισης προσωπικής πληροφορίας (PIM-tool) ορίζεται ως «ένα εργαλείο λογισμικού που επιτρέπει στον χρήστη τη διαχείριση μιας συλλογής από στοιχεία προσωπικής πληροφορίας» (Boardman, 2004). Τα εργαλεία αυτά προσφέρουν μηχανισμούς:

- για την προσθήκη στοιχείων πληροφορίας στη συλλογή (απόκτηση)
- για την οργάνωση των στοιχείων πληροφορίας στη συλλογή (κατηγοριοποίηση)
- για τη διαγραφή στοιχείων από τη συλλογή (συντήρηση)
- για την πρόσβαση στα στοιχεία (ανάκτηση)

Ο όρος κινητός υπολογισμός (mobile computing) αναφέρεται «στη χρήση υπολογιστικών πόρων σε ένα κινητό περιβάλλον, χωρίς δηλαδή ο χρήστης να βρίσκεται συνεχώς στην ίδια τοποθεσία (χώρος εργασίας)» (Ancona, Dodero & Gianuzzi, 1999).

Οι συσκευές που έχουν καταστήσει εφικτό τον κινητό υπολογισμό, επιτρέποντας στους χρήστες τους να τις χρησιμοποιούν ενώ βρίσκονται σε κίνηση, ονομάζονται συσκευές κινητού υπολογισμού (mobile computing devices). Σε αυτή την κατηγορία περιλαμβάνονται συσκευές όπως τα έξυπνα κινητά τηλέφωνα (smart phones), τα PDAs, τα Tablets ή ακόμα και φορετές συσκευές όπως τα έξυπνα ρολόγια (smart watches).

Όσον αφορά το πλαίσιο, δεν υπάρχει ένας κοινά αποδεκτός ορισμός. Στη βιβλιογραφία μπορεί να βρει κανείς πολλούς διαφορετικούς ορισμούς, μεγάλο μέρος των οποίων είναι αρκετά ευρείς, γεγονός που υποδεικνύει ότι πρόκειται για μια πιο

«χαλαρή» έννοια που προσαρμόζεται ανάλογα με το πεδίο έρευνας. Ίσως ο πιο δημοφιλής είναι αυτός που προτάθηκε από τον Dey (2001):

«Το πλαίσιο (context) είναι οποιαδήποτε πληροφορία μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει την κατάσταση μιας οντότητας. Μια οντότητα είναι ένα άτομο, μια τοποθεσία ή ένα αντικείμενο που θεωρείται σχετικό με την αλληλεπίδραση ανάμεσα σε έναν χρήστη και μια εφαρμογή, συμπεριλαμβανομένου του χρήστη και της εφαρμογής».

Ένα σύστημα είναι ενήμερο για το πλαίσιο (context aware) εάν χρησιμοποιεί το πλαίσιο για να παρέχει σχετική πληροφορία και/ή υπηρεσίες στον χρήστη, όπου η σχετικότητα εξαρτάται από την εργασία του χρήστη (Dey, 2001). Όταν ο χρήστης βρίσκεται σε κίνηση και η συσκευή του είναι κινητή συσκευή, η ενημερότητα πλαισίου ονομάζεται κινητή ενημερότητα πλαισίου (mobile context awareness) (Dey & Häkkinen, 2008).

Σε ακόμα μια δημοφιλή δημοσίευση (Chen & Kotz, 2000), διακρίνονται δύο πλευρές που αφορούν το πλαίσιο για την περίπτωση των κινητών εφαρμογών: η μια πλευρά περιλαμβάνει τα χαρακτηριστικά του περιβάλλοντος που καθορίζουν τη συμπεριφορά των κινητών εφαρμογών (ενεργό πλαίσιο), ενώ η άλλη πλευρά αφορά το πλαίσιο που σχετίζεται με την εφαρμογή, αλλά δεν είναι κρίσιμο (παθητικό πλαίσιο). Το πλαίσιο ορίζεται ως εξής:

«Πλαίσιο είναι το σύνολο των περιβαλλοντικών καταστάσεων και ρυθμίσεων που είτε καθορίζει τη συμπεριφορά μιας εφαρμογής, είτε στο οποίο ένα γεγονός της εφαρμογής λαμβάνει χώρα και έχει ενδιαφέρον για τον χρήστη».

Οι ερευνητές δίνουν δύο ορισμούς για την ενήμερη για το πλαίσιο υπολογιστική (context aware computing):

«Ενεργή ενημερότητα πλαισίου (active context awareness): μια εφαρμογή αυτόματα προσαρμόζεται στο ανακαλυφθέν πλαίσιο, αλλάζοντας τη συμπεριφορά της».

«Παθητική ενημερότητα πλαισίου (passive context awareness): μια εφαρμογή παρουσιάζει το νέο ή ενημερωμένο πλαίσιο στον ενδιαφερόμενο χρήστη ή του επιτρέπει να το ανακτήσει αργότερα».

1.2 Προβλήματα διαχείρισης προσωπικής πληροφορίας και συσκευές κινητού υπολογισμού

1.2.1 Προβλήματα στη διαχείριση προσωπικής πληροφορίας

Σύμφωνα με τον Whittaker (2011), ο κύκλος ζωής για τη διαχείριση προσωπικής πληροφορίας περιλαμβάνει 3 στάδια:

- *Διατήρηση (Keeping)*: οι άνθρωποι αποφασίζουν να κρατήσουν ένα μόνο μέρος από το σύνολο της πληροφορίας με την οποία έρχονται αντιμέτωποι καθημερινά. Πληροφορία την οποία δεν έχουν κρατήσει, δεν μπορούν ούτε να διαχειριστούν ούτε να αξιοποιήσουν.
- *Διαχείριση (Management)*: το στάδιο αυτό αναφέρεται στον τρόπο με τον οποίο οι άνθρωποι οργανώνουν την πληροφορία που αποφασίζουν να κρατήσουν, έτσι ώστε να διευκολύνουν τη μελλοντική ανάκτησή της. Μια απόφαση – κλειδί που καλούνται να πάρουν αφορά την αντιστάθμιση ανάμεσα στην προσπάθεια που θα αφιερώσουν για την οργάνωση της πληροφορίας και το προβλεπόμενο κόστος για την αξιοποίησή της.
- *Αξιοποίηση (Exploitation)*: αυτό το στάδιο αναφέρεται στην εύκολη πρόσβαση στην κατάλληλη πληροφορία την κατάλληλη στιγμή, ώστε αυτή να αξιοποιηθεί.

Σε μια ιδανική εφαρμογή της διαχείρισης προσωπικής πληροφορίας, οι χρήστες θα έπρεπε να έχουν πάντα διαθέσιμη την κατάλληλη πληροφορία στο σωστό μέρος και στην κατάλληλη μορφή, σε εκείνη την ποιότητα και την πληρότητα που απαιτείται για να ικανοποιήσουν τις ανάγκες τους (Jones, 2007). Η πραγματικότητα, εντούτοις, απέχει πολύ από αυτό το ιδανικό σενάριο, καθώς συχνά οι άνθρωποι αδυνατούν να βρουν την κατάλληλη πληροφορία για την ικανοποίηση αυτών των αναγκών (Teenan & Jones, 2008). Η απαραίτητη πληροφορία μπορεί να μη βρεθεί ποτέ ή να βρεθεί πολύ αργά για να είναι χρήσιμη. Επίσης, η πληροφορία μπορεί να αγνοηθεί, να ξεχαστεί ή να αποθηκευθεί σε λάθος σημείο, ώστε να μην είναι δυνατή η αξιοποίησή της όταν χρειάζεται. Η πληροφορία δε βρίσκεται πάντα στο σωστό σημείο: η πληροφορία που χρειάζεται κανείς όταν είναι στο γραφείο μπορεί να βρίσκεται στο σπίτι και το αντίστροφο. Μπορεί να είναι σε άλλον υπολογιστή, έξυπνο τηλέφωνο, PDA ή άλλη συσκευή. Η ψηφιακή πληροφορία μπορεί να είναι περαιτέρω διεσπαρμένη σε «νησίδες πληροφορίας», που η κάθε μια υποστηρίζεται από διαφορετική εφαρμογή και διαφορετική συσκευή, ένα πρόβλημα που ονομάζεται

«κατακερματισμός της πληροφορίας» (information fragmentation) (Teevan, Jones & Bederson, 2006; Teevan & Jones, 2008; Jones, 2004; Jones, 2007).

Οι άνθρωποι αποθηκεύουν και διατηρούν συνεχώς προσωπική πληροφορία. Για παράδειγμα, κανείς δε σβήνει κάθε μήνυμα ηλεκτρονικού ταχυδρομείου που λαμβάνει, ούτε διαγράφει ένα έγγραφο ή μια παρουσίαση που μόλις δημιούργησε. Υπάρχουν αρκετές ερευνητικές εργασίες που παρουσιάζουν παραδείγματα από την τεράστια ποσότητα πληροφορίας που οι χρήστες διατηρούν στους προσωπικούς τους χώρους αποθήκευσης: μελέτες που αφορούν τα μηνύματα ηλεκτρονικού ταχυδρομείου αποκάλυψαν ότι οι χρήστες αποθήκευαν χιλιάδες μηνύματα (Whittaker & Sidner, 1996), ο αριθμός των οποίων σχεδόν δεκαπλασιάστηκε μέσα σε μια δεκαετία (Fisher, Brush, Gleave & Smith, 2006). Σε μια άλλη έρευνα, οι Boardman και Sasse (2004) βρήκαν ότι οι χρήστες αποθήκευαν κατά μέσο όρο 2.200 αρχεία στους σκληρούς τους δίσκους. Μια πιο πρόσφατη έρευνα (Whittaker, Bergman & Clough, 2010) που αφορούσε ψηφιακές φωτογραφίες εντόπισε κατά μέσο όρο περισσότερες από 4.000 προσωπικές φωτογραφίες. Άλλες μελέτες δείχνουν ότι οι χρήστες αποθηκεύουν εκατοντάδες σελιδοδείκτες (Aula, Jhaveri & Käki, 2005) και επαφές (Bentley & Chen, 2015).

Είναι γνωστό ότι οι άνθρωποι δυσκολεύονται να αποφασίσουν ποια πληροφορία χρειάζεται να κρατήσουν (Bergman, Tucker, Beyth-Marom, Cutrell & Whittaker, 2009; Boardman & Sasse, 2004; Jones, 2007). Ένας λόγος γι' αυτό το φαινόμενο είναι η δυσκολία να προβλέψουν το μέλλον και συγκεκριμένα τις μελλοντικές τους πληροφοριακές ανάγκες (Whittaker, 2011). Η απόφαση για την πληροφορία που πρέπει να διατηρήσει κανείς ή όχι μπορεί να οδηγήσει σε δύο τύπους λαθών που έχουν σημαντικό κόστος (Jones, 2004): πληροφορία που δεν αποθηκεύεται δεν είναι δυνατό να χρησιμοποιηθεί αν χρειαστεί, ενώ η διατήρηση άσχετης πληροφορίας δυσκολεύει την οργάνωση (Boardman & Sasse, 2004; Whittaker & Sidner, 1996), αλλά ταυτόχρονα αυξάνει το χρόνο ανάκτησης, αποσπώντας την προσοχή του χρήστη. Είναι γνωστό στον τομέα της ψυχολογίας ότι στην οπτική αναζήτηση το πλήθος των άσχετων αντικειμένων που αποσπούν την προσοχή αυξάνει το χρόνο που απαιτείται για να εντοπίσει κανείς ένα αντικείμενο-στόχο (Treisman & Gelade, 1980). Πιο συγκεκριμένα, στον τομέα της διαχείρισης προσωπικής πληροφορίας έχει αποδειχθεί ότι υπάρχει μια θετική συσχέτιση ανάμεσα στο πλήθος των αρχείων στους φακέλους και στον χρόνο ανάκτησης ενός αρχείου – στόχου (Bergman, Whittaker, Sanderson, Nachmias & Ramamoorthy, 2010). Επιπλέον, εμφανίζεται το «παράδοξο της διαγραφής» (Bergman, Beyth Marom & Nachmias, 2003; Bergman, Tucker, Beyth-Marom, Cutrell & Whittaker, 2009): ενώ τα ασήμαντα ή άσχετα στοιχεία

πληροφορίας αποσπούν την προσοχή και αυξάνουν το χρόνο ανάκτησης της σημαντικής πληροφορίας, η αξιολόγησή τους από τους χρήστες για το αν χρειάζεται να διαγραφούν ή όχι απαιτεί χρόνο και προσπάθεια, με αποτέλεσμα να την αναβάλλουν συνεχώς. Συνήθως, πάντως, οι χρήστες αποφασίζουν να κρατήσουν την πληροφορία (Bergman, Komninos, Liaroikaris & Clarke, 2012) και αυτό συμβαίνει γιατί κανείς μπορεί πάντα να σκεφθεί μια περίπτωση που να του χρειαστεί (Whittaker & Hirschberg, 2001). Μάλιστα, αρκετοί ερευνητές έχουν προβλέψει πως λόγω του φθηνού αποθηκευτικού χώρου και του κόστους για την απόφαση διατήρησης ή διαγραφής πληροφορίας, η ανθρώπινη συμπεριφορά θα μετατοπιστεί σε ένα μοντέλο όπου δε θα διαγράφεται καθόλου πληροφορία (Jones, 2004; Marshall 2008a,b), αλλά θα διατηρούνται τα πάντα και σε αυτή την περίπτωση η διαχείριση της προσωπικής πληροφορίας θα υποστηρίζεται από διεπαφές που θα ξεχωρίζουν τα σημαντικά και πολύτιμα δεδομένα.

1.2.2 Συσκευές κινητού υπολογισμού και προκλήσεις

Η πρώτη επανάσταση στη διαχείριση πληροφορίας έλαβε χώρα με τη διάδοση των υπολογιστών και η δεύτερη με τη διάδοση των δικτύων υπολογιστών και κυρίως του διαδικτύου. Στη σημερινή εποχή, η κοινωνία μας γίνεται μάρτυρας μιας μεγάλης ανάπτυξης στον τομέα της προσωπικής πληροφορίας, καθώς βιώνει το τρίτο κύμα τεχνολογικής καινοτομίας: την άφιξη των πληροφοριακών συσκευών κινητού υπολογισμού (Tungare & Perez-Quinones, 2008b).

Η διάδοση των συσκευών κινητού υπολογισμού είχε ως αποτέλεσμα, όσον αφορά τη διαχείριση προσωπικής πληροφορίας, τη μεγέθυνση των παραπάνω προκλήσεων και τη δημιουργία νέων, λόγω των ιδιοτήτων που τις χαρακτηρίζουν σε σχέση με τους σταθερούς υπολογιστές. Παρόλο που οι συσκευές αυτές έχουν σημαντικά αυξημένη υπολογιστική ισχύ σε σχέση με τα προηγούμενα χρόνια, η αποθηκευτική χωρητικότητά τους αυξάνεται διαρκώς, υπάρχει η δυνατότητα να είναι συνεχώς συνδεδεμένες στο διαδίκτυο και είναι ιδιαίτερα προσιτές από οικονομικής άποψης στον μέσο χρήστη, υπάρχουν περιορισμοί που δεν αναμένεται να εξαλειφθούν σύντομα. Οι οθόνες τους παραμένουν σημαντικά μικρότερες από αυτές των σταθερών υπολογιστών, περιορίζοντας την ποσότητα πληροφορίας που μπορούν να παρουσιάσουν ταυτόχρονα. Η ανάγνωση μεγάλης ποσότητας πληροφορίας σε αυτές τις συσκευές απαιτεί συχνή οριζόντια κύλιση και εστιασμένη συγκέντρωση (Gong & Tarasewich, 2004). Το γεγονός αυτό, σε συνδυασμό με άλλους περιορισμούς των συσκευών που παρατίθενται παρακάτω, τη μεγάλη ποσότητα προσωπικής πληροφορίας που όπως παρουσιάστηκε αποθηκεύουν οι χρήστες και την απροθυμία τους να αφιερώσουν χρόνο και προσπάθεια για τη διαγραφή προσωπικών δεδομένων

(Boardman & Sasse, 2004), έχει ως αποτέλεσμα να δυσχεραίνεται η ανάκτηση και επομένως η αξιοποίηση της πληροφορίας (Zhou, Mohammed & Zhang, 2012). Επιπλέον, ο εξοπλισμός και οι εφαρμογές που ενσωματώνονται σε αυτές τις συσκευές επιτρέπουν την εύκολη δημιουργία ακόμα μεγαλύτερης ποσότητας προσωπικής πληροφορίας (Viana, Miron, Moisuc, Gensel, Villanova-Oliver & Martin, 2011; Jan, Li & Al-Raweshidy, 2011; Hang, Von Zezschwitz, De Luca & Hussmann, 2012), όπως για παράδειγμα στην περίπτωση των φωτογραφιών και των βίντεο που οι χρήστες αποθηκεύουν μέσω της κάμερας (Whittaker, Bergman & Clough, 2010) ή των σημειώσεων που δημιουργούν με κάποια αντίστοιχη εφαρμογή (Dai, Lutters & Bower, 2005), οπότε η πρόσβαση στην κατάλληλη πληροφορία την κατάλληλη στιγμή γίνεται ακόμα πιο δύσκολη (Jan, Li & Al-Raweshidy, 2011).

Επιπλέον, οι τρόποι εισόδου δεδομένων διαφέρουν σημαντικά: συνήθως αυτές οι συσκευές είναι εξοπλισμένες με οθόνες αφής ή πληκτρολόγια πολύ μικρότερου μεγέθους σε σχέση με αυτά των σταθερών υπολογιστών, ενώ άλλοι τρόποι εισόδου πληροφορίας περιλαμβάνουν τη χρήση φωνής ή την αναγνώριση χειρονομιών. Αν και η τεχνολογία των συσκευών κινητού υπολογισμού έχει αναπτυχθεί με γρήγορους ρυθμούς, οι μέθοδοι εισόδου δεν έχουν καταφέρει να φτάσουν σε επίπεδο εφάμιλλο των παραδοσιακών πληκτρολογίων (Page, 2013). Συνεπώς, σε συνδυασμό και με τα υπόλοιπα χαρακτηριστικά των συσκευών (π.χ. κινητικότητα χρήστη) προκαλούνται δυσκολίες στη δημιουργία στοιχείων προσωπικής πληροφορίας (στις περιπτώσεις που απαιτείται καταχώρηση δεδομένων) και στην οργάνωσή τους (π.χ. οργάνωση σε ιεραρχικές δομές, μετακίνηση σε άλλη τοποθεσία, προσθήκη ετικετών κ.α.) (Karypidis & Lalis, 2007).

Επίσης, το πλαίσιο χρήσης αυτών των συσκευών είναι συχνά μεταβαλλόμενο (π.χ. καθώς οι χρήστες κινούνται αλλάζει η τοποθεσία τους, οι συσκευές και οι άνθρωποι με τους οποίους αλληλεπιδρούν, καθώς και οι στόχοι και οι ανάγκες τους) (Dey & Häkklä, 2008) και πιο απαιτητικό σε σχέση με τη χρήση των σταθερών υπολογιστών (Kristoffersen & Ljungberg, 1999; Pascoe, Ryan & Morse, 2000) και αυτό επηρεάζει τη διαχείριση προσωπικής πληροφορίας (Tungare & Perez-Quinones, 2008b):

- Η χρήση της συσκευής είναι συχνά δευτερεύουσα εργασία. Εργασίες εξωτερικές σε σχέση με τη λειτουργία της συσκευής είναι πιο σημαντικές.
- Τα χέρια των χρηστών χρησιμοποιούνται συχνά για να διαχειρίζονται φυσικά αντικείμενα, σε αντίθεση με τους χρήστες που εργάζονται σε περιβάλλον γραφείου που τα χέρια τους βρίσκονται εργονομικά τοποθετημένα στο πληκτρολόγιο.

- Οι εργασίες που αναλαμβάνουν οι χρήστες μπορεί να απαιτούν υψηλό επίπεδο οπτικής προσοχής σε αντίθεση με το παραδοσιακό περιβάλλον γραφείου, όπου μεγάλο μέρος της προσοχής του χρήστη αφιερώνεται στον υπολογιστή.
- Οι χρήστες μπορεί να παρουσιάζουν υψηλή κινητικότητα κατά τη διάρκεια εκτέλεσης των εργασιών.

Επιπροσθέτως, οι χρήστες δεν έχουν συνήθως την ικανότητα να αφιερώσουν επαρκή προσήλωση στη διατήρηση και οργάνωση της πληροφορίας που συναντούν σε ένα κινητό πλαίσιο (Oulasvirta, Tamminen, Roto & Kuorelahti, 2005).

Καθώς οι χρήστες χρησιμοποιούν πολλαπλές διαφορετικές συσκευές για τη διαχείριση της προσωπικής τους πληροφορίας, το πρόβλημα του κατακερματισμού γίνεται ολοένα πιο έντονο (Tungare & Pérez-Quinones, 2008a). Κανείς μπορεί στο γραφείο να χρησιμοποιεί έναν σταθερό υπολογιστή, ενώ κατά τη διάρκεια ενός συνεδρίου ή μιας συνάντησης ένα λάπτοπ. Ο ίδιος χρήστης μπορεί χειρίζεται ένα έξυπνο κινητό τηλέφωνο ή ένα PDA στο δρόμο της επιστροφής στο σπίτι. Αυτή η κατάσταση προκαλεί δυσκολίες στους χρήστες όσον αφορά την εκτέλεση εργασιών που σχετίζονται με την πληροφορία (Tungare & Pérez-Quinones, 2008a), ενώ ταυτόχρονα η διαχείριση της προσωπικής πληροφορίας μετατρέπεται σε μια περίπλοκη προσπάθεια (Tungare & Pérez-Quinones, 2009). Ο συγχρονισμός της πληροφορίας σε όλες τις συσκευές είναι μια πολύπλοκη διαδικασία και πολλές φορές όχι επιθυμητή (Tungare & Perez-Quinones, 2008b). Συχνά, δεν είναι καν εφικτός, καθώς οι συσκευές αποτελούν νησίδες όπου η πληροφορία αποθηκεύεται σε διαφορετικές μορφές και η διαλειτουργικότητα δεν υποστηρίζεται πλήρως. Ένα τέτοιο παράδειγμα είναι η φορητή συσκευή πολυμέσων iPod της Apple που ο κατασκευαστής έχει επιλέξει να συνεργάζεται με τη βιβλιοθήκη iTunes. Άλλο παράδειγμα είναι η δυσκολία μεταφοράς επαφών από μια συσκευή Android σε συσκευές με παλιότερες εκδόσεις Windows Phone.

Ακόμα όμως και στην ίδια συσκευή, τα δομημένα αποθετήρια πληροφορίας παραμένουν «περιφραγμένοι κήποι» (walled gardens) (Kominos, Plessas, Stefanis & Garofalakis, 2011a), απαιτώντας από τον χρήστη να επισκεφθεί κάθε ένα με τη σειρά, ώστε να συγκεντρώσει τα στοιχεία πληροφορίας που χρειάζεται στα πλαίσια της τρέχουσας δραστηριότητας σε μια συνεκτική διανοητική συλλογή. Καθώς το μέγεθος της προσωπικής πληροφορίας αυξάνει, ακόμα και θεωρώντας κάθε αποθετήριο ξεχωριστά, γίνεται γρήγορα εμφανές ότι οι παραδοσιακές μέθοδοι πρόσβασης στην πληροφορία είναι ανεπαρκείς, λαμβάνοντας υπόψη το συνεχώς αυξανόμενο

αποθηκευτικό χώρο των συσκευών και τον πλούτο της προσωπικής πληροφορίας που αποθηκεύεται εκεί.

1.3 Αξιοποίηση του κινητού κοινωνικού πλαισίου

Η εξέλιξη της τεχνολογίας επιτρέπει στις συσκευές κινητού υπολογισμού να συλλέγουν πλήθος από δεδομένα που περιγράφουν το φυσικό περιβάλλον – πλαίσιο στο οποίο βρίσκεται ο χρήστης τους, δηλαδή το κινητό πλαίσιο του χρήστη (mobile context) (Koshman, 2011). Τέτοια δεδομένα περιλαμβάνουν την τοποθεσία (σχετική ή απόλυτη), την τρέχουσα ώρα, το αν ο χρήστης κινείται και την ταχύτητά του, τη δραστηριότητά του, τον προσανατολισμό της συσκευής, εάν η δόνηση ή το σιωπηλό προφίλ είναι ενεργοποιημένα κ.α. (Beach, Gartrell, Xing, Han, Lv, Mishra & Seada, 2010). Η συλλογή των δεδομένων μπορεί να γίνει είτε με άμεσο τρόπο από τους αισθητήρες που είναι ενσωματωμένοι στις συσκευές, είτε με έμμεσο (π.χ. μετά από διαδικτυακά ερωτήματα). Αισθητήρες που περιλαμβάνονται στις σύγχρονες συσκευές κινητού υπολογισμού είναι οι παρακάτω (Lane, Miluzzo, Lu, Peebles, Choudhury & Campbell, 2010):

- Μικρόφωνο
- Κάμερα
- GPS
- Ψηφιακή πυξίδα
- Γυροσκόπιο
- Επιταχυνσιόμετρο
- Αισθητήρας εγγύτητας
- Αισθητήρας φωτός περιβάλλοντος
- Αισθητήρας θερμοκρασίας και υγρασίας
- Αισθητήρας κίνησης
- Bluetooth
- WiFi
- Μετρητής πίεσης της ατμόσφαιρας
- Αναγνώστης δακτυλικών αποτυπωμάτων

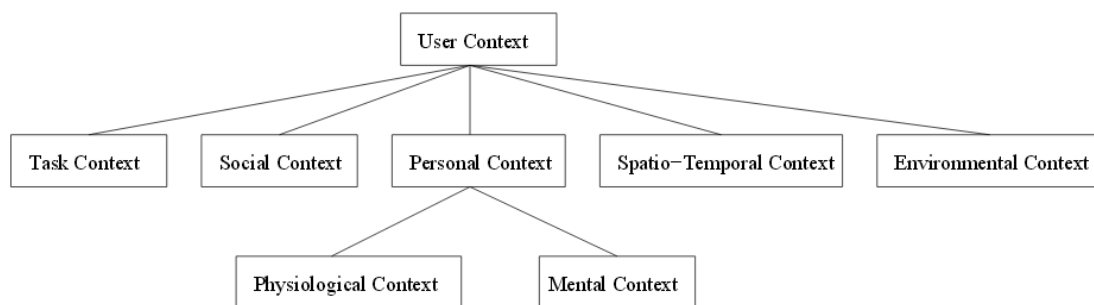
Οι κατηγορίες των πληροφοριών πλαισίου που μπορούν να εξαχθούν από τις συσκευές είναι (Abowd, Dey, Brown, Davies, Smith & Steggles, 1999; Chamasemani & Affendey, 2014):

- *Περιβαλλοντικό ή φυσικό πλαίσιο (Environmental or Physical Context).* Περιλαμβάνει τις περιβαλλοντικές συνθήκες της τρέχουσας τοποθεσίας.

Αισθητήρες επιτρέπουν τη συλλογή δεδομένων θερμοκρασίας, υγρασίας, επιπέδων θορύβου, συνθηκών φωτισμού κ.α.

- *Πλαίσιο δραστηριότητας (Activity Context)*. Ορίζει την τρέχουσα δραστηριότητα του χρήστη. Για παράδειγμα, η συσκευή μπορεί να αναγνωρίσει ότι ο χρήστης περπατάει, τρέχει ή οδηγεί, ότι μιλάει ή χρησιμοποιεί κάποια εφαρμογή.
- *Χρονικό πλαίσιο (Temporal Context)*. Αποτελείται από χρονικούς παράγοντες όπως η ώρα της ημέρας, η μέρα της εβδομάδας, η ημερομηνία κ.λπ.
- *Πλαίσιο προσωπικής ταυτότητας (Personal Identity Context)*. Προσδιορίζει τα χαρακτηριστικά του χρήστη, όπως την ηλικία του, τα στοιχεία επικοινωνίας του, τις προτιμήσεις του κ.α.
- *Χωρικό πλαίσιο (Spatial Context)*. Αφορά οποιαδήποτε πληροφορία σχετίζεται με την τοποθεσία του χρήστη, όπως φυσική τοποθεσία, σημασιολογική τοποθεσία, προσανατολισμός, ταχύτητα, επιτάχυνση.
- *Πλαίσιο ζωτικών σημάτων (Vital Signs Context)*. Καλύπτει όλες τις πληροφορίες που σχετίζονται με την κατάσταση της υγείας του χρήστη, όπως ο καρδιακός ρυθμός, η πίεση κ.λπ.

Ένας εναλλακτικός, αλλά παρόμοιος διαχωρισμός των κατηγοριών πλαισίου που προτάθηκε στα πλαίσια του έργου AmbieSense (Mikalsen & Kofod-Petersen, 2004) παρουσιάζεται στην εικόνα 1.



Εικόνα 1. Κατηγορίες πλαισίου στο έργο AmbieSense (Mikalsen & Kofod-Petersen, 2004)

Οι χρήστες θεωρούν τις συσκευές κινητού υπολογισμού «έμπιστες συσκευές», συνήθως τις μεταφέρουν μαζί τους, ενώ συχνά βρίσκονται σε 24ωρη λειτουργία (Raento, Oulasvirta, Petit & Toivonen, 2005). Επιπλέον, οι συσκευές κινητού υπολογισμού περιέχουν μεγάλες ποσότητες προσωπικής πληροφορίας που σχετίζεται με το κοινωνικό περιβάλλον των χρηστών (Toninelli, Khushraj, Lassila & Montanari, 2008). Η πληροφορία αυτή συχνά δημιουργείται αυτόματα από τις ίδιες τις συσκευές. Για παράδειγμα, για κάθε κλήση που πραγματοποιείται σε ένα κινητό τηλέφωνο

αποθηκεύεται στη λίστα κλήσεων μια σχετική εγγραφή, η οποία περιλαμβάνει την ημερομηνία και την ώρα, την επαφή με την οποία έγινε και τη διάρκεια της. Επίσης, μέρος της πληροφορίας δημιουργείται από τους ίδιους τους χρήστες: SMS και MMS, φωτογραφίες, βίντεο και αρχεία ήχου, γεγονότα ημερολογίου, επαφές κ.α. Συνεπώς, οι συσκευές αυτές εκτός από το κινητό πλαίσιο των χρηστών, είναι ενήμερες και για το κοινωνικό περιβάλλον τους (κοινωνικό πλαίσιο – social context) (Absar, O'Brien & Webster, 2014) και τις κοινωνικές τους αλληλεπιδράσεις. Ο συνδυασμός του κινητού και του κοινωνικού περιβάλλοντος έχει ως αποτέλεσμα ένα δυναμικά ορισμένο κοινωνικό πλαίσιο, γνωστό ως κινητό κοινωνικό πλαίσιο (mobile social context) (Gilbert, Cuervo & Cox, 2009).

Η προσωπική πληροφορία σχετίζεται σε μεγάλο βαθμό με το πλαίσιο, κάτι που ισχύει και για τις δραστηριότητες διαχείρισής της (Huvila, Eriksen, Häusner & Jansson, 2014). Οι πληροφοριακές ανάγκες των χρηστών εξαρτώνται όχι μόνο από το κινητό πλαίσιο (Church & Smyth, 2009) αλλά και από τις κοινωνικές τους αλληλεπιδράσεις (Heimonen, 2009), ενώ το ίδιο ισχύει και για την πληροφοριακή συμπεριφορά τους όταν χρησιμοποιούν κινητές συσκευές (Koshman, 2011; Absar, O'Brien & Webster, 2014). Επιπλέον, το πλαίσιο έχει αναγνωριστεί ως ένας σημαντικός παράγοντας που επηρεάζει τις αποφάσεις διατήρησης και οργάνωσης της πληροφορίας (Barreau, 1995). Ο Whittaker (2011) υπονόησε ότι το κινητό και κοινωνικό πλαίσιο μπορεί να χρησιμοποιηθεί για τη δημιουργία διαφορετικών προβολών προσωπικής πληροφορίας ανάλογα με το πλαίσιο στο οποίο βρίσκεται ο χρήστης, βελτιώνοντας έτσι την αξιοποίηση της προσωπικής πληροφορίας. Άλλοι ερευνητές επισήμαναν πως τα συστήματα ανάκτησης πληροφορίας χρειάζεται να λαμβάνουν υπόψη το πλαίσιο των χρηστών για να ανταποκριθούν αποτελεσματικά στις ανάγκες τους για πληροφορία (Allan, Croft, Moffat & Sanderson, 2012; Bergman, Beyth Marom & Nachmias, 2008). Ο Singh (2006) υπογράμμισε πως το πλαίσιο του κινητού χρήστη διαδραματίζει ένα ρόλο κλειδί στην αναγνώριση και διαχείριση πληροφορίας που χρήζει άμεσης αντιμετώπισης.

Από τα παραπάνω μπορεί να συμπεράνει κανείς πως είναι θεμελιωμένη στην επιστημονική κοινότητα η άποψη πως η αξιοποίηση του πλαισίου μπορεί να διαδραματίσει καθοριστικό ρόλο στην αντιμετώπιση των προκλήσεων στη διαχείριση προσωπικής πληροφορίας που προκύπτουν από τη διάδοση των συσκευών κινητού υπολογισμού.

1.4 Πλαίσιο εκπόνησης διατριβής

1.4.1 Κίνητρο - στόχος

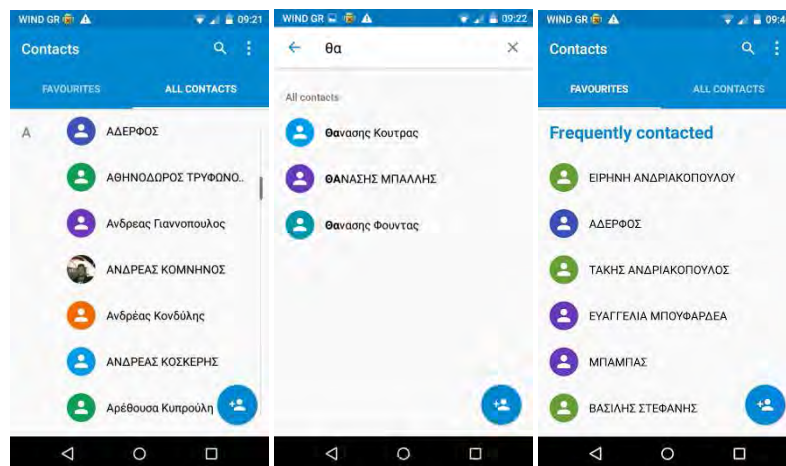
Κίνητρο για την εκπόνηση της έρευνας που παρουσιάζεται στην παρούσα διδακτορική διατριβή αποτέλεσαν οι προκλήσεις και τα προβλήματα που προέκυψαν στον τομέα της διαχείρισης προσωπικής πληροφορίας λόγω της ευρείας χρήσης συσκευών κινητού υπολογισμού, αλλά και οι δυνατότητες που προσφέρει η αξιοποίηση του κινητού κοινωνικού πλαισίου στην αντιμετώπιση τους, όπως παρουσιάστηκε αναλυτικά στην προηγούμενη ενότητα.

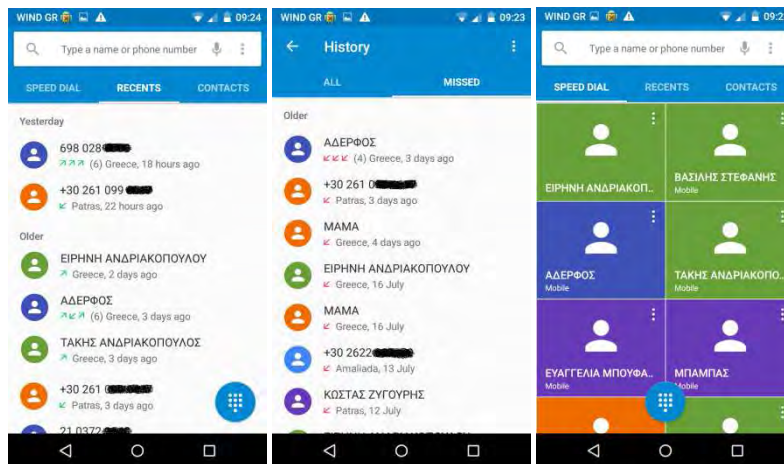
Στόχος της διδακτορικής έρευνας ήταν η δημιουργία ενός μοντέλου που αξιοποιεί την πληροφορία πλαισίου που συλλέγουν οι συσκευές κινητού υπολογισμού για να βελτιώσει τη διαχείριση της προσωπικής πληροφορίας από τους χρήστες. Το στάδιο εκείνο της διαχείρισης προσωπικής πληροφορίας στο οποίο δόθηκε έμφαση είναι η **αξιοποίηση**. Οι Myers, Zapata και Singh (2007) έχουν τονίσει την ανάγκη γρήγορης πρόσβασης στη σωστή πληροφορία τη σωστή στιγμή για τους χρήστες συσκευών κινητού υπολογισμού και έχουν επισημάνει πόσο σημαντικό είναι τα εργαλεία διαχείρισης προσωπικής πληροφορίας να τους βοηθούν να πραγματοποιούν τις εργασίες τους αποτελεσματικά. Από την άλλη, τα υπάρχοντα εργαλεία διαχείρισης προσωπικής πληροφορίας απαιτούν εκτενή συμμετοχή των χρηστών, καθιστώντας τη διαχείριση των στοιχείων προσωπικής πληροφορίας, όπως οι επαφές ή οι σημειώσεις μια χρονοβόρα διαδικασία (Zhou, Mohammed & Zhang, 2012). Όπως ήδη έχει αναφερθεί, η ποσότητα προσωπικής πληροφορίας που αποθηκεύουν οι χρήστες έχει αυξηθεί σε πολύ μεγάλο βαθμό, ενώ καθώς ο αποθηκευτικός χώρος γίνεται ολοένα και πιο φθηνός, στο προσεχές μέλλον αναμένεται να μη διαγράφουν καθόλου πληροφορία, μια διαδικασία που απαιτεί κόπο και χρόνο. Παράλληλα, τα χαρακτηριστικά των συσκευών κινητού υπολογισμού θέτουν σημαντικά προβλήματα στην εύρεση και αξιοποίηση της προσωπικής πληροφορίας. Επομένως, νέα μοντέλα διαχείρισης πληροφορίας, ενήμερα για το πλαίσιο, απαιτούνται ώστε οι χρήστες να έχουν άμεσα διαθέσιμη την απαιτούμενη πληροφορία την κατάλληλη στιγμή για να ικανοποιήσουν τις πληροφοριακές τους ανάγκες.

Σε αυτή την έρευνα μελετήθηκε η **περίπτωση των επαφών ως πρόβλημα τομέα** για λόγους που παρουσιάζονται στην επόμενη ενότητα. Σκοπός ήταν να διερευνηθεί πως το **πλαίσιο** μπορεί να βελτιώσει την **ανάκτηση των επαφών για την πραγματοποίηση επικοινωνίας** και συγκεκριμένα **εξερχόμενων κλήσεων**. Εντούτοις, η προσέγγιση που παρουσιάζεται μπορεί να επεκταθεί και να εφαρμοστεί σε οποιοδήποτε άλλο στοιχείο προσωπικής πληροφορίας.

1.4.2 Οι επαφές ως πρόβλημα τομέα

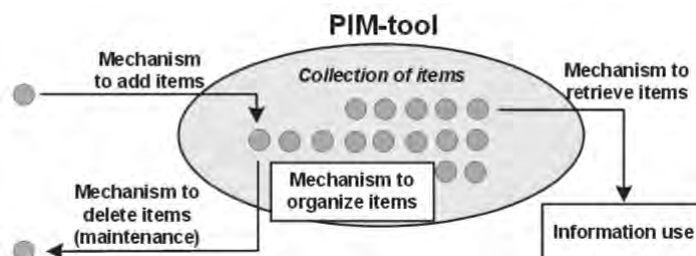
Δεδομένου ότι η κύρια αποστολή των συσκευών κινητού υπολογισμού, στην πιο διαδεδομένη μορφή τους (smart phones), είναι η διευκόλυνση της επικοινωνίας των χρηστών (Bentley & Chen, 2015), γίνεται σαφές πόσο σημαντική είναι η ευχρηστία των παρεχόμενων διαδικασιών για την έναρξη μιας συνόδου επικοινωνίας. Ενδεικτικό του ρόλου που έχουν τα κινητά τηλέφωνα ως εργαλεία επικοινωνίας είναι πως εφαρμογές όπως η λίστα επαφών ή η εφαρμογή σύντομων μηνυμάτων (SMS) είναι οι πιο συχνά χρησιμοποιούμενες από τους χρήστες τους (LaRue, Mitchell, Terhorst & Karimi, 2010), ενώ γενικά η χρήση εφαρμογών επικοινωνίας κυριαρχεί κατά τη διάρκεια της ημέρας (Böhmer, Hecht, Schöning, Krüger & Bauer, 2011). Σε μια κινητή συσκευή, μια τυπική διαδικασία επικοινωνίας (π.χ. η πραγματοποίηση μιας εξερχόμενης κλήσης ή η αποστολή ενός σύντομου μηνύματος ή ενός email) περιλαμβάνει συνήθως την ανάκτηση της επιθυμητής επαφής (δηλαδή την αναζήτηση και εύρεση του ονόματος και του αριθμού ή της ηλεκτρονικής διεύθυνσης επικοινωνίας), ως θεμελιώδες μέρος της επιτυχούς ολοκλήρωσης της. Ως επαφή (contact) μπορεί να οριστεί μια εγγραφή που αποθηκεύεται στη μνήμη της συσκευής κινητού υπολογισμού και περιλαμβάνει ένα σύνολο από πεδία (όνομα, επώνυμο, αριθμούς τηλεφώνων, email, διεύθυνση κ.α.) που αντιστοιχούν σε ένα πρόσωπο και κυρίως σχετίζονται με την επικοινωνία μαζί του. Οι επαφές είναι από τα πιο διαδεδομένα στοιχεία προσωπικής πληροφορίας και δεν αφορούν μόνο τις κινητές συσκευές, αφού εμφανίζονται προγενέστερα στους σταθερούς υπολογιστές (π.χ. στα προγράμματα διαχείρισης ηλεκτρονικής αλληλογραφίας).





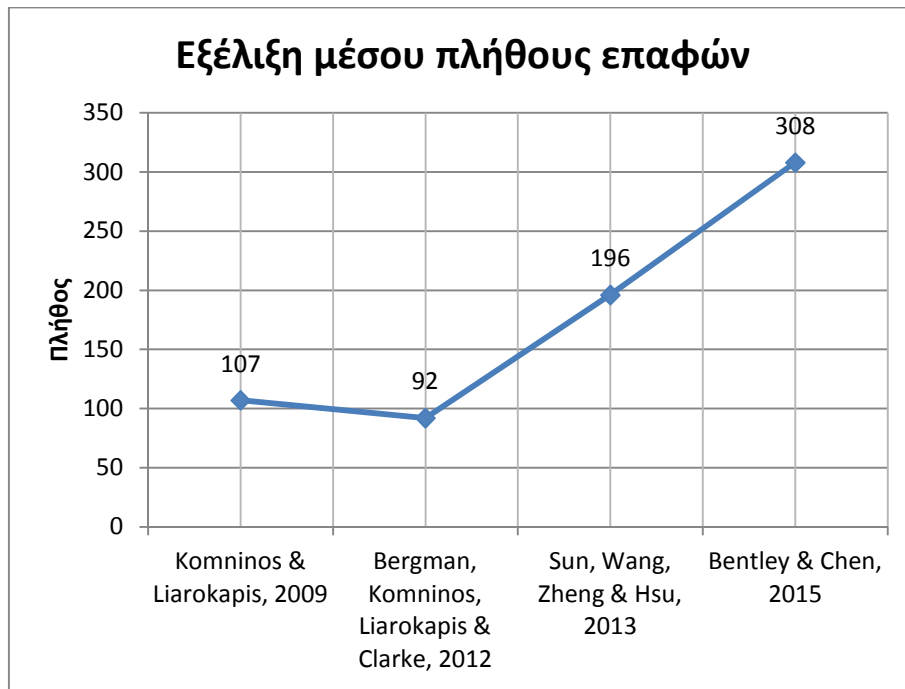
Εικόνα 2. Από αριστερά προς τα δεξιά και από πάνω προς τα κάτω: αποθετήριο επαφών, αναζήτηση στο αποθετήριο, λίστα πιο συχνά χρησιμοποιούμενων επαφών, λίστα πρόσφατων κλήσεων, λίστα χαμένων κλήσεων, λίστα γρήγορης επανάκλησης (speed dial)

Η ανάκτηση μιας επαφής μπορεί να γίνει με αρκετούς διαφορετικούς τρόπους (όπως η αναζήτηση με ένα αλφαριθμητικό που αποτελεί μέρος του ονόματος της επαφής ή το φιλτράρισμα στις ομάδες επαφών) (Lee, Seo & Lee, 2010) και από διάφορες διαθέσιμες διεπαφές (Bergman, Komninos, Liarakis & Clarke, 2012): λίστα επαφών (contact list), λίστα πρόσφατων κλήσεων (recent calls lists), λίστα αναπάντητων κλήσεων (missed calls list), λίστα αγαπημένων επαφών (favorites), λίστες γρήγορης επανάκλησης (speed dial lists), ειδοποιήσεις στην οθόνη (notifications) (εικόνα 2). Εντούτοις, το προκαθορισμένο αποθετήριο στο οποίο αποθηκεύονται όλες οι επαφές και αποτελεί την προεπιλεγμένη εφαρμογή πρόσβασης σε αυτές είναι η λίστα ή βιβλίο επαφών (contact list ή phonebook). Πρόκειται για ένα εργαλείο διαχείρισης προσωπικής πληροφορίας που παρέχει τις κύριες λειτουργίες που έχουν τα τυπικά αυτά εργαλεία (εικόνα 3) σύμφωνα με τον Boardman (2004): μηχανισμό προσθήκης νέων επαφών, διαγραφής επαφών, οργάνωσης (μέσω δημιουργίας ομάδων επαφών ή χαρακτηρισμού ως αγαπημένων) και ανάκτησης επαφών (μέσω αλφαριθμητικής αναζήτησης ή πλοήγησης στη γραμμική λίστα).



Εικόνα 3. Ένα αφαιρετικό εργαλείο διαχείρισης προσωπικής πληροφορίας (Boardman, 2004)

Ένα από τα πλεονεκτήματα των «ψηφιακών» λιστών επαφών είναι ότι προσφέρουν στους χρήστες τους πολύ μεγάλο αποθηκευτικό χώρο (πρακτικά απεριόριστο στις σύγχρονες συσκευές κινητού υπολογισμού καθώς διαθέτουν εκατοντάδες ή χιλιάδες θέσεις αποθήκευσης) για την αποθήκευση των λεπτομερειών που αφορούν τις επαφές. Συνεπώς, οι χρήστες των συσκευών έχουν τη δυνατότητα να δημιουργήσουν πολυπληθή αποθετήρια επαφών, τα οποία συνήθως περιλαμβάνουν και πολλές επαφές που είναι χρήσιμες μόνο προσωρινά ή χρησιμοποιούνται σπάνια, καθώς και άσχετα στοιχεία, όπως PINs, κωδικούς πρόσβασης (passwords) και άλλες πληροφορίες (Komninos & Liarokapis, 2009; Jung, Anttila & Blom, 2008). Επιπλέον, τα αποθετήρια των επαφών τείνουν να γίνουν ακόμα μεγαλύτερα τα τελευταία χρόνια, καθώς οι σύγχρονες κινητές συσκευές επιτρέπουν την εισαγωγή επαφών από υπηρεσίες επικοινωνίας (π.χ. Email providers, 3rd party εφαρμογές ανταλλαγής μηνυμάτων) και από τις υπηρεσίες κοινωνικής δικτύωσης στις οποίες συμμετέχει ο χρήστης (Bentley & Chen, 2015). Μάλιστα, το μέγεθος του δικτύου επαφών από τα κοινωνικά δίκτυα στα οποία συμμετέχουν οι χρήστες είναι συνήθως μεγαλύτερο από το μέγεθος του δικτύου επαφών που χρησιμοποιούν για τις τυπικές λειτουργικότητες των κινητών συσκευών, όπως η πραγματοποίηση κλήσεων και η αποστολή μηνυμάτων (Salehan & Negahban, 2013). Έτσι, εκτός από τη λίστα των επαφών που καλεί ο χρήστης, μπορεί κανείς να βρει στο αποθετήριο επαφές από το Gmail, το Viber και το WhatsApp, το Facebook και το Twitter, το Skype κ.λπ. Απόδειξη για τη συνεχή αύξηση του μεγέθους του αποθετηρίου επαφών είναι πως σε μια πρόσφατη μελέτη (Bentley & Chen, 2015) ο μέσος αριθμός επαφών στο βιβλίο επαφών ήταν 308 επαφές. Αυτό το νούμερο είναι σημαντικά μεγαλύτερο σε σχέση με αυτά που αναφέρονται σε προηγούμενες μελέτες (π.χ. 107 επαφές στο (Komninos & Liarokapis, 2009), 92 επαφές στο (Bergman, Komninos, Liarokapis & Clarke, 2012) και 196 επαφές στο σύνολο δεδομένων Rice LiveLab (Shepard, Rahmati, Tossell, Zhong & Kortum, 2011) που χρησιμοποιήθηκε στο (Sun, Wang, Zheng & Hsu, 2013) (γράφημα 1). Από την άλλη, οι χρήστες είναι συνήθως απρόθυμοι να διαγράψουν επαφές που δε χρειάζονται, λόγω του κόστους αυτής της διαδικασίας σε χρόνο και προσπάθεια (Bergman, Komninos, Liarokapis & Clarke, 2012).



Γράφημα 1. Εξέλιξη του μέσου πλήθους επαφών στο βιβλίο επαφών σε διαφορετικές ερευνητικές εργασίες

Τα μοναδικά χαρακτηριστικά των συσκευών κινητού υπολογισμού σε σχέση με τους σταθερούς υπολογιστές που παρουσιάστηκαν και νωρίτερα, όπως οι μικρές οθόνες και οι περιορισμοί στην ανάλυση, τα δύσχρηστα πληκτρολόγια και οι εναλλακτικοί τρόποι αλληλεπίδρασης, η ταυτόχρονη εμπλοκή του χρήστη σε άλλες δραστηριότητες (π.χ. περπάτημα), η έλλειψη σταθερού σημείου χρήσης καθώς οι χρήστες τις κρατούν στο χέρι τους κ.α., θέτουν σημαντικές δυσκολίες στην πρόσβαση και την ανάκτηση της πληροφορίας (Zhou, Mohammed & Zhang, 2012) για την περαιτέρω αξιοποίησή της. Η ύπαρξη πολυπληθών αποθετηρίων επαφών επιδεινώνει το πρόβλημα της ανάκτησης (Bergman, Komninos, Liarokapis & Clarke, 2012; Bentley & Chen, 2015), υπό την έννοια ότι η εύρεση της επιθυμητής επαφής για την έναρξη μιας επικοινωνίας συνήθως απαιτεί σχετικά εκτενή αλληλεπίδραση και συχνά οδηγεί σε εκνευρισμό εξ' αιτίας σφαλμάτων (π.χ. κλήση λάθος προσώπου και επανέναρξη της διαδικασίας).

1.4.3 Ερευνητικά ερωτήματα και αντιμετώπισή τους

Τα ερευνητικά ερωτήματα που τέθηκαν προς διερεύνηση είναι τα παρακάτω:

- **E1: Ποιο μέρος του πλαισίου των συσκευών κινητού υπολογισμού σχετίζεται με την ανάκτηση επαφών από τους χρήστες για την έναρξη επικοινωνίας;** Καθώς δεν είναι όλα τα στοιχεία πλαισίου που συγκεντρώνει μια συσκευή κινητού υπολογισμού χρήσιμα για τις ενήμερες για το πλαίσιο

εφαρμογές (Korpiä, Häkkinen, Kela, Ronkainen & Känslä, 2004), η διερεύνηση αυτού του ερωτήματος θα πρέπει να αποκαλύψει εκείνο το μέρος του πλαισίου που μπορεί να αξιοποιηθεί για τη βελτίωση της πρόσβασης στις επαφές. Για την απάντηση αυτού του ερωτήματος πραγματοποιήθηκε ένα πείραμα ανάλυσης των αποθετηρίων των επαφών και των προτύπων επικοινωνίας των χρηστών συσκευών κινητού υπολογισμού (κεφάλαιο 3).

- **E2: Πώς μπορεί το πλαίσιο αυτό να μοντελοποιηθεί και να αξιοποιηθεί για τη βελτίωση της πρόσβασης στις επαφές;** Στα πλαίσια αυτού του ερωτήματος χρειάζεται να προταθεί ένα κατάλληλο μοντέλο αναπαράστασης του πλαισίου και ένας αλγόριθμος που να το χρησιμοποιεί ώστε να βελτιώσει την αξιοποίηση των επαφών. Για τον σκοπό αυτό, αναπτύχθηκε ένα μοντέλο αναπαράστασης του πλαισίου (κεφάλαιο 3) και με βάση αυτό υλοποιήθηκε αλγόριθμος πρόβλεψης εξερχόμενων κλήσεων (κεφάλαιο 4) που μπορεί να χρησιμοποιηθεί σε προσαρμοζόμενες διεπαφές για τη βελτίωση της ανάκτησης των επαφών.
- **E3: Ποια τα αποτελέσματα αξιολόγησης ενός συστήματος που αξιοποιεί το πλαίσιο για να διευκολύνει την πρόσβαση στις επαφές;** Ένα ενήμερο για το πλαίσιο σύστημα είναι σημαντικό να αξιολογηθεί στο πεδίο, καθώς η αξιολόγηση παρόμοιων προσπαθειών σε πραγματικές συνθήκες αποτελεί μια έλλειψη στη σχετική ερευνητική βιβλιογραφία. Για την απάντηση αυτού του ερωτήματος πραγματοποιήθηκαν τόσο πειραματικές προσομοιώσεις, όσο και ένα πείραμα πεδίου (κεφάλαιο 4).
- **E4: Ποια είναι η επίπτωση της τοποθεσίας ως παράγοντα πλαισίου για την πρόσβαση στις επαφές;** Καθώς, όπως παρουσιάζεται στη συνέχεια, η τοποθεσία είναι ένας από τους σημαντικότερους και συχνότερα χρησιμοποιούμενους παράγοντες πλαισίου, είναι απαραίτητο να μελετηθεί η επίπτωσή της στο πρόβλημα της αξιοποίησης των επαφών, ώστε να καλυφθεί το κενό που υπάρχει σε αυτόν τον τομέα της ερευνητικής βιβλιογραφίας. Για τον σκοπό αυτό πραγματοποιήθηκε μια σειρά πειραμάτων (κεφάλαιο 5) που αναλύουν την επίπτωση τόσο της φυσικής, όσο και της σημασιολογικής τοποθεσίας στην πρόβλεψη κλήσεων.

2. Επισκόπηση Σχετικής Ερευνητικής Δραστηριότητας

2.1 Διαχείριση προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού

Η ευρεία εισαγωγή του κινητού υπολογισμού, ειδικά στις ανεπτυγμένες χώρες, ως μέσο διαχείρισης της προσωπικής, κοινωνικής και επαγγελματικής ζωής των χρηστών, καθώς και τα ιδιαίτερα χαρακτηριστικά της χρήσης συσκευών κινητού υπολογισμού σε σχέση με τους παραδοσιακούς σταθερούς υπολογιστές, έχει οδηγήσει σε σημαντική ερευνητική δραστηριότητα στον τομέα της διαχείρισης προσωπικής πληροφορίας. Στη συνέχεια, παρουσιάζονται μελέτες σχετικά με τις πληροφοριακές ανάγκες των χρηστών συσκευών κινητού υπολογισμού και πως διαφοροποιούνται σε σχέση με τη χρήση σταθερών υπολογιστών. Επίσης, παρατίθενται εργασίες και μελέτες που παρουσιάζουν προσπάθειες για εργαλεία διαχείρισης προσωπικής πληροφορίας που ανταποκρίνονται στις προκλήσεις των συσκευών κινητού υπολογισμού.

2.1.1 Ανάγκες των κινητών χρηστών για πληροφορία

Η μελέτη των πληροφοριακών αναγκών των κινητών χρηστών είναι ένα απαραίτητο βήμα για την ανάπτυξη νέων προσεγγίσεων για τη διαχείριση προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού.

Οι Komninos, Baillie και Barrie (2008) υποστήριξαν ότι οι μέθοδοι διαχείρισης προσωπικής πληροφορίας που έχουν μεταφερθεί από τις παραδοσιακές υπολογιστικές πρακτικές, αδυνατούν να ικανοποιήσουν τις απαιτήσεις των κινητών χρηστών και πρότειναν μια ολιστική προσέγγιση για τη διαχείριση της προσωπικής πληροφορίας που βασίζεται στην αλληλεπίδραση των συσκευών με το περιβάλλον τους. Καθώς οι χρήστες έχουν συνεχώς τις συσκευές μαζί τους και αυτές ενσωματώνουν ολοένα και περισσότερους αισθητήρες, οι ερευνητές διέκριναν νέους τύπους προσωπικής πληροφορίας που ονόμασαν εγω-δεδομένα (me-Data): δεδομένα τοποθεσίας, κοινωνικά δεδομένα, δεδομένα που αφορούν άλλους χρήστες και δεδομένα που σχετίζονται με το περιβάλλον του χρήστη και τις υπηρεσίες που χρησιμοποιεί. Εξετάζοντας τις ανάγκες προσωπικής πληροφορίας των χρηστών, παρατήρησαν ότι οι χρήστες πρέπει να μπορούν να επιτελούν τις εργασίες που επιθυμούν σε οποιαδήποτε τοποθεσία και οποιαδήποτε χρονική στιγμή. Η κινητικότητα εισάγει τη έννοια των μεταβαλλόμενων περιβαλλόντων και καταστάσεων που έχουν ως αποτέλεσμα να ποικίλουν οι απαιτήσεις των χρηστών για δεδομένα και ο τρόπος με τον οποίο τα δεδομένα δημιουργούνται ή παρουσιάζονται:

- *Δημιουργία δεδομένων*: οι χρήστες δεν έχουν πάντα τη δυνατότητα να αφιερώσουν τον ίδιο χρόνο για τη δημιουργία δεδομένων, όπως για παράδειγμα στην περίπτωση δημιουργίας μιας σημείωσης ενώ βρίσκονται στην αποβάθρα του μετρό και αυτό καταφθάνει πριν την ολοκλήρωση της εργασίας. Τέτοιες καταστάσεις αποτελούν ενδείξεις για την ανάγκη αυξημένης υποστήριξης της εισόδου δεδομένων στις συσκευές: από τη βελτίωση της ταχύτητας εισόδου (π.χ. T9 ή Swype) έως την υλοποίηση νέων τρόπων εισόδου (π.χ. φωνητική είσοδος ή χειρονομίες) και τη βελτίωση της σχεδίασης φορμών εισόδου.
- *Αναμετάδοση πληροφορίας*: καθώς οι χρήστες αλληλεπιδρούν με τις κινητές συσκευές με διαφορετικό τρόπο σε σχέση με τους σταθερούς υπολογιστές, μια σημαντική απαίτηση είναι να μην είναι απαραίτητο να αφιερώσουν πολύ χρόνο για να βρουν την πληροφορία που χρειάζονται. Μάλιστα, οι χρήστες εκνευρίζονται πολύ πιο γρήγορα (χρόνος εργασίας > 1 λεπτό) κατά την εργασία αναζήτησης σε συσκευές κινητού υπολογισμού απ' ότι στους σταθερούς υπολογιστές (χρόνος εργασίας > 2,5 λεπτά). Είναι, επομένως, πολύ σημαντικό να παρουσιάζονται στους χρήστες τα δεδομένα που σχετίζονται με την κατάσταση στην οποία βρίσκονται, καθώς και να φιλτράρονται και να παρουσιάζονται σε τέτοιο χρόνο που να τους επιτρέπουν να πραγματοποιήσουν τις ενέργειες που επιθυμούν.
- *Διακοπή εργασιών*: Ένα σημαντικό ερώτημα που προκύπτει είναι πότε είναι κατάλληλη στιγμή για τη διακοπή του χρήστη από την τρέχουσα εργασία και την παρουσίαση πληροφορίας (π.χ. στην περίπτωση υπενθυμίσεων). Η έξυπνη αξιοποίηση του πλαισίου του χρήστη και ειδικότερα η τρέχουσα δραστηριότητά του και οι κοινωνικές προεκτάσεις που προκύπτουν από την κατάσταση στην οποία βρίσκεται είναι κρίσιμοι παράγοντες για την πραγματοποίηση ή όχι μιας διακοπής (π.χ. μια ηχητική υπενθύμιση κατά τη διάρκεια μιας επαγγελματικής συνάντησης είναι συνήθως ακατάλληλη).

Οι Church και Smyth (2009) αναγνώρισαν επίσης τις προκλήσεις που προκύπτουν για τις ανάγκες πληροφόρησης των χρηστών όταν αυτοί κινούνται και ειδικότερα την επίπτωση που μπορεί να έχει το πλαίσιο στο οποίο βρίσκονται (τοποθεσία, χρόνος, δραστηριότητα, κοινωνικές αλληλεπιδράσεις κ.λπ.) και οι αλλαγές του. Πραγματοποίησαν μια μελέτη ημερολογίου για την έρευνα της πρόθεσης πίσω από τις ανάγκες πληροφόρησης των χρηστών όταν αυτοί βρίσκονταν στο σπίτι, το γραφείο ή σε κίνηση. Τα αποτελέσματα της μελέτης κατέδειξαν τις σημαντικές διαφορές των αναγκών για τους κινητούς χρήστες. Ένα πολύ σημαντικό εύρημα ήταν

πως όταν οι χρήστες βρίσκονταν σε κίνηση υπήρχε ξεκάθαρη χωρική και χρονική εξάρτηση των πληροφοριακών τους αναγκών. Επιπλέον, οι ανάγκες των χρηστών για πληροφόρηση ήταν αυξημένες όταν κινούνταν και βρίσκονταν μακριά από οικεία περιβάλλοντα, όπως είναι το σπίτι ή το γραφείο. Η ανάλυση της πρόθεσης πίσω από τις ανάγκες πληροφόρησης έδειξε ότι ένα σημαντικό ποσοστό αναγκών (σχεδόν 11%) αφορούσε τη διαχείριση προσωπικής πληροφορίας, από τις οποίες το 65% προέκυψε όταν οι χρήστες ήταν σε κίνηση. Αυτές οι ανάγκες εστιάζουν στο στόχο εύρεσης προσωπικών δεδομένων που αφορούν ατομικά τους χρήστες και συνήθως πρόκειται για ερωτήματα που μπορούν να απαντηθούν μόνο από τους ίδιους ή που για να απαντηθούν απαιτείται είσοδος από τους ίδιους. Περιλαμβάνουν αντικείμενα προσωπικής πληροφορίας όπως επαφές, σημειώσεις, εργασίες ημερολογίου, μηνύματα ηλεκτρονικού ταχυδρομείου κ.α. Εντούτοις, πολλές από αυτές τις ανάγκες δεν μπορούν να καλυφθούν επιτυχώς από τις σημερινές κινητές συσκευές, ένα γεγονός που τονίζει την αναγκαιότητα για νέες υπηρεσίες διαχείρισης προσωπικής πληροφορίας.

Σε μια παρόμοια εργασία (Sohn, Li, Griswold & Hollan, 2008), οι ερευνητές επισήμαναν την ανάγκη για εφαρμογές και υπηρεσίες που είναι σχεδιασμένες με βάση τις απαιτήσεις του κινητού πλαισίου χρήσης. Και σε αυτή την έρευνα πραγματοποιήθηκε μελέτη ημερολογίου, όπου οι χρήστες κατέγραφαν τις πληροφοριακές τους ανάγκες όταν βρίσκονταν σε κίνηση. Αν και η κατηγοριοποίηση των αναγκών ήταν διαφορετική σε σύγκριση με αυτή που παρουσιάστηκε στο (Church & Smyth, 2009), το άθροισμα των ποσοστών για τις κατηγορίες που αφορούν προσωπικά στοιχεία, προγραμματισμό εργασιών, αριθμούς επαφών και μηνύματα ηλεκτρονικού ταχυδρομείου, δηλαδή για τη διαχείριση προσωπικής πληροφορίας, πλησιάζει το 20% των συνολικών πληροφοριακών αναγκών. Σύμφωνα με τα αποτελέσματα, περισσότερες από τις μισές (55%) από τις ανάγκες πληροφόρησης των κινητών χρηστών δεν ικανοποιήθηκαν τη στιγμή που προέκυψαν ή και ποτέ, ενώ η τρέχουσα εργασία του χρήστη φάνηκε να παίζει σημαντικό ρόλο σε αυτό. Επιπλέον, το 72% των αναγκών σχετιζόνταν με το πλαίσιο των χρηστών, δηλαδή την τοποθεσία, το χρόνο, τη συνομιλία και τη δραστηριότητα. Τέλος, η ικανοποίηση του 38% των απαιτήσεων των αναγκών πληροφόρησης των χρηστών περιελάμβανε πρόσβαση σε προσωπικά τους δεδομένα που ήταν διαθέσιμα είτε μέσα από web υπηρεσίες είτε απευθείας από τις συσκευές τους. Αυτό το εύρημα οδήγησε τους ερευνητές να τονίσουν την ανάγκη εξερεύνησης βελτιωμένων τρόπων πρόσβασης στις διαθέσιμες πηγές προσωπικών δεδομένων των συσκευών.

Ο Heimonen (2009), μελέτησε επίσης τις ανάγκες πληροφόρησης των κινητών χρηστών. Σύμφωνα με τα ευρήματά του, οι χρήστες χρησιμοποιούσαν κατά κύριο λόγο την πρόσβαση στον παγκόσμιο ιστό και εργαλεία διαχείρισης προσωπικής πληροφορίας για να τις ικανοποιήσουν. Επιπλέον, μελέτησε το πλαίσιο στο οποίο προκύπτουν οι πληροφοριακές ανάγκες και βρήκε ότι τα 2/3 των αναγκών προέκυψαν μακριά από ένα σταθερό πλαίσιο (όπως το σπίτι), ενώ περίπου το 40% από αυτές δημιουργήθηκαν όταν οι χρήστες ήταν κυριολεκτικά σε κίνηση. Ακόμα και στην περίπτωση που οι χρήστες είχαν διαθέσιμο σταθερό υπολογιστή, συχνά επέλεγαν να χρησιμοποιήσουν την κινητή τους συσκευή για να ικανοποιήσουν τις ανάγκες τους. Τέλος, ένα σημαντικό μέρος των αναγκών για πληροφορία φάνηκε να δημιουργείται σε σχέση με το κοινωνικό πλαίσιο των χρηστών.

2.1.2 Εργαλεία διαχείρισης προσωπικής πληροφορίας και συσκευές κινητού υπολογισμού

Οι Dai, Lutters και Bower (2005) μελέτησαν την πρακτική της λήψης άτυπων σημειώσεων (informal note taking) σε κινητές συσκευές, που είναι μια ουσιώδης δραστηριότητα διαχείρισης προσωπικής πληροφορίας. Οι περισσότερες συσκευές κινητού υπολογισμού την υποστηρίζουν υιοθετώντας τόσο δομημένες (ημερολόγιο, λίστα εργασιών, επαφές) όσο και μη δομημένες μορφές δεδομένων (σημειώματα). Η έρευνά τους έδειξε ότι τα δομημένα εργαλεία αποφεύγονταν για χάρη των μη δομημένων, αν και αυτό καθιστούσε τη διαχείριση της πληροφορίας πιο δύσκολη. Οι ερευνητές κατέληξαν πως εργαλεία που συνδυάζουν τα χαρακτηριστικά των δύο κατηγοριών είναι απαραίτητα, δίνοντας κυρίως έμφαση στη δυνατότητα διασύνδεσης των δεδομένων ανάμεσα στα διαφορετικά εργαλεία. Όπως παρουσιάστηκε και στην εισαγωγή, οι εφαρμογές διαχείρισης προσωπικής πληροφορίας λειτουργούν ως «περιφραγμένοι κήποι» μη επιτρέποντας την ανταλλαγή πληροφορίας μεταξύ τους. Επιπλέον, οι Chande, Vartiainen και Rämö (2007) αφού μελέτησαν τις ανάγκες των κινητών χρηστών για τη λήψη σημειώσεων δημιούργησαν ένα ερευνητικό πρωτότυπο που επέτρεπε τη δημιουργία σημειώσεων, και τη συσχέτισή τους με δεδομένα από άλλα, κυρίως δομημένα αποθετήρια προσωπικής πληροφορίας (ημερολόγιο, επαφές, φωτογραφίες, αρχεία, εφαρμογή χάρτη, φυλλομετρητή ιστού). Γεγονότα που σχετίζονται με το πλαίσιο των χρηστών (π.χ. μια κλήση με κάποια επαφή ή μια υπενθύμιση στο ημερολόγιο) είχαν ως αποτέλεσμα την αυτόματη παρουσίαση συσχετισμένων σημειώσεων. Ακόμα μια ερευνητική εργασία που αφορά τη λήψη σημειώσεων σε κινητές συσκευές παρουσιάστηκε στο (Bernstein, Van Kleeck, Schraefel & Karger, 2008), όπου οι ερευνητές δημιούργησαν μια εφαρμογή για κινητές συσκευές που επέτρεπε τη λήψη σημειώσεων σε συνδυασμό με την επισύναψη

εικόνας και ήχου, ενώ ταυτόχρονα συγχρόνιζε τα δεδομένα ανάμεσα στην κινητή εφαρμογή και την εφαρμογή που είχε σχεδιαστεί ειδικά για φορητούς υπολογιστές. Η σχεδίαση ακόμα ενός πειραματικού πρωτοτύπου για τη διαχείριση σημειώσεων σε συσκευές κινητού υπολογισμού παρουσιάστηκε στο (Falke, 2008).

Οι Leino, Finnberg και Rähä (2010) μελέτησαν τη χρήση εργαλείων διαχείρισης λιστών εργασιών (to-do lists) και βρήκαν ότι τα ηλεκτρονικά εργαλεία είναι πλέον εξίσου δημοφιλή με τη χρήση χαρτιού, ενώ περίπου το 60% των συμμετεχόντων χρησιμοποιούσαν συσκευές κινητού υπολογισμού για το σκοπό αυτό. Επίσης, μελέτησαν χρήστες που χρησιμοποιούσαν τις κινητές συσκευές τους για τη διαχείριση προσωπικής πληροφορίας και ανέφεραν ότι οι εφαρμογές ημερολογίου χρησιμοποιούνταν περισσότερο από τα έντυπα ημερολόγια, ενώ και η χρήση εργαλείων για διαχείριση σημειώσεων και εργασιών αυξανόταν με τις κινητές συσκευές. Εντούτοις, παρατήρησαν ότι η αύξηση της δημοφιλίας των κινητών εργαλείων διαχείρισης προσωπικής πληροφορίας οφείλεται στη βελτίωση των τεχνολογικών χαρακτηριστικών των συσκευών (μεγαλύτερες οθόνες, βελτιωμένες μέθοδοι εισόδου, συνδεσιμότητα) και όχι σε πρόοδο που αφορούσε το σχεδιασμό των εφαρμογών.

Οι Komninos, Baillie και Dunlop (2007) μελέτησαν τρόπους βελτίωσης του ημερολογίου με το οποίο είναι εξοπλισμένες οι συσκευές κινητού υπολογισμού και βρήκαν ότι η σημασιολογική κατηγοριοποίηση των εγγραφών με τη χρήση χρώματος βελτιώνει την ανάκτηση τους. Πρότειναν, επίσης, την επαύξηση και οργάνωσή τους με βάση γεωχωρική πληροφορία, όπως η απόσταση από την τρέχουσα θέση του χρήστη ή ο χρόνος που απαιτείται για τη μετάβαση στην τοποθεσία όπου είναι προγραμματισμένο το γεγονός. Άλλη μια εργασία που αφορά τη μελέτη εναλλακτικής αναπαράστασης γεγονότων ημερολογίου στις περιορισμένες σε μέγεθος οθόνες των κινητών συσκευών (ειδικότερα σε PDAs) είναι η (Bederson, Clamage, Czerwinski & Robertson, 2004).

Όπως έχει ήδη αναφερθεί, οι σύγχρονες συσκευές κινητού υπολογισμού είναι εξοπλισμένες με κάμερες υψηλής ανάλυσης που επιτρέπουν τη δημιουργία και αποθήκευση μεγάλου πλήθους από πολυμεσικά δεδομένα (φωτογραφίες και βίντεο), δημιουργώντας ταυτόχρονα δυσκολία στη διαχείριση και την εύρεσή τους. Μια λύση για το πρόβλημα αυτό είναι η αυτοματοποιημένη εξαγωγή μεταδεδομένων για τον χαρακτηρισμό των πολυμεσικών αρχείων. Για παράδειγμα, η εφαρμογή Photo Cube (Kim, Lee, Won & Moon, 2011) χρησιμοποιεί τους αισθητήρες των συσκευών ή τις δυνατότητες δικτύωσης για να εξάγει αυτόματα μεταδεδομένα για τις φωτογραφίες

(ημερομηνία, ώρα, τοποθεσία, καιρός, προσωπικό γεγονός κ.λπ.). Τα μεταδεδομένα μπορούν να χρησιμοποιηθούν για την οργάνωση των φωτογραφιών και για την αναζήτησή τους. Σε μια παρόμοια προσέγγιση, παρουσιάστηκε η εφαρμογή TagSense (Qin, Bao, Roy Choudhury & Nelakuditi, 2011) που επιτρέπει την αυτοματοποιημένη προσθήκη ετικετών σε φωτογραφίες αναλύοντας τόσο στοιχεία από το πλαίσιο (π.χ. ώρα, τοποθεσία) της συσκευής που λαμβάνει τη φωτογραφία, αλλά και των συσκευών που βρίσκονται κοντά (π.χ. δραστηριότητα χρηστών), όσο και από το περιεχόμενο τους (π.χ. αναγνώριση προσώπου ή αντικειμένου). Οι ετικέτες μπορούν να χρησιμοποιηθούν για τη βελτίωση της αναζήτησης. Σε μια άλλη σχετική εργασία (Anguera, Xu & Oliver, 2008), προτάθηκε εκτός από τα μεταδεδομένα που παράγονται αυτόματα (ημερομηνία και ώρα, αναγνώριση χρήστη, τοποθεσία), ο χαρακτηρισμός φωτογραφιών με τη χρήση φωνητικής εισόδου, οπότε και γίνεται εφικτή η πολυτροπική ανάκτησή τους. Στη βιβλιογραφία υπάρχουν και άλλες αντίστοιχες ερευνητικές προσπάθειες δημιουργίας μεταδεδομένων που διευκολύνουν την ανάκτηση τόσο για την περίπτωση των φωτογραφιών (Sarvas, Herrarte, Wilhelm, & Davis, 2004; Gurrin, Jones, Lee, O'Hare, Smeaton, & Murphy, 2005) όσο και για τα βίντεο (Patel & Abowd, 2004), τους ήχους ή γενικά τα πολυμέσα (Davis, Van House, Towle, King, Ahern, Burgener, ... & Rothenberg, 2005; Viana, Miron, Moisuc, Gensel, Villanova-Oliver & Martin, 2011) που παράγονται στις συσκευές κινητού υπολογισμού.

Οι Myers, Zapata και Singh (2007) πρότειναν μια νέα διεπαφή για την αντιμετώπιση των προβλημάτων διαχείρισης προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού. Η διεπαφή προσέφερε ένα ενοποιημένο περιβάλλον πρόσβασης στα διαθέσιμα αποθετήρια προσωπικής πληροφορίας. Επιπλέον, επέτρεπε στους χρήστες να ομαδοποιούν στοιχεία πληροφορίας διαφορετικού τύπου, αλλά και να τα διασυνδέουν μεταξύ τους (π.χ. μια επαφή με ένα ραντεβού). Ακόμα ένα πρωτότυπο που προσπάθησε να ενοποιήσει διαφορετικές πηγές προσωπικής πληροφορίας σε μια διεπαφή ήταν το TapGlance (Robbins, 2009).

Μια διαφορετική προσέγγιση στο ζήτημα της διαχείρισης προσωπικής πληροφορίας περιλαμβάνει την προσαρμογή λύσεων για σταθερούς υπολογιστές, ώστε να μπορούν να χρησιμοποιηθούν σε συσκευές κινητού υπολογισμού. Μια τέτοια περίπτωση είναι το σύστημα του σημασιολογικού σταθερού υπολογιστή (Semantic Desktop) (Woerndl & Woehrl, 2008; Woerndl & Hristov, 2009).

2.2 Κινητό πλαίσιο και μοντελοποίηση

Υπάρχουν πάρα πολλά παραδείγματα πρωτοτύπων για συσκευές κινητού υπολογισμού που αφορούν διαφορετικούς τομείς και επιδεικνύουν ότι η ενημερότητα πλαισίου μπορεί να βελτιώσει τις εφαρμογές και τη ζωή των ανθρώπων. Υπάρχουν προσπάθειες που χρησιμοποίησαν το πλαίσιο για να παρέχουν καλύτερες υπηρεσίες επικοινωνίας (Schmidt & Gellersen, 2001; Barkhuus, 2003; Brown & Randell, 2004; Chihani, Bertin, Jeanne & Crespi, 2011). Άλλες επικεντρώθηκαν στη δημιουργία συστημάτων συστάσεων που αξιοποιούσαν το πλαίσιο για την παροχή προτάσεων στους χρήστες. Τέτοιες περιπτώσεις μπορεί να αφορούν, ενδεικτικά, τουριστικούς οδηγούς (π.χ. Park, Hong & Cho 2007; Baltrunas, Ludwig, Peer & Ricci, 2011), προτάσεις αναπαραγωγής μουσικής (π.χ. Lehtiniemi, 2008; Wang, Rosenblum & Wang, 2012) ή παρουσίασης φωτογραφιών (Yoon, Ahn, Lee, Hong & Kim, 2008). Σε άλλες περιπτώσεις το πλαίσιο χρησιμοποιήθηκε για την πρόβλεψη χρήσης κινητών εφαρμογών (π.χ. Shin, Hong & Dey, 2012; Liao, Li, Peng, Yu & Liu, 2013). Στην προηγούμενη ενότητα παρουσιάστηκαν πρωτότυπα που χρησιμοποιούσαν το πλαίσιο για την παραγωγή μεταδεδομένων που χαρακτηρίζουν πολυμεσικά αρχεία και μπορούν να βελτιώσουν την ανάκτησή τους. Σε επόμενη ενότητα παρουσιάζονται βιβλία επαφών που έχουν επαυξηθεί με στοιχεία πλαισίου και προωθούν την επικοινωνία ανάμεσα στους χρήστες.

Η λίστα των ενήμερων για το πλαίσιο εφαρμογών που έχουν προταθεί στη βιβλιογραφία είναι μακροσκελής και η περαιτέρω παρουσίασή τους ξεφεύγει από τους σκοπούς αυτής της διατριβής. Σε όλες τις περιπτώσεις, είναι απαραίτητο να μοντελοποιηθεί με κάποιο τρόπο το πλαίσιο ώστε να μπορεί να χρησιμοποιηθεί από την εφαρμογή. Οι Strang και Linnhoff-Popien (2004) παρουσίασαν διάφορες προσεγγίσεις που μπορούν να εφαρμοστούν στη μοντελοποίηση πλαισίου, ενώ μια πιο πρόσφατη αναλυτική παρουσίαση των μοντέλων περιλαμβάνεται στο (Bettini, Brdiczka, Henricksen, Indulska, Nicklas, Ranganathan & Riboni, 2010):

- **Ζεύγη κλειδιού – τιμής:** τα ζεύγη κλειδιού – τιμής είναι η πιο απλή δομή για τη μοντελοποίηση πληροφοριών πλαισίου. Σε αυτές τις περιπτώσεις το πλαίσιο περιγράφεται ως μια λίστα από απλά χαρακτηριστικά (π.χ. τοποθεσία) με τις τιμές τους. Τα ζεύγη κλειδιού – τιμής μπορούν να παρουσιαστούν και με τη μορφή διανύσματος (vector). Αυτό το μοντέλο είναι εύκολο στη διαχείριση, αλλά λείπουν οι δυνατότητες που προσφέρουν πιο σύνθετες δομές για αποτελεσματικούς αλγορίθμους ανάκτησης.

Μια τέτοια μοντελοποίηση υιοθετήθηκε στο (Du & Wang, 2008), όπου ο χώρος του πλαισίου (context space) αναπαραστάθηκε ως το καρτεσιανό γινόμενο πολλαπλών διαστάσεων πλαισίου. Μια διάσταση πλαισίου αντιστοιχεί σε έναν τύπο πλαισίου από τον οποίο η εφαρμογή εξαρτάται (π.χ. θερμοκρασία). Ένα στιγμιότυπο του πλαισίου είναι ένα σημείο στον πολυδιάστατο χώρο πλαισίου. Ορίζεται ως μια n -πλειάδα ($d_1, d_2, \dots, d_n$) για τον n -διάστατο χώρο πλαισίου, όπου d_i είναι η συντεταγμένη της διάστασης i . Για παράδειγμα, το πλαίσιο που περιγράφεται από το διάνυσμα (υψηλή, σπίτι) στο 2-διάστατο χώρο, αντιστοιχεί στην κατάσταση με τιμές θερμοκρασία (d_1) = υψηλή και τοποθεσία (d_2) = σπίτι. Επιπλέον, στο (Yoon, Ahn, Lee, Hong & Kim, 2008) οι ερευνητές χρησιμοποίησαν για την αναπαράσταση του πλαισίου μια πολυδιάστατη σύνθετη μεταβλητή, της οποίας τα συστατικά περιελάμβαναν τις διαστάσεις του χρόνου, της τοποθεσίας, του ιστορικού αλληλεπίδρασης, του προγράμματος του χρήστη κ.λπ.

- **Σχήματα σήμανσης:** αυτά τα μοντέλα χρησιμοποιούν ιεραρχικές δομές δεδομένων που αποτελούνται από ετικέτες σήμανσης με ιδιότητες και περιεχόμενο. Ειδικότερα, το περιεχόμενο των ετικετών σήμανσης ορίζεται αναδρομικά από άλλες ετικέτες σήμανσης. Αντιπροσωπευτικά παραδείγματα αυτού του είδους μοντελοποίησης είναι τα προφίλ περιγραφής του πλαισίου. Σε αυτές τις περιπτώσεις χρησιμοποιείται κάποια παράγωγη γλώσσα της SGML, η οποία αποτελεί υπερκλάση όλων των γλωσσών σήμανσης, όπως η XML, για τον χαρακτηρισμό του πλαισίου. Μερικά από αυτά τα προφίλ (π.χ. Ferscha, Hechinger, Riener, Schmitzberger, Franz, Rocha & Zeidler, 2006) ορίζονται ως επεκτάσεις των προτύπων Composite Capabilities / Preferences Profile (CC/PP)¹ και User Agent Profile (UAProf)², ενώ άλλα ακολουθούν τη δική τους σήμανση (π.χ. Ryan, 1999).

Ένα ενδεικτικό παράδειγμα χρήσης σχημάτων σήμανσης για την περιγραφή του πλαισίου είναι η ContextML (Knappmeyer, Kiani, Frà, Moltchanov & Baker, 2010), ένα ελαφρύ XML σχήμα κατάλληλο για την αναπαράσταση του πλαισίου σε συσκευές κινητού υπολογισμού. Στην εικόνα 4 παρουσιάζεται το ContextML σχήμα για την αναπαράσταση ενός στοιχείου πλαισίου (ctxEl) που περιλαμβάνει πληροφορία σχετικά με το ποιος πάροχος πλαισίου (contextProvider) αναγνώρισε την κατάσταση του στοιχείου πλαισίου, με ποια οντότητα (entity) σχετίζεται, πότε δημιουργήθηκε (timestamp) και το

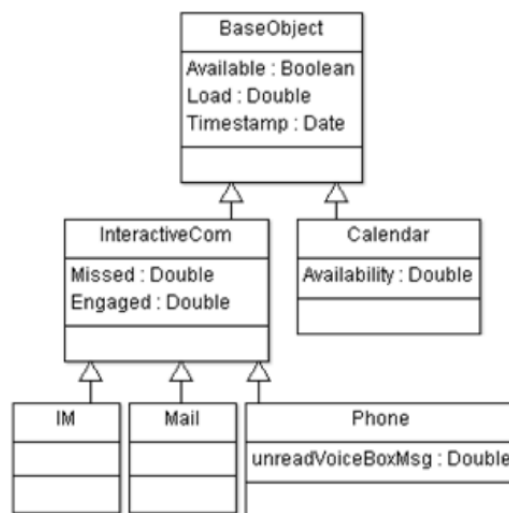
¹ <http://www.w3.org/Mobile/CCPP/>

² www.wapforum.org/what/technical/SPEC-UAProf-19991110.pdf

πραγματικό περιεχόμενο του πλαισίου (dataPart). Τα στοιχεία par, pars, parA είναι απλές κατασκευές που αποθηκεύουν ζεύγη κλειδιών – τιμών ή συλλογές χαρακτηριστικών (δομές και πίνακες).



Εικόνα 4. ContextML σχήμα αναπαράστασης πλαισίου (Knappmeyer, Kiani, Frà, Moltchanov & Baker, 2010)



Εικόνα 5. Μοντελοποίηση πλαισίου με UML για το σύστημα HEP (Chihani, Bertin, Jeanne & Crespi, 2011)

- **Γραφικά μοντέλα:** ένα δημοφιλές, γενικού σκοπού εργαλείο είναι η γλώσσα UML (Unified Modeling Language), η οποία διαθέτει ένα ισχυρό γραφικό

συστατικό (UML διαγράμματα). Λόγω της γενικής δομής της, η UML είναι κατάλληλη για τη μοντελοποίηση του πλαισίου.

Μια τέτοια περίπτωση, όπου το πλαίσιο μοντελοποιήθηκε με τη χρήση UML ήταν το σύστημα ενήμερης για το πλαίσιο επικοινωνίας HEP (Chihani, Bertin, Jeanne & Crespi, 2011). Η μοντελοποίηση αυτή με UML παρουσιάζεται στην εικόνα 5.

- **Αντικειμενοστρεφή μοντέλα:** ένα κοινό σημείο στις προσεγγίσεις αντικειμενοστρεφούς μοντελοποίησης του πλαισίου είναι η πρόθεση αξιοποίησης των κύριων οφελών της αντικειμενοστρέφειας, δηλαδή της ενθυλάκωσης και της επαναχρησιμοποίησης. Οι λεπτομέρειες της επεξεργασίας του πλαισίου ενθυλακώνονται σε επίπεδο αντικειμένου και παραμένουν κρυφές σε άλλα συστατικά. Η πρόσβαση στην πληροφορία του πλαισίου παρέχεται μόνο μέσα από υλοποιημένες διεπαφές.

Ένα παράδειγμα εφαρμογής του αντικειμενοστρεφούς μοντέλου παρουσιάστηκε στο (Hofer, Schwinger, Pichler, Leonhartsberger, Altmann & Retschitzegger, 2003). Η αφαιρετική περιγραφή του πλαισίου πραγματοποιήθηκε με UML (διάγραμμα κλάσεων), ενώ στην υλοποίησή του το σύστημα χρησιμοποιούσε κλάσεις και αντικείμενα.

- **Λογικά μοντέλα:** τα μοντέλα που βασίζονται στη λογική χαρακτηρίζονται από μεγάλο βαθμό τυπικότητας. Συνήθως, γεγονότα, εκφράσεις και κανόνες χρησιμοποιούνται για να οριστεί ένα μοντέλο πλαισίου. Ένα σύστημα που βασίζεται στη λογική χρησιμοποιείται τότε για τη διαχείριση των παραπάνω όρων και επιτρέπει την προσθήκη, ανανέωση ή διαγραφή γεγονότων. Μια διαδικασία εξαγωγής συμπερασμάτων (inference ή reasoning) χρησιμοποιείται για τον εντοπισμό νέων γεγονότων με βάση κανόνες που υλοποιούνται στο σύστημα. Η πληροφορία πλαισίου αναπαρίσταται με έναν τυπικό τρόπο με τη μορφή γεγονότων. Μια τέτοια προσέγγιση παρουσιάστηκε από τους Gray και Salber με την ονομασία Sensed Context Model (2001).
- **Μοντέλα οντολογιών:** Οι οντολογίες αποτελούν εργαλεία που περιγράφουν έννοιες και σχέσεις – διασυνδέσεις. Λόγω της υψηλής και τυπικής τους εκφραστικότητας, καθώς και των τεχνικών συμπερασμού που υποστηρίζουν, αποτελούν σημαντικά εργαλεία για την αναπαράσταση πληροφορίας πλαισίου. Σύμφωνα με τους Strang και Linnhoff-Popien (2004), οι οντολογίες ικανοποιούν τις περισσότερες απαιτήσεις για τη μοντελοποίηση του πλαισίου.

Οι Korpiää και Mäntyjärvi (2003) παρουσίασαν μερικές απαιτήσεις και στόχους που πρέπει να λαμβάνονται υπόψη κατά το σχεδιασμό των οντολογιών για την αναπαράσταση του πλαισίου:

- Ο *Απλότητα*. Οι εκφράσεις και οι σχέσεις πρέπει να είναι όσο το δυνατόν πιο απλές, ώστε να διευκολύνεται η εργασία των προγραμματιστών εφαρμογών.
- Ο *Ευελιξία και επεκτασιμότητα*. Η οντολογία πρέπει να υποστηρίζει την εύκολη προσθήκη νέων στοιχείων πλαισίου και σχέσεων.
- Ο *Γενικότητα*. Η οντολογία δεν πρέπει να είναι περιορισμένη, αλλά να υποστηρίζει διαφορετικά είδη πλαισίου.
- Ο *Εκφραστικότητα*. Η οντολογία πρέπει να επιτρέπει την περιγραφή όσο το δυνατόν περισσότερων καταστάσεων με αφαιρετική λεπτομέρεια.

Οι οντολογίες είναι αρκετά δημοφιλές εργαλείο, καθώς ένα μεγάλο πλήθος από ερευνητικές εργασίες τις αξιοποιούν για τη μοντελοποίηση του πλαισίου (π.χ. Korpiää, Jani, Kela & Malm, 2003; Korpiää & Mäntyjärvi, 2003; Korpiää, Häkkinen, Kela, Ronkainen & Käsälä, 2004; Wang, Zhang, Gu & Pung, 2004; Moore, Hu & Wan, 2010).

2.3 Μελέτες για το πρόβλημα της ανάκτησης των επαφών

Οι επαφές ως στοιχεία προσωπικής πληροφορίας έχουν απασχολήσει σε μεγάλο βαθμό τους ερευνητές, λόγω της σημασίας των συσκευών κινητού υπολογισμού ως εργαλεία επικοινωνίας. Αρκετές εργασίες ασχολούνται με το κύριο αποθετήριο επαφών (βιβλίο επαφών ή λίστα επαφών) και τις δυσκολίες που τίθενται στην πρόσβαση – ανάκτηση της επιθυμητής επαφής.

2.3.1 Μελέτες ευχρηστίας της λίστας επαφών

Παρότι η λίστα επαφών είναι από τις πιο συχνά χρησιμοποιούμενες εφαρμογές μιας κινητής συσκευής (Korninos & Liarakis, 2009) και η βελτίωση της σχεδίασής της θα έπρεπε να θεωρείται άμεση προτεραιότητα δεδομένων της σημασίας της και των περιορισμών της (Lugano, 2008), ένας μόνο μικρός αριθμός από ερευνητικές δημοσιεύσεις έχουν ασχοληθεί με την ευχρηστία της όσον αφορά την ανάκτηση και τη χρησιμοποίηση επαφών.

Σε μια από τις πρώτες σχετικές εργασίες, οι Böcker και Suwita (1999) εξέτασαν την ευχρηστία ενός κινητού τηλεφώνου Siemens C10. Στο εργαστηριακό πείραμα που πραγματοποίησαν ανέθεσαν στα υποκείμενά τους, ανάμεσα σε άλλα, τις παρακάτω εργασίες που σχετίζονται με τις επαφές:

1. Πραγματοποίηση μιας κλήσης από το βιβλίο επαφών της συσκευής (94% επιτυχία)
2. Αναζήτηση μιας επαφής από το βιβλίο επαφών κατά τη διάρκεια μιας κλήσης (73% επιτυχία)
3. Πραγματοποίηση μιας κλήσης από τη λίστα κλήσεων (73% επιτυχία)
4. Πραγματοποίηση μιας κλήσης από τη λίστα γρήγορων κλήσεων (87% επιτυχία)

Μπορεί να παρατηρήσει κανείς ότι ενώ σχεδόν όλοι οι χρήστες δεν αντιμετώπισαν πρόβλημα με την ανάκτηση και κλήση της ζητούμενης επαφής από το βιβλίο επαφών (94% επιτυχία), το ποσοστό επιτυχίας μειώθηκε στο 73% όταν τους ζητήθηκε να καλέσουν την ίδια επαφή από τη λίστα κλήσεων. Σε κάθε περίπτωση, οι σύγχρονες συσκευές παρέχουν πολύ πιο προχωρημένες διεπαφές σε σχέση με το Siemens C10, εφόσον διαθέτουν μεγαλύτερο διαθέσιμο χώρο στην οθόνη τους, ενώ οι χρήστες τους είναι πολύ πιο εξοικειωμένοι με τα εικονίδια σε σχέση με τους χρήστες την εποχή του πειράματος, στο οποίο οι ερευνητές κατέληξαν ότι είχαν καλύτερα αποτελέσματα αντικειμενικών μετρήσεων όταν χρησιμοποιήθηκε κείμενο αντί για εικονίδια, αλλά ταυτόχρονα χειρότερα αποτελέσματα υποκειμενικών μετρήσεων.

Σε ακόμα ένα πείραμα εκείνης της πρώιμης περιόδου (Klockar, Carr, Hedman, Johansson & Bengtsson, 2003) οι ερευνητές μελέτησαν την ευχρηστία κινητών τηλεφώνων Nokia και Siemens. Ζήτησαν από 9 συμμετέχοντες να πραγματοποιήσουν 26 εργασίες χρησιμοποιώντας τις συσκευές τους και μέτρησαν τα επιπλέον κλικ σε σχέση με τη βέλτιστη διαδρομή ενεργειών για την πραγματοποίηση της κάθε εργασίας. Το σύνολο των εργασιών περιελάμβανε τόσο συχνά πραγματοποιούμενες εργασίες σε μια κινητή συσκευή, όπως η πραγματοποίηση μιας κλήσης, όσο και λιγότερο συνηθισμένες εργασίες, όπως η απόκρυψη του αριθμού κλήσης. Σύμφωνα με τα αποτελέσματα του πειράματος, οι συμμετέχοντες ακολούθησαν διαδρομές κοντά στη βέλτιστη και επομένως δεν αντιμετώπισαν προβλήματα για εργασίες όπως η πραγματοποίηση μιας κλήσης, η αποστολή ενός μηνύματος και ο έλεγχος της τελευταίας κλήσης. Προβλήματα αναφέρθηκαν για την εργασία του ελέγχου των αναπάντητων κλήσεων (missed calls), ειδικά στην περίπτωση των συσκευών Nokia όπου οι κλήσεις τις οποίες απέρριπτε ρητά ο χρήστης καταχωρούνταν στη λίστα των ληφθέντων κλήσεων. Τέλος, κάποιοι συμμετέχοντες φαίνεται να αντιμετώπισαν μικρές δυσκολίες για την εύρεση ενός αριθμού από το βιβλίο επαφών.

Τα αποτελέσματα των δύο παραπάνω ερευνητικών εργασιών αποκτούν ξεχωριστή σημασία σε συνδυασμό με τις συζητήσεις των Bergman, Komninos, Liarokapis και

Clarke (2012) με χρήστες, σύμφωνα με τους οποίους για να αποφύγουν την αναζήτηση για μια επαφή με την οποία ήθελαν να επικοινωνήσουν, επισκέπτονταν πρώτα τη λίστα (αναπάντητων ή πρόσφατων) κλήσεων. Αυτή η πρακτική ακολουθείτο σε μεγαλύτερο βαθμό όταν θυμόντουσαν ότι επρόκειτο για επαφή με την οποία είχαν επικοινωνήσει πρόσφατα, καθώς θεωρούσαν αυτή τη στρατηγική πιο γρήγορη και αποτελεσματική από την αναζήτηση στο βιβλίο επαφών. Ο Whittaker (2011) είχε, επίσης, επισημάνει ότι η προηγούμενη γνώση μπορεί να κάνει πιο εύκολη την ανάκτηση προσωπικής πληροφορίας. Με βάση τα αποτελέσματα των αξιολογήσεων που πραγματοποιήθηκαν στις εργασίες που παρουσιάστηκαν παραπάνω, μια τέτοια στρατηγική δεν είναι πάντα η καλύτερη από άποψη αποτελεσματικότητας, καθώς μπορεί να δει κανείς ότι στην πρώτη περίπτωση η επιτυχία ανάκτησης μιας επαφής και κλήσης της έπεσε από το 94% στο 73% όταν χρησιμοποιήθηκε η λίστα κλήσεων, ενώ στη δεύτερη περίπτωση παρατηρήθηκαν προβλήματα (σημαντικά μεγαλύτερος αριθμός από πατήματα πλήκτρων) όταν ζητήθηκε από τους χρήστες να ελέγξουν τη λίστα με τις αναπάντητες κλήσεις.

Ακόμα μια μελέτη για την ευχρηστία του βιβλίου επαφών παρουσιάζεται στο (Bhamidipaty & Deepak, 2007), όπου δόθηκε έμφαση στους ημι-αναλφάβητους χρήστες και συγκεκριμένα σε όσους μπορούν να αναγνωρίσουν και να διαβάσουν αριθμούς αλλά όχι γράμματα. Το περιβάλλον του βιβλίου επαφών βασίζεται κυρίως σε αναπαραστάσεις κειμένου, γεγονός που δυσκολεύει γι' αυτή την κατηγορία χρηστών την ανάκτηση των επαφών. Σύμφωνα με τους ερευνητές, τα χαρακτηριστικά επικοινωνίας αυτής της κατηγορίας χρηστών μοιάζουν με αυτά των υπόλοιπων χρηστών κινητών συσκευών, καθώς έχουν ένα μικρό σύνολο επαφών που καλούν πολύ συχνά, ενώ με την πλειοψηφία των επαφών τους επικοινωνούν σπάνια. Στα πλαίσια της έρευνας υλοποιήθηκε μια διεπαφή που αντικαθιστά τα γράμματα με σύμβολα και ένας μηχανισμός ανάκτησης που βασίζεται σε αυτά τα σύμβολα. Σύμφωνα με τα συμπεράσματα της αξιολόγησης, το περιβάλλον του βιβλίου επαφών που χρησιμοποιεί σύμβολα αντί για γράμματα έτυχε θετικής αποδοχής από τους χρήστες, καθώς έκανε πιο απλή και εύκολη στη χρήση τη διαχείριση των επαφών.

2.3.2 Σύνθεση του αποθετηρίου επαφών και πρότυπα χρήσης επαφών και επικοινωνίας

Η μελέτη των προτύπων και της σχέσης επικοινωνίας των χρηστών με τις επαφές τους μπορεί να αναδείξει σημαντικά χαρακτηριστικά, η αξιοποίηση των οποίων μπορεί να χρησιμοποιηθεί για την αποτελεσματικότερη ανάκτηση των επαφών. Σε δυο από τις πρώτες μελέτες (Whittaker, Jones & Terveen, 2002a,b) από την ίδια ερευνητική ομάδα σχετικά με τη διαχείριση επαφών που χρησιμοποιούνται για την

ανταλλαγή μηνυμάτων ηλεκτρονικού ταχυδρομείου, αναγνωρίστηκε η δυσκολία διαχείρισής τους λόγω του εξαιρετικά μεγάλου πλήθους τους. Ανάμεσα στις πρακτικές που ακολουθούσαν οι χρήστες ήταν η χρήση διαφορετικών εργαλείων (π.χ. ψηφιακά και έντυπα βιβλία επαφών και αρχεία κειμένου) για την οργάνωσή τους ή η διατήρηση περισσότερων από ένα βιβλίο επαφών (π.χ. διαφορετικά βιβλία επαφών σε σταθερό υπολογιστή και PDA). Οι ερευνητές ανέδειξαν την ανάγκη διάκρισης των *σημαντικών* επαφών από τις υπόλοιπες και υποστήριξαν ότι κριτήρια που αφορούν το ιστορικό επικοινωνίας παίζουν σημαντικό ρόλο γι' αυτή τη διάκριση: συχνότητα, αμοιβαιότητα, προσφατότητα και μακροβιότητα επικοινωνίας. Μάλιστα, ανέφεραν ότι ανέμεναν μεγαλύτερη συχνότητα επικοινωνίας με τις σημαντικές επαφές, μεγαλύτερη αμοιβαιότητα, πιο πρόσφατη επικοινωνία και αλληλεπίδραση για μεγαλύτερα χρονικά διαστήματα. Αντίθετα, υποστήριξαν ότι μηνύματα που συχνά δεν απαντώνται υποδεικνύουν μη σημαντικές επαφές. Επιπλέον, υποστήριξαν ότι χρήστες με μεγαλύτερη συχνότητα χρήσης θα είχαν και μεγαλύτερο ποσοστό σημαντικών επαφών. Οι εικασίες τους επιβεβαιώθηκαν από τα πειράματα που πραγματοποίησαν. Ακόμα ένα σημαντικό εύρημα ήταν πως οι χρήστες αναγνώρισαν ένα μικρό αριθμό από επαφές ως σημαντικές, ενώ τα αποθετήριά τους περιείχαν από εκατοντάδες έως χιλιάδες εγγραφές.

Οι Bergman, Komninos, Liarokapis και Clarke (2012) μελέτησαν την επίπτωση των μη χρησιμοποιούμενων επαφών στα αποθετήρια και πρότειναν ένα πρωτότυπο που βελτιώνει τη λίστα επαφών ως προς την αποτελεσματικότητα της ανάκτησης. Πιο συγκεκριμένα, πρότειναν το διαχωρισμό των επαφών σε τρεις κατηγορίες ανάλογα με τη συχνότητα επικοινωνίας μαζί τους: συχνά χρησιμοποιούμενες, περιστασιακά χρησιμοποιούμενες και μη χρησιμοποιούμενες. Στην τελευταία κατηγορία ανήκουν επαφές που οι χρήστες δεν έχουν καλέσει για πολύ μεγάλο χρονικό διάστημα (οι ερευνητές το όρισαν στους 6 μήνες) ή και ποτέ. Μερικές από αυτές τις μη χρησιμοποιούμενες επαφές έχουν πάψει να σχετίζονται με τους χρήστες (π.χ. ένας πρώην εργοδότης), άλλες είναι επαφές που δε χρησιμοποιήθηκαν ποτέ (π.χ. ένας πρώην συμμαθητής με τον οποίο ο χρήστης αντάλλαξε στοιχεία επικοινωνίας από ευγένεια), ενώ μερικές επαφές μπορεί να είναι τόσο παλιές που ο χρήστης να μην είναι σίγουρος αν τα στοιχεία τους είναι ακόμα έγκυρα. Σε ένα εργαστηριακό πείραμα, αφού εξήγαγαν τις λίστες επαφών από τα κινητά τηλέφωνα των συμμετεχόντων, τους ζήτησαν να υποδείξουν από μια προκαθορισμένη λίστα απαντήσεων για κάθε επαφή πότε επικοινωνήσαν για τελευταία φορά μαζί της. Τα αποτελέσματα έδειξαν ότι το 47% των συνολικών επαφών ήταν μη χρησιμοποιούμενες επαφές (καμία επικοινωνία τους τελευταίους 6 μήνες), ενώ το

μέσο ποσοστό των μη χρησιμοποιούμενων επαφών ανά χρήστη ήταν 39%. Επιπλέον, το ποσοστό των συχνά χρησιμοποιούμενων (τουλάχιστον μια φορά την εβδομάδα) επαφών ήταν 15%, ενώ το ποσοστό των περιστασιακά χρησιμοποιούμενων (για χρονικό διάστημα από μια βδομάδα έως 6 μήνες) επαφών 38%. Ένα άλλο σημαντικό εύρημα ήταν πως από τους 18 συμμετέχοντες μόνο ένας (που δεν είχε καμία μη χρησιμοποιούμενη επαφή) ανέφερε ότι έσβηνε επαφές από τη λίστα. Όσον αφορά το μέσο μέγεθος του αποθετηρίου επαφών, αυτό ήταν 92,47 (sd=56,93, min=17, max=202), ενώ βρέθηκε ισχυρή συσχέτιση ανάμεσα στο μέγεθος του αποθετηρίου και το πλήθος των μη χρησιμοποιούμενων επαφών.



Εικόνα 6. Η διαδικασία ανάκτησης με ένα τυπικό NOKIA UI (αριστερά) και το πρωτότυπο των Bergman, Komninos, Liarakis και Clarke (2012) (δεξιά)

Επομένως, σχεδόν οι μισές επαφές από το βιβλίο επαφών μιας συσκευής κινητού υπολογισμού δε χρησιμοποιούνται, εντούτοις δυσκολεύουν την ανάκτηση από τις

υπόλοιπες επαφές. Βασιζόμενοι σε αυτή την παρατήρηση και ακολουθώντας την αρχή του υποβιβασμού (demotion principal) (Bergman, Tucker, Beyth-Marom, Cutrell & Whittaker, 2009) από την υποκειμενική για τον χρήστη προσέγγιση (user subjective approach) (Bergman, Beyth Marom, & Nachmias, 2003; Bergman, Beyth Marom, & Nachmias, 2008), οι ερευνητές ανέπτυξαν ένα πρωτότυπο (εικόνα 6) που υποβιβάζει τις μη χρησιμοποιούμενες επαφές, καθώς σε κάθε οθόνη του βιβλίου επαφών τις τοποθετεί στο τέλος της λίστας και με μικρότερη γραμματοσειρά, ταξινομημένες με αλφαβητική σειρά. Σε αυτή την περίπτωση το κριτήριο υποκειμενικότητας είναι η έλλειψη σημαντικότητας για κάποιες επαφές, η οποία υπολογίζεται αυτόματα καθώς το σύστημα ερμηνεύει τις πράξεις των χρηστών. Όταν οι χρήστες εισέρχονται στο βιβλίο επαφών, εμφανίζεται μια οθόνη με όλες τις τακτικά χρησιμοποιούμενες επαφές και από κάτω όλες τις μη χρησιμοποιούμενες επαφές (οι οποίες συνήθως δε φαίνονται στην πρώτη οθόνη). Καθώς οι χρήστες πληκτρολογούν γράμματα που αντιστοιχούν στο όνομα της ζητούμενης επαφής, τα αποτελέσματα φιλτράρονται, αλλά και πάλι εμφανίζονται πρώτα με αλφαβητική σειρά οι τακτικά χρησιμοποιούμενες επαφές και από κάτω οι μη χρησιμοποιούμενες. Η εργαστηριακή αξιολόγηση του πρωτοτύπου έδειξε ότι για την ανάκτηση των επαφών απαιτείτο στατιστικά σημαντικά μικρότερο πλήθος από πατήματα πλήκτρων (κατά μέσο όρο μείωση 1,96 πατήματα) και λιγότερος χρόνος (βελτίωση κατά μέσο όρο 340 msec) σε σχέση με την παραδοσιακή διεπαφή. Επιπλέον, βρέθηκε στατιστικά σημαντική συσχέτιση ανάμεσα στο ποσοστό των υποβιβαζόμενων επαφών και τη μείωση στο χρόνο ανάκτησης, καθώς και ανάμεσα στο μέγεθος του αποθετηρίου και τον αριθμό των βημάτων για την πραγματοποίηση μιας κλήσης. Επίσης, με βάση τα σχόλια των υποκειμένων φαίνεται ότι το πρωτότυπο είχε θετική αποδοχή.

Οι ερευνητές τόνισαν ότι, σύμφωνα με τα αποτελέσματα της έρευνάς τους, το πρωτότυπό τους και η λίστα πρόσφατων κλήσεων είναι περισσότερο συμπληρωματικά εργαλεία παρά ανταγωνιστικά. Οι χρήστες μπορούν να χρησιμοποιούν τη λίστα πρόσφατων κλήσεων για να ανακτήσουν κάποια από τις συχνά χρησιμοποιούμενες επαφές, που αποτελούν το 15% του συνολικού αριθμού των επαφών και με τις οποίες επικοινωνούν τουλάχιστον μια φορά την εβδομάδα. Για την ανάκτηση από το 38% των περιστασιακά χρησιμοποιούμενων επαφών, καθώς οι παραδοσιακές διεπαφές δεν τις ξεχωρίζουν από το 47% των μη χρησιμοποιούμενων επαφών, μπορούν να χρησιμοποιήσουν το πρωτότυπο που ανέπτυξαν.

Σε μια άλλη εργασία (Komninos & Liarokapis, 2009), οι ερευνητές παρουσίασαν ένα μοντέλο σχεδίασης που διευκολύνει την ανάκτηση των σημαντικών επαφών, εντοπίζοντάς τες και προωθώντας τες σε σχέση με τις υπόλοιπες. Αφορμή για την

ανάπτυξή του αποτέλεσε η πρακτική ορισμένων χρηστών να προωθούν επαφές τις οποίες θεωρούν σημαντικές στην κορυφή της λίστας επαφών, προσθέτοντας κάποιο γράμμα (συνήθως το α), αριθμό ή σύμβολο στην αρχή του ονόματός τους, μια στρατηγική που αναφέρεται και στο (Jung, Anttila & Blom, 2008). Πρότειναν τέσσερα κριτήρια πλαισίου που θα μπορούσαν να χρησιμοποιηθούν για την αναγνώριση των σημαντικών επαφών:

- C1 – Συχνότητα και προσφατότητα επικοινωνίας
- C2 – Τοποθεσία και χρονικό πλαίσιο
- C3 – Πλαίσιο εργασίας και δραστηριότητας
- C4 – Προσωπική προτίμηση

Οι ερευνητές μελέτησαν το κριτήριο C1, εξάγοντας σε ένα εργαστηριακό πείραμα τις επαφές από τις συσκευές των συμμετεχόντων. Κατά μέσο όρο κάθε χρήστης είχε 107,7 επαφές ($sd=81,46$, $min=16$, $max=372$) στο αποθετήριό του, γεγονός που επιβεβαίωσε την πεποίθηση ότι η αναζήτηση πρέπει να πραγματοποιηθεί σε ένα σημαντικά μεγάλο σύνολο. Κάθε χρήστης υπέδειξε για κάθε επαφή πότε επικοινωνήσε τελευταία φορά μαζί της, καθώς και ποιες επαφές θεωρούσε ως πολύ σημαντικές. Προέκυψε ότι 31% των επαφών χρησιμοποιήθηκαν τον τελευταίο μήνα, ανεβαίνοντας στο 40% για τους τελευταίους δύο μήνες. Αυτό το ποσοστό αντιστοιχεί μόλις στο 22% των επαφών που χρησιμοποιήθηκαν τους τελευταίους 6 μήνες. Σημαντικό εύρημα ήταν ότι σχεδόν οι μισές επαφές (46%) δεν είχαν χρησιμοποιηθεί τους τελευταίους 6 μήνες ή και ποτέ. Όσον αφορά τις σημαντικές επαφές, βρέθηκε ότι το 43% χρησιμοποιείται σε εβδομαδιαία βάση, ενώ ακόμα ένα 37% σε μηνιαία. Αυτά τα ευρήματα φαίνεται να επιβεβαιώνουν ότι η συχνότητα επικοινωνίας είναι ισχυρός δείκτης της σημαντικότητας. Επιπλέον, το γεγονός ότι σημαντικές επαφές υπάρχουν σε όλες τις κατηγορίες συχνότητας υποδεικνύει ότι το κριτήριο C4 είναι πιθανώς σημαντικός παράγοντας σχεδίασης αλγορίθμων και διεπαφών ανάκτησης. Τέλος, οι ερευνητές ανέφεραν ότι το πλήθος των σημαντικών επαφών παραμένει σχετικά σταθερό, άσχετα από το μέγεθος της λίστας επαφών και κατά συνέπεια μια σχετική διεπαφή θα πρέπει να προσφέρει γρήγορη πρόσβαση σε περίπου 20 επαφές.

Μια μελέτη σχετικά με την κοινωνική δικτύωση μέσω συσκευών κινητού υπολογισμού (Lugano, 2008) αποκάλυψε πως οι χρήστες δεν εκμεταλλεύονται πλήρως τις δυνατότητες των νέων βιβλίων επαφών. Μόνο το 22% ανέφερε πως κατά την εισαγωγή μιας επαφής αποθήκευε τη διεύθυνση ηλεκτρονικού ταχυδρομείου, ενώ μόλις το 56% των χρηστών βρέθηκε να έχει συνδέσει περισσότερους από έναν αριθμούς σε μια επαφή. Τα βιβλία επαφών στο 50% των περιπτώσεων περιείχαν ένα

πλήθος ανάμεσα σε 100 και 200 επαφές, στο 33% των περιπτώσεων λιγότερες από 100 επαφές και στο 17% των περιπτώσεων περισσότερες από 200 επαφές. Η ανάλυση των αρχείων κλήσεων (με έκταση 1-2 εβδομάδες) έδειξε ότι οι χρήστες επικοινωνούσαν κατά μέσο όρο με 21 διαφορετικές επαφές (24% των επαφών στο αποθετήριο), ακολουθώντας την αρχή Pareto 80/20 (Pareto, 1935). Οι συμμετέχοντες με τα μεγαλύτερα βιβλία επαφών ήταν και οι περισσότερο ενεργοί κοινωνικά, με αλληλεπιδράσεις κατά μέσο όρο με 30 επαφές. Σύμφωνα με τη μελέτη, οι κλήσεις και τα μηνύματα χρησιμοποιούνταν για τακτική επικοινωνία με την οικογένεια, τους φίλους και τους συντρόφους. Το 44% των χρηστών ανέφεραν ότι επικοινωνούσαν με τους φίλους τους κάθε μέρα, ενώ το 50% κάθε βδομάδα. Η ανάλυση των αρχείων επικοινωνίας του συγγραφέα (που συνδεόταν με ισχυρές ή ασθενείς σχέσεις με όλους τους συμμετέχοντες) έδειξε ότι η πραγματική επικοινωνία πραγματοποιούνταν με μικρότερη συχνότητα. Αυτό το εύρημα υποδηλώνει ότι η πραγματική επικοινωνία είναι μικρότερη σε σχέση με αυτή που αντιλαμβάνονται οι χρήστες και επομένως οι μελέτες που βασίζονται σε ποσοτικά στοιχεία παρέχουν πιο αξιόπιστη εικόνα σε σχέση με όσες βασίζονται σε ποιοτικά.

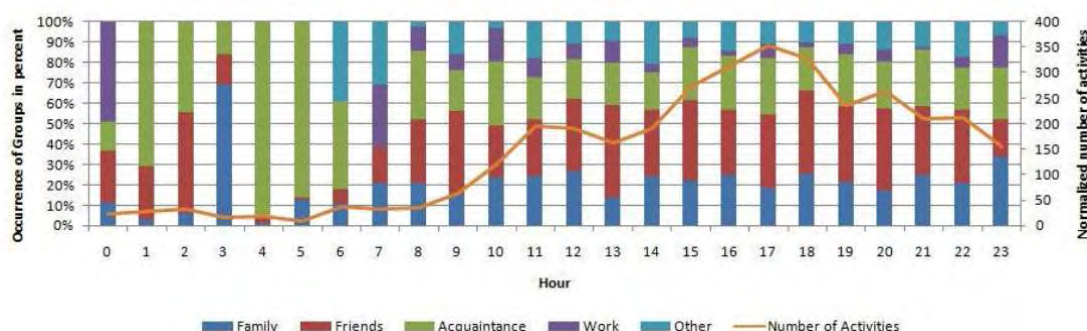
Πίνακας 1. Έρευνα για τις μεθόδους πραγματοποίησης κλήσεων από τους Lee, Seo και Lee (2010)

Πώς πραγματοποιείτε μια κλήση;	Ποσοστό απαντήσεων
Με αναζήτηση του ονόματος στο βιβλίο επαφών	72%
Επιλέγοντας μια επαφή από τη λίστα πρόσφατων κλήσεων	64%
Χρησιμοποιώντας μια συντόμευση πληκτρολογίου	42%
Πληκτρολογώντας μερικώς ένα όνομα στην αρχική οθόνη	28%
Πληκτρολογώντας μερικώς ένα νούμερο στην αρχική οθόνη	25%
Με αναζήτηση κατηγοριών στο βιβλίο επαφών	17%
Επιλέγοντας έναν αριθμό από τη λίστα των SMS	16%
Πληκτρολογώντας ολόκληρο το νούμερο	9%

Στην ερευνητική εργασία που παρουσιάζεται στο (Lee, Seo & Lee, 2010) μελετήθηκαν τα πρότυπα επικοινωνίας των χρηστών κινητών τηλεφώνων. Για το σκοπό αυτό πραγματοποιήθηκαν δύο διαφορετικές έρευνες: η πρώτη περιελάμβανε τη συμπλήρωση ερωτηματολογίων, ενώ η δεύτερη την ανάλυση των αρχείων κλήσεων από τα τηλέφωνα των συμμετεχόντων. Προέκυψε ότι από όλους τους πιθανούς τρόπους πραγματοποίησης μιας κλήσης, η αναζήτηση πληκτρολογώντας το όνομα της επαφής στο βιβλίο επαφών και η επιλογή της επαφής από τη λίστα πρόσφατων κλήσεων αποτελούσαν τις δημοφιλέστερες μεθόδους με ποσοστά απαντήσεων 72% και 64%, όπως φαίνεται και στον πίνακα 1. Επιπροσθέτως, το 67% των χρηστών δήλωσαν ότι έχουν ένα αριθμό που καλούν τακτικά, με τους γονείς (51%) και

συντρόφους (44%) να βρίσκονται στην κορυφή των επιλογών και φίλους (21%) και αδέρφια (15%) να ακολουθούν. Η ανάλυση των αρχείων κλήσεων έδειξε ότι η επικοινωνία των χρηστών με τις επαφές τους ακολουθεί περιοδικά πρότυπα και αντιπροσωπεύει τα πρότυπα της ζωής τους.

Στο (Le, Henze & Schmidt, 2013) από την ανάλυση των απαντήσεων χρηστών συσκευών κινητού υπολογισμού σε ένα ερωτηματολόγιο προέκυψαν τα εξής αποτελέσματα για το μέγεθος του αποθετηρίου επαφών: 20% των συμμετεχόντων ανέφεραν ότι έχουν λιγότερες από 50 επαφές, 29% μεταξύ 50 και 100 και 33% ανέφεραν ότι έχουν περισσότερες από 100 επαφές. Το 78% έστειλαν καθημερινά μηνύματα και το 57% πραγματοποιούσαν καθημερινά κλήσεις. Στην εικόνα 7 παρουσιάζεται η ανάλυση της δραστηριότητας φωνητικής επικοινωνίας των χρηστών με 5 ομάδες επαφών που αναγνωρίστηκαν στη διάρκεια της ημέρας. Η μεγαλύτερη δραστηριότητα λάμβανε χώρα το απόγευμα κυρίως με επαφές από την ομάδα των φίλων.



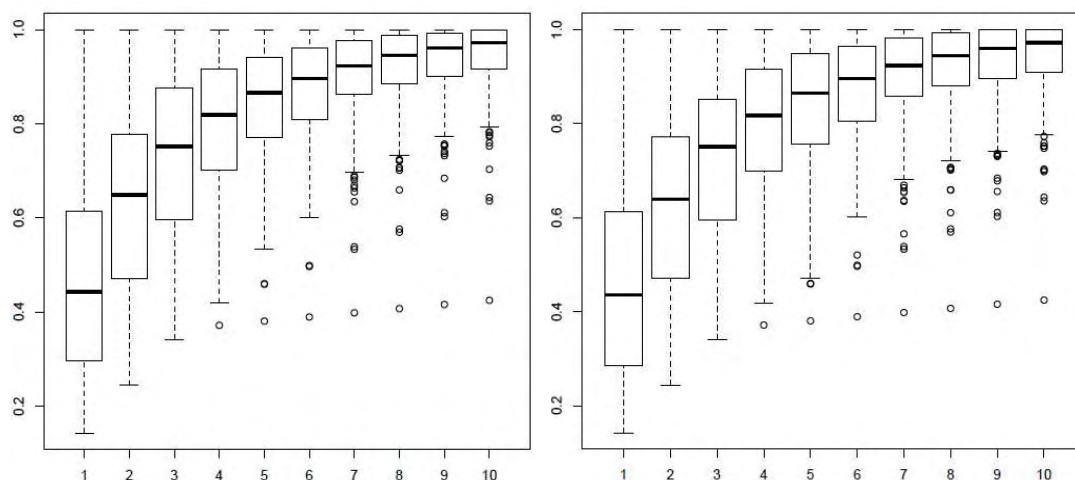
Εικόνα 7. Κλάσματα κλήσεων με επαφές από τις πέντε ομάδες κατά τη διάρκεια της ημέρας (Le, Henze & Schmidt, 2013)

Σε μια από τις πιο πρόσφατες και πιο σημαντικές εργασίες (Bentley, & Chen, 2015) που σχετίζονται με τη χρήση των επαφών από τους χρήστες συσκευών κινητού υπολογισμού, οι ερευνητές μελέτησαν διεξοδικά το βιβλίο επαφών ως προς το περιεχόμενο και τα χαρακτηριστικά επικοινωνίας. Τα ερευνητικά τους ερωτήματα σχετίζονται με τη σύνθεση του αποθετηρίου επαφών, τις ακολουθούμενες πρακτικές επικοινωνίας στις συσκευές κινητού υπολογισμού και τη χρήση των επαφών. Έναυσμα γι' αυτή τη μελέτη αποτέλεσε η δυνατότητα των κινητών συσκευών να συγχρονίζουν τα βιβλία επαφών τους με άλλες υπηρεσίες (π.χ. Twitter, Facebook, Skype κ.λπ.) μετατρέποντας τα αποθετήρια από λίστες με το σχετικά περιορισμένο πλήθος των προσώπων με τα οποία επικοινωνούν με μηνύματα ή κλήσεις σε λίστες χιλιάδων επαφών που έχουν συλλεγεί από τα διαδικτυακά κοινωνικά δίκτυα στα οποία συμμετέχουν οι χρήστες. Οι ερευνητές πραγματοποίησαν μια διαδικτυακή

έρευνα με 200 συμμετέχοντες με Android συσκευές, από τους οποίους ζήτησαν να εγκαταστήσουν μια εφαρμογή που εξήγαγε στοιχεία από τη λίστα επαφών και μεταδεδομένα σχετικά με την επικοινωνία. Το δείγμα τους είχε χαρακτηριστικά αντιπροσωπευτικά του πληθυσμού των Η.Π.Α. Κατά τη διάρκεια της διαδικασίας παρουσιάστηκαν στους συμμετέχοντες 25 τυχαίες επαφές τους και τους ζητήθηκε να απαντήσουν σε ερωτήσεις όπως πότε επικοινωνήσαν για τελευταία φορά, τι είδους σχέση είχαν (π.χ. φίλοι, οικογένεια, επαγγελματική κ.λπ.) κ.α.

Όσον αφορά τη σύνθεση του αποθετηρίου επαφών, το μέσο μέγεθος του ήταν 308 επαφές με 25% των συμμετεχόντων να έχουν λίστες με περισσότερες από 360 επαφές, ενώ 14 (7%) από αυτούς είχαν περισσότερες από 1.000 επαφές. Τα αποθετήρια των επαφών κυμαίνονταν σε μέγεθος από 26 έως 3.038 επαφές. Σχετικά με την προέλευση των επαφών, το 84% των χρηστών δήλωσαν ότι έχουν συνδέσει με το αποθετήριο το λογαριασμό τους στο Gmail/Google (γεγονός αναμενόμενο σε τηλέφωνα με το λειτουργικό σύστημα Android), το 70% το λογαριασμό τους στο Facebook και το 17% το λογαριασμό τους στο Yahoo. Καθώς τα ποσοστά των Αμερικανών που χρησιμοποιούν το Facebook και το Yahoo είναι 71% και 19% αντίστοιχα, είναι φανερό ότι ένα μεγάλο ποσοστό χρηστών επιλέγουν να συγχρονίζουν τις υπηρεσίες αυτές στο τηλέφωνό τους. Το 11% των συμμετεχόντων δε συγχρόνιζαν τη λίστα επαφών με καμία υπηρεσία. Ένα ενδιαφέρον εύρημα της έρευνας είναι ότι το 29% των επαφών που παρουσιάζονταν στους χρήστες ήταν επαφές που δεν αναγνώριζαν, κάτι το οποίο δεν ανέμεναν και τους προβλημάτισε. Πιθανότατα αυτές οι επαφές προέρχονταν από διάφορες υπηρεσίες στις οποίες οι χρήστες έχουν συνδέσεις με άτομα τα οποία δεν έχουν συναντήσει ποτέ από κοντά και δεν επικοινωνούν (π.χ. επαφές από το επαγγελματικό κοινωνικό δίκτυο LinkedIn ή από το Facebook). Τις επαφές τις οποίες αναγνώρισαν οι χρήστες, τις κατέταξαν κυρίως στην κατηγορία φίλοι (23%), οικογένεια (13%) και συνάδελφοι (9%), ενώ το 11% των επαφών αναφέρθηκε ως επιχειρήσεις (π.χ. εστιατόρια). Οι αναφορές σχετικά με την ισχύ των δεσμών με κάθε επαφή έδειξαν ότι σχεδόν οι μισές επαφές (46%) θεωρούνταν εντελώς ασήμαντες. Το 6% των επαφών αξιολογήθηκαν ως επαφές με τους πιο ισχυρούς δεσμούς με τον χρήστη. Επομένως, σε ένα βιβλίο επαφών μέσου μεγέθους το ποσοστό αυτό αντιστοιχεί σε περίπου 10 επαφές που θεωρούνται πολύ σημαντικές. Μελετώντας τα πεδία των εγγραφών που αντιστοιχούν στις επαφές, οι ερευνητές παρατήρησαν ότι το 55% των εγγραφών του βιβλίου επαφών είχαν συμπληρωμένο αριθμό τηλεφώνου, ενώ το 45% είχαν διεύθυνση email. Το 5% των επαφών είχαν και διεύθυνση email και αριθμό τηλεφώνου, ενώ ένα άλλο 5% δεν είχαν ούτε αριθμό τηλεφώνου ούτε διεύθυνση email. Αυτές οι τελευταίες επαφές μπορεί να

προέρχονταν από συνδεδεμένες πηγές όπως το Facebook, το Twitter και το LinkedIn, όπου οι συνδεδεμένοι χρήστες δε μοιράζονταν στοιχεία επικοινωνίας με τον συμμετέχοντα κάτοχο του τηλεφώνου.

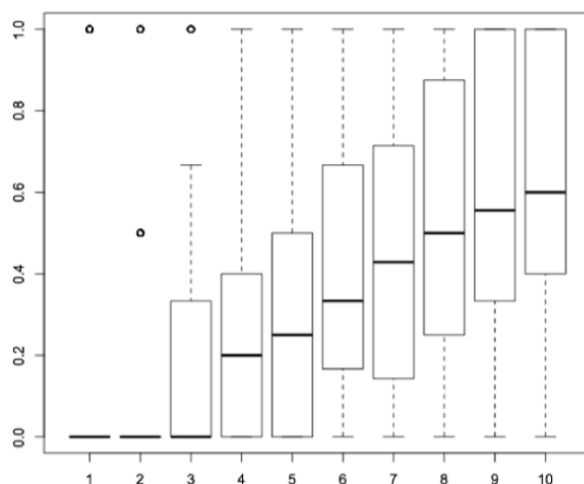


Εικόνα 8. Το ποσοστό των κλήσεων (αριστερά) και SMS (δεξιά) για τις N πιο συχνά χρησιμοποιούμενες επαφές (Bentley & Chen, 2015)

Όσον αφορά τα πρότυπα επικοινωνίας, η ανάλυση των αρχείων καταγραφής των κλήσεων και των μηνυμάτων έδειξε ότι το 84% των κλήσεων και το 81% των μηνυμάτων με άτομα που εμφανίζονταν ως εγγραφές του βιβλίου επαφών πραγματοποιείται με τους 5 ανθρώπους με τη συχνότερη επικοινωνία. Μελετώντας τις 10 επαφές με τη συχνότερη επικοινωνία, τα παραπάνω ποσοστά γίνονται 94% για τις κλήσεις και 92% για τα μηνύματα (εικόνα 8). Εντούτοις, τα περισσότερα βιβλία επαφών παρέχουν την αλφαβητική λίστα επαφών ως τον προκαθορισμένο μηχανισμό επιλογής. Σε μερικές περιπτώσεις υπάρχει η δυνατότητα χαρακτηρισμού επαφών από τους χρήστες ως «αγαπημένες» αλλά αυτή η δυνατότητα φαίνεται να χρησιμοποιείται σπάνια. Οι ερευνητές, επιπλέον, ήθελαν να διερευνήσουν ποιες επαφές δεν είχαν χρησιμοποιηθεί ποτέ, καθώς και πότε έγινε η τελευταία επικοινωνία με την κάθε επαφή από τις υπόλοιπες. Για την περίπτωση των κλήσεων, οι Android συσκευές συνήθως περιορίζουν το αρχείο καταγραφής κλήσεων στις 500 εγγραφές. Αυτές οι 500 κλήσεις αναπαριστούν επικοινωνία με 17 (6% των συνολικών επαφών) διακριτές επαφές κατά μέσο όρο. Για τα μηνύματα δεν υπάρχει κάποιος περιορισμός καταγραφής, επομένως εξαρτάται από τον χρήστη αν θα σβήσει κάποια ή όχι (κάτι που δε φαίνεται να λαμβάνουν υπόψη οι ερευνητές). Στα αρχεία καταγραφής των μηνυμάτων των συμμετεχόντων βρέθηκαν κατά μέσο όρο 19 (6% των συνολικών επαφών) διακριτές επαφές. Ιδιαίτερο ενδιαφέρον έχει ο πληθυσμός του 25% των συμμετεχόντων που είχαν καλέσει λιγότερες από 8 επαφές ή είχαν επικοινωνήσει με μηνύματα με λιγότερες από 6 επαφές.

Ενδιαφέρον παρουσιάζουν και τα ευρήματα που αφορούν την προσφατότητα της επικοινωνίας. Για τις επαφές με τις οποίες ανταλλάχθηκαν μηνύματα SMS, ο μέσος χρόνος από την τελευταία επικοινωνία ήταν 26 μέρες, ενώ με το 25% αυτών των επαφών υπήρξε επικοινωνία τις τελευταίες 8 μέρες. Για τις επαφές με τις οποίες υπήρξε τηλεφωνική επικοινωνία, ο μέσος χρόνος από την τελευταία κλήση ήταν περίπου 36 μέρες, ενώ το 25% αυτών των επαφών κλήθηκαν τις τελευταίες 17 μέρες. Από το σύνολο των κλήσεων που αναλύθηκαν ποσοστό 27% ήταν εισερχόμενες και απαντημένες, 19% εισερχόμενες και αναπάντητες και το 54% ήταν εξερχόμενες. Το 31% των κλήσεων ήταν με αριθμούς που δεν ήταν αποθηκευμένοι σε κάποια επαφή του βιβλίου επαφών. Από αυτές τις κλήσεις το 55% αντιστοιχεί σε εξερχόμενες επικοινωνίες. Αξιοσημείωτο είναι επίσης πως το 65% των εισερχόμενων κλήσεων από άγνωστους αριθμούς δεν απαντήθηκαν.

Μελετώντας τις αναφορές που συμπληρώθηκαν, οι ερευνητές βρήκαν ότι οι συμμετέχοντες ανέφεραν πως με το 33% από το σύνολο των επαφών δεν είχαν επικοινωνήσει ποτέ. Επίσης, ανέφεραν ότι με το 55% των επαφών δεν είχαν επικοινωνήσει με κανένα τρόπο (συμπεριλαμβάνοντας συναντήσεις πρόσωπο με πρόσωπο ή επικοινωνίες μέσω των υπηρεσιών κοινωνικής δικτύωσης) τους τελευταίους 6 μήνες. Το αντίστοιχο ποσοστό που προκύπτει από το πεδίο “last contacted” των επαφών (που όμως περιλαμβάνει μόνο επικοινωνίες που πραγματοποιήθηκαν μέσω της συσκευής) είναι 80%.



Εικόνα 9. Ποσοστό των N πιο συχνά χρησιμοποιούμενων επαφών που περιλαμβάνονται στις N πιο πρόσφατες επαφές (Bentley & Chen, 2015)

Στην εικόνα 9 απεικονίζεται η σχέση ανάμεσα στην προσφατότητα και τη συχνότητα επικοινωνίας. Παρουσιάζεται για N=1...10 το ποσοστό των N πιο συχνά χρησιμοποιούμενων επαφών που εμφανίζονται στη λίστα με τις N πιο πρόσφατες

επαφές. Για $N=10$ ο διάμεσος των πιο συχνά χρησιμοποιούμενων επαφών που εμφανίζονται στη λίστα των πρόσφατων επαφών είναι 6. Επομένως, η λίστα των πρόσφατων κλήσεων μπορεί να είναι ένα πολύ καλό εργαλείο για την επικοινωνία «προσωρινών» επαφών που είναι σχετικές μόνο για μια μέρα ή για ένα σαββατοκύριακο, αλλά αποτυγχάνει σε μεγάλο βαθμό να προσφέρει πρόσβαση στις πιο συχνά χρησιμοποιούμενες επαφές.

Τέλος, αρκετοί από τους συμμετέχοντες ανέφεραν ότι θα ήθελαν να «καθαρίσουν» το βιβλίο επαφών τους, είτε διαγράφοντας επαφές που δε χρησιμοποιούν είτε αποσυνδέοντας υπηρεσίες που συγχρονίζονται με το βιβλίο επαφών. Όμως, οι περισσότεροι αντιλαμβάνονταν ότι πρόκειται για μια προσπάθεια που απαιτεί χρόνο και προσπάθεια και κατέληξαν πως πιθανότατα δε θα πραγματοποιούσαν αυτή την εργασία.

2.3.3 Εναλλακτικά πρωτότυπα για τη λίστα επαφών

Η αδυναμία των παραδοσιακών μεθόδων ανάκτησης επαφών να ικανοποιήσουν τις ανάγκες των χρηστών συσκευών κινητού υπολογισμού έχει οδηγήσει την ερευνητική κοινότητα σε μελέτη εναλλακτικών πρωτοτύπων που θα μπορούσαν να τις αντικαταστήσουν.

Οι Jung, Anttila και Blom (2008) αναγνώρισαν την αναγκαιότητα για επανασχεδιασμό της λίστας επαφών των συσκευών κινητού υπολογισμού. Ανάμεσα στις οδηγίες στις οποίες θεώρησαν πως πρέπει να βασιστεί ένας τέτοιος σχεδιασμός είναι:

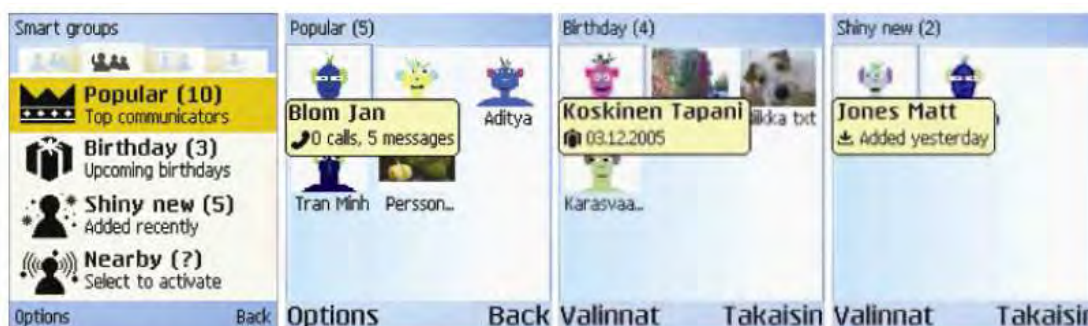
- Αποτελεσματικότητα της πρόσβασης στις επαφές: η ευκολία και ταχύτητα πρόσβασης στις πληροφορίες των επαφών αποτελούν παράγοντες κλειδιά για την εμπειρία του χρήστη, επομένως αυτό το χαρακτηριστικό θα πρέπει να είναι η κύρια προτεραιότητα για τον σχεδιασμό νέων διεπαφών για τη λίστα επαφών.
- Διαφοροποίηση σημαντικών επαφών: οι χρήστες έχουν εκφράσει την ανάγκη για διαφοροποίηση των σημαντικών επαφών από τις υπόλοιπες για συναισθηματικούς και πρακτικούς λόγους, ενώ έχει παρατηρηθεί η προώθηση επαφών στην κορυφή της λίστας από τους ίδιους τους χρήστες προσθέτοντας νούμερα ή σύμβολα στην αρχή του ονόματός τους, στρατηγική που όπως σημειώθηκε νωρίτερα αναφέρεται και στο (Komninos & Liaroikaris, 2009). Καθώς η φύση των σχέσεων ανάμεσα στους ανθρώπους αλλάζει με το χρόνο και όλες οι επαφές δεν έχουν την ίδια σημασία, οι συγγραφείς τόνισαν την

ιδέα μιας δυναμικής λίστας επαφών που βασίζεται στη συχνότητα και προσφατότητα της επικοινωνίας.

Οι ερευνητές βασιζόμενοι στις οδηγίες που πρότειναν, σχεδίασαν και υλοποίησαν ένα πρωτότυπο που αντικαθιστά το βιβλίο επαφών. Το πρωτότυπο περιλαμβάνει μια οθόνη με τις “top” επαφές (εικόνα 10): πρόκειται για μια υβριδική διεπαφή που παρουσιάζει είτε τις 10 πιο συχνά χρησιμοποιούμενες επαφές, είτε έως 10 επαφές που έχει επιλέξει ο χρήστης πάνω από την αλφαβητική παραδοσιακή λίστα των επαφών.



Εικόνα 10. Η λειτουργία των top επαφών από το (Jung, Anttila & Blom, 2008). Από αριστερά προς τα δεξιά: αυτόματα επιλεγμένες top επαφές, επιλεγμένες από τον χρήστη top επαφές, ρυθμίσεις.



Εικόνα 11. Έξυπνες ομάδες επαφών που δείχνουν τις επαφές που ταιριάζουν σε συγκεκριμένες συνθήκες (Jung, Anttila & Blom, 2008)

Ακόμα ένα ενδιαφέρον χαρακτηριστικό του πρωτοτύπου ήταν μια οθόνη που παρουσίαζε ομάδες επαφών που είχαν δημιουργηθεί αυτόματα (smart groups) με βάση κοινά χαρακτηριστικά για τα οποία άξιζε οι χρήστες να ενημερωθούν (εικόνα 11). Οι ομάδες που υλοποιήθηκαν στο πρωτότυπο ήταν: δημοφιλείς επαφές σε επικοινωνία, επαφές που έχουν πρόσφατα γενέθλια και επαφές που προστέθηκαν πρόσφατα. Στόχος αυτής της διεπαφής ήταν να ξεχωρίσει πληροφορίες που πιθανόν να διαφεύγουν της προσοχής των χρηστών.

Αν και οι ερευνητές πραγματοποίησαν πείραμα αξιολόγησης του πρωτοτύπου τους, παραθέτουν ελάχιστα στοιχεία για τα αποτελέσματα που αφορούν τις παραπάνω λειτουργικότητες: συγκεκριμένα αναφέρουν πως οι “top” επαφές έτυχαν πολύ θετικής αποδοχής, ενώ η ομάδα των δημοφιλών επαφών επέτρεπε στους χρήστες, σύμφωνα με τα σχόλιά τους, να κατανοήσουν καλύτερα τη συμπεριφορά τους στην επικοινωνία.

Ένα σημαντικό κομμάτι της βιβλιογραφίας που αφορά εναλλακτικά πρωτότυπα για τις λίστες επαφών των συσκευών κινητού υπολογισμού εστιάζει στην προσθήκη πληροφορίας πλαισίου, όπως η κατάσταση (status) και η τοποθεσία των επαφών.

Μια από τις πρώτες προσπάθειες επαύξησης των επαφών με εικονίδια που δηλώνουν τη συνδεσιμότητα, τη διαθεσιμότητα για επικοινωνία και την τοποθεσία/δραστηριότητα παρουσιάζεται στο (Schmidt, Gellersen & Stuhr, 2001). Στο (Oulasvirta, Raento & Tiitta, 2005) παρουσιάζεται ένα βιβλίο επαφών (ContextContacts) που έχει επαυξηθεί με εικονίδια που προσφέρουν πληροφορίες για το πλαίσιο στο οποίο βρίσκονται οι επαφές (π.χ. προφίλ ήχου, αθόρυβο, τοποθεσία και χρόνος που έχει περάσει σε αυτή, πρόσφατη χρήση κινητού, ενδείξεις για την κοινωνική δραστηριότητα). Οι ερευνητές επιχείρησαν με αυτό το σχεδιασμό να λύσουν το πρόβλημα της επίγνωσης της κατάστασης των επαφών των χρηστών, η έλλειψη πληροφοριών για το πλαίσιο των οποίων οδηγεί σε υψηλό ποσοστό αποτυχημένων προσπαθειών επικοινωνίας ή σε συχνές αναφορές στην τοποθεσία στην αρχή μιας επικοινωνίας. Σε μια άλλη δημοσίευση, παρουσιάστηκε η εφαρμογή Friendlee (Ankolekar, Szabo, Luon, Huberman, Wilkinson & Wu, 2009), η οποία ανέλυε το ιστορικό επικοινωνίας μέσω κλήσεων και μηνυμάτων (συχνότητα, προσφατότητα και διάρκεια φωνητικής επικοινωνίας) για την εύρεση των επαφών με τις οποίες οι χρήστες έχουν κοντινούς και ισχυρούς κοινωνικούς δεσμούς. Οι επαφές αυτές παρουσιάζονταν σε περίοπτη θέση, επιτρέποντας στους χρήστες την άμεση ανάκτησή τους χωρίς να χρειάζεται να ψάξουν σε μια μακριά λίστα. Επιπλέον, η εφαρμογή επέτρεπε το διαμοιρασμό πληροφοριών πλαισίου (τοποθεσία, μήνυμα κατάστασης, προφίλ χρήσης συσκευής κ.α.) με αυτές τις επαφές. Ακόμα μια εφαρμογή που επέτρεπε την ανταλλαγή πληροφοριών σχετικά με το πλαίσιο των επαφών (τοποθεσία, χρόνος στην τρέχουσα τοποθεσία, προφίλ ήχου) ήταν το Connecto (Barkhuus, Brown, Bell, Sherwood, Hall & Chalmers, 2008). Και σε αυτή την περίπτωση η διεπαφή αποτελούσε μια λίστα επαφών που έχει επαυξηθεί με τις πληροφορίες πλαισίου.

Η πρώτη εμπορική εφαρμογή που αντικατέστησε την παραδοσιακή στατική λίστα επαφών με ένα κεντρικό σημείο επικοινωνίας ήταν το Motoblur της Motorola (Bentley, Kames, Ahmed, Zivin & Schwendimann, 2010). Στόχος αυτού του project ήταν να συνδυάσει την επικοινωνία που πραγματοποιείται στη συσκευή με την επικοινωνία από τις υπηρεσίες κοινωνικής δικτύωσης, επιτρέποντας τη διασύνδεση των επαφών της συσκευής με τις επαφές των χρηστών από τα κοινωνικά δίκτυα στα οποία συμμετέχουν. Ακόμα μια ερευνητική εργασία όπου επαυξάνεται η διεπαφή των επαφών με περιεχόμενο από υπηρεσίες κοινωνικής δικτύωσης παρουσιάστηκε στο (Cui & Honkala, 2013).

Εντούτοις, ενώ οι περισσότερες από τις προσπάθειες που παρουσιάστηκαν σε αυτή την ενότητα στοχεύουν στη δημιουργία επαυξημένων βιβλίων επαφών που εξυπηρετούν τις ανάγκες των χρηστών για επικοινωνία ή αυξάνουν την επίγνωση για το πλαίσιο στο οποίο βρίσκονται, καμία από αυτές δε σχεδιάστηκε με γνώμονα τη διευκόλυνση της ανάκτησης των επαφών.

2.3.4 Άλλες τεχνικές για τη βελτίωση της ανάκτησης των επαφών

Εκτός από τις τροποποιήσεις στο UI του βιβλίου επαφών έχουν προταθεί και άλλοι τρόποι για τη βελτίωση της οργάνωσης των επαφών με σκοπό την καλύτερη διαχείριση και ανάκτησή τους.

Στο (Le, Henze & Schmidt, 2013) αναφέρεται ότι η ταξινόμηση των επαφών σε ομάδες (groups) μπορεί να βελτιώσει την ανάκτησή τους. Παρότι οι συσκευές κινητού υπολογισμού παρέχουν τη δυνατότητα δημιουργίας ομάδων και ταξινόμησης των επαφών σε αυτές, οι χρήστες δε χρησιμοποιούν αυτή τη λειτουργικότητα, πιθανότητα επειδή δε θεωρούν ότι έχει αξία να κατηγοριοποιήσουν τις εκατοντάδες επαφές τους. Οι ερευνητές πρότειναν ένα μοντέλο μηχανικής μάθησης που αναλύει το ιστορικό επικοινωνίας (κλήσεις και μηνύματα) του χρήστη και κατηγοριοποιεί αυτόματα τις επαφές σε 5 κατηγορίες (οικογένεια, φίλοι, γνωστοί, δουλειά, άλλοι). Στα πειράματά τους βρήκαν ότι το μοντέλο έχει αντικειμενική ακρίβεια περίπου 59%, ενώ όταν οι χρήστες αξιολογούν την ταξινόμηση η ακρίβεια αυξάνεται σε περίπου 75%.

Η αποτελεσματικότητα της ταξινόμησης των επαφών σε ομάδες μελετήθηκε σε μια έρευνα που παρουσιάστηκε στο (Nguyen & Oh, 2010). Το 52% των χρηστών που συμμετείχαν στην έρευνα δήλωσε ότι χρησιμοποιεί τις ομάδες για την οργάνωση των επαφών, σε αντίθεση με όσα ανέφεραν οι Le, Henze και Schmidt (2013). Εντούτοις, οι περισσότεροι συμμετέχοντες δήλωσαν ότι η διαδικασία οργάνωσης του βιβλίου επαφών σε ομάδες είναι ενοχλητική. Το 63% δεν μπορούσε να αποφασίσει σε ποια

ομάδα να τοποθετήσει μια επαφή, ενώ το 12% δεν μπορούσε να βρει αντιπροσωπευτικό όνομα για μια ομάδα και επίσης 12% δεν μπορούσε να πραγματοποιήσει αναζήτηση για επαφές χρησιμοποιώντας τα ονόματα των ομάδων. Σε ερωτήσεις ανοιχτού τύπου κάποιοι συμμετέχοντες δήλωσαν ότι θα ήθελαν να μπορούν να κατηγοριοποιούν μια επαφή σε περισσότερες από μια ομάδες, να υπάρχει η δυνατότητα για ιεραρχικές ομάδες και να πραγματοποιούν αναζήτηση με το όνομα της ομάδας. Οι ερευνητές επικεντρώθηκαν στη διαχείριση επαφών που αντιπροσωπεύουν επιχειρήσεις (εστιατόρια, καφετέριες κ.λπ.), οι οποίες χρησιμοποιούνται συνήθως σε όχι συχνή βάση. Ενώ σχεδόν οι μισοί χρήστες χρειάζεται να καλέσουν μια τέτοια επαφή τουλάχιστον μια φορά το μήνα και τα $\frac{3}{4}$ τουλάχιστον μια φορά στους 3 μήνες, συνήθως δε θυμούνται αν την έχουν αποθηκεύσει στο βιβλίο επαφών και με ποιο όνομα, οπότε καταλήγουν να εξετάζουν ολόκληρη τη λίστα επαφών. Οι ερευνητές, για την επίλυση αυτών των προβλημάτων πρότειναν ένα πρωτότυπο (Trung & Oh, 2010) που επέτρεπε το χαρακτηρισμό των επαφών με ετικέτες, το διαμοιρασμό των ετικετών με άλλους χρήστες και διευκόλυνε την αναζήτηση, επιτρέποντας ερωτήματα με κριτήριο τις ετικέτες. Μια προκαταρκτική αξιολόγηση του πρωτοτύπου, που σχεδιάστηκε σε χαρτί και όχι ως πραγματική εφαρμογή, έδειξε θετική ανταπόκριση των χρηστών στα προσφερόμενα χαρακτηριστικά.

2.4 Προσαρμοζόμενες λίστες επαφών και πρόβλεψη κλήσεων

Διάφορες προσεγγίσεις έχουν προταθεί για τη βελτίωση των διεπαφών των συσκευών κινητού υπολογισμού και ειδικότερα για τη διευκόλυνση της πρόσβασης στις λειτουργικότητες που προσφέρουν, όπως η ανάκτηση των επαφών. Μια τέτοια προσέγγιση είναι και ο σχεδιασμός προσαρμοζόμενων διεπαφών (adaptive user interfaces). Η έννοια των προσαρμοζόμενων διεπαφών ορίζεται στο (Langley, 1999) ως εξής: «Μια προσαρμοζόμενη διεπαφή χρήστη είναι ένα τεχνούργημα λογισμικού που βελτιώνει την ικανότητά του να αλληλεπιδρά με το χρήστη δημιουργώντας ένα μοντέλο χρήσης που βασίζεται στη μερική εμπειρία με αυτόν τον χρήστη». Στην περίπτωση των επαφών, μια προσαρμοζόμενη διεπαφή είναι συνήθως μια δυναμική λίστα επαφών που δημιουργείται με κάποια κριτήρια κατάταξης. Οι επαφές παρουσιάζονται με σειρά φθίνουσας πιθανότητας να χρησιμοποιηθούν στο τρέχον πλαίσιο, οπότε ο χρήστης δε χρειάζεται να πραγματοποιήσει αναζήτηση, αλλά είναι πιθανό να βρει την επαφή που τον ενδιαφέρει στις κορυφαίες θέσεις της λίστας. Με αυτόν τον τρόπο μειώνεται ο χρόνος ανάκτησης και τα πατήματα πλήκτρων που απαιτούνται (χρειάζονται μόνο δύο πατήματα: ένα για την πρόσβαση στη λίστα και ένα για την πραγματοποίηση της κλήσης). Για τη δημιουργία της προσαρμοζόμενης

λίστας χρειάζεται κάποιος αλγόριθμος κατάταξης ή αλλιώς πρόβλεψης για το ποια επαφή είναι πιο πιθανό να χρησιμοποιηθεί. Στη συνέχεια παρουσιάζονται ερευνητικές εργασίες όπου μελετώνται προσαρμοζόμενες λίστες επαφών και αλγόριθμοι πρόβλεψης κλήσεων.

Στο (Barzaïq & Loke, 2011) παρουσιάστηκε η ATF (adaptive task-based functionality), μια προσαρμοζόμενη λειτουργικότητα που αφορά εργασίες (tasks), όπου η εργασία στην οποία εστίασαν οι ερευνητές ήταν η πρόβλεψη της επόμενης επαφής που ο χρήστης είναι πιθανό να καλέσει. Το μοντέλο που παρουσιάστηκε θα μπορούσε να προσαρμοστεί και σε άλλες εργασίες (π.χ. στην πρόβλεψη της εφαρμογής που είναι πιο πιθανό να χρησιμοποιήσει ο χρήστης). Το πρωτότυπο περιελάμβανε μια δυναμική λίστα επαφών που δημιουργείται μετά από την ανάλυση του πλαισίου που προκύπτει από το ιστορικό των εξερχόμενων κλήσεων του χρήστη. Τα κριτήρια που χρησιμοποιήθηκαν για την κατάταξη των επαφών ήταν η συχνότητα και η τακτικότητα της επικοινωνίας. Ο σχεδιασμός του πρωτοτύπου βασίστηκε στη θεωρία Multiple Attribute Utility Theory (Butler, Morrice & Mullarkey, 2001) και για τον υπολογισμό του σκορ κάθε επαφής που θα χρησιμοποιηθεί για την κατάταξή της πρέπει να απαντηθούν τα παρακάτω ερωτήματα:

- Πόσες εγγραφές που αφορούν την επαφή υπάρχουν στο σύνολο δεδομένων;
- Έχει κληθεί τουλάχιστον μια φορά κάθε μήνα;
- Έχει κληθεί τουλάχιστον μια φορά κάθε μέρα;

Οι τρεις αυτές ερωτήσεις σηματοδοτούν τις διαστάσεις του αλγορίθμου πρόβλεψης και αντιπροσωπεύουν τη συχνότητα και την τακτικότητα της επικοινωνίας. Ο αλγόριθμος αναλύει τις επαφές των δύο τελευταίων ετών, αφαιρώντας όσες είναι πιο παλιές. Μια τέτοια επιλογή δεν είναι ιδιαίτερα ρεαλιστική, καθώς συνήθως στις συσκευές αποθηκεύεται ένας συγκεκριμένος αριθμός από κλήσεις. Για παράδειγμα στις περισσότερες συσκευές με το λειτουργικό σύστημα Android διατηρούνται οι 500 τελευταίες κλήσεις (εισερχόμενες και εξερχόμενες), αριθμός που συνήθως αντιστοιχεί σε μικρότερο χρονικό διάστημα. Επιπλέον, στα (Phithakkitnukoon & Dantu, 2008) και (Phithakkitnukoon, Dantu, Claxton & Eagle, 2011) αποδείχθηκε ότι για την πρόβλεψη της μελλοντικής συμπεριφοράς των χρηστών ως προς την επικοινωνία απαιτείται ανάλυση ενός πρόσφατου μόνο τμήματος του ιστορικού κλήσεων. Τελικά, ο αλγόριθμος υπολογίζει ένα σκορ για κάθε επαφή που χρησιμοποιείται για την ταξινόμησή τους από τις παρακάτω διαστάσεις:

- Μετρητής – C(X): αντιστοιχεί στη συχνότητα επικοινωνίας και υπολογίζεται ως το πλήθος των εξερχόμενων κλήσεων με την επαφή προς το συνολικό πλήθος εξερχόμενων κλήσεων. Σε αυτή τη διάσταση αντιστοιχεί βάρος 20%.
- Κάθε μήνα – EM(X): υπολογίζεται ως το πλήθος των μηνών για τους οποίους έχει πραγματοποιηθεί μια εξερχόμενη κλήση με την επαφή προς τους συνολικούς μήνες (24 για διάστημα δύο ετών). Σε αυτή τη διάσταση δίνεται βάρος 30%.
- Κάθε μέρα – ED(X): υπολογίζεται ως το πλήθος των ημερών για τις οποίες έχει πραγματοποιηθεί μια εξερχόμενη κλήση με την επαφή προς τις συνολικές ημέρες (730 για διάστημα δύο ετών). Σε αυτή τη διάσταση δίνεται βάρος 50%.

Το συνολικό σκορ υπολογίζεται ως:

$$S(X) = C(X) * W_C + EM(X) * W_{EM} + ED(X) * W_{ED}$$

όπου τα βάρη για κάθε διάσταση είναι αυτά που αναφέρθηκαν παραπάνω. Η επιλογή των συγκεκριμένων βαρών είναι αμφιλεγόμενη, καθώς έγινε αυθαίρετα και όχι ύστερα από κάποια ανάλυση ή μελέτη. Το πρωτότυπο περιελάμβανε τρεις διεπαφές για την πραγματοποίηση μιας κλήσης: μια δυναμική λίστα που προκύπτει μόνο από τη συχνότητα επικοινωνίας (συχνότερα χρησιμοποιούμενες επαφές), τη δυναμική λίστα με βάση τη μέθοδο που παρουσιάστηκε και τη λίστα με όλες τις επαφές.

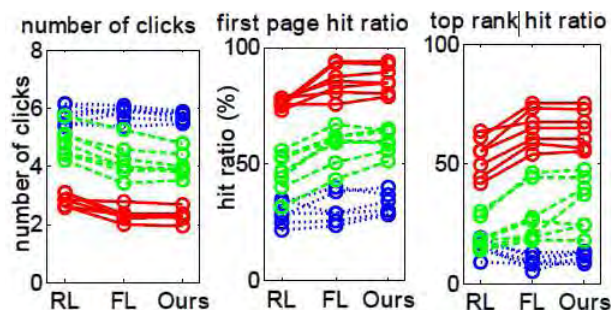
Οι ερευνητές προκειμένου να μελετήσουν τη βελτίωση του αλγορίθμου με την πάροδο του χρόνου πραγματοποίησαν αξιολόγηση με έναν χρήστη για 5 εβδομάδες. Η ακρίβεια του αλγορίθμου (κλήσεις από τη διεπαφή/συνολικές κλήσεις) αυξήθηκε από 10% την πρώτη εβδομάδα σε 35% την πέμπτη, εντούτοις τα αποτελέσματα δεν μπορούν να ληφθούν σοβαρά υπόψη λόγω του ελάχιστου πλήθους του δείγματος. Οι ερευνητές, ισχυρίζονται πως ο συνδυασμός διαστάσεων βελτιώνει τα αποτελέσματα πρόβλεψης, όμως πραγματοποίησαν τα πειράματα που τους οδήγησαν σε αυτό το συμπέρασμα σε συνθετικό σύνολο δεδομένων και όχι σε πραγματικό. Τέλος, πραγματοποίησαν αξιολόγηση του αλγορίθμου χρησιμοποιώντας ένα δημόσια διαθέσιμο σύνολο δεδομένων που έχει συλλεχθεί από χρήστες κινητών τηλεφώνων στα πλαίσια του Reality Mining Project (Eagle & Pentland, 2006), επαληθεύοντας την παρατήρηση για βελτίωση της αποτελεσματικότητας του αλγορίθμου με την πάροδο του χρόνου. Το συγκεκριμένο σύνολο περιλαμβάνει επικοινωνία μόνο για 9 μήνες, κάτι που δεν είναι συμβατό με την απαίτησή του αλγορίθμου για ανάλυση δεδομένων δύο ετών. Επιπλέον, η πρόβλεψη πραγματοποιήθηκε για όλες τις εξερχόμενες κλήσεις που έγιναν την πρώτη μέρα κάθε μήνα, μια επιλογή που δεν επιτρέπει την άμεση σύγκριση των αποτελεσμάτων ανά μήνα μεταξύ τους. Τέλος, οι ερευνητές πρότειναν

την προσθήκη και άλλων διαστάσεων (αναγνωρίζοντας την τοποθεσία ως μια σημαντική διάσταση που επηρεάζει κάτω από συγκεκριμένες καταστάσεις την επικοινωνία των χρηστών), επιπλέον χαρακτηριστικών (π.χ. διάρκεια κλήσεων) στις υπάρχουσες διαστάσεις, καθώς και την αξιοποίηση των εισερχόμενων κλήσεων στη διαδικασία πρόβλεψης.

Ένας αλγόριθμος για τη δημιουργία προσαρμοζόμενης λίστας γρήγορης κλήσης επαφών παρουσιάστηκε στο (Lee, Seo & Lee, 2010). Ο αλγόριθμος αναλύει τις εξερχόμενες κλήσεις των χρηστών και χρησιμοποιεί 5 διαστάσεις για την ταξινόμηση των επαφών (πίνακας 2).

Πίνακας 2. Πιθανές τιμές για τις διαστάσεις που χρησιμοποιούνται για την πρόβλεψη στον αλγόριθμο των Lee, Seo και Lee (2010)

Μεταβλητές	Τιμές
Μέρα της εβδομάδας	Δευτέρα - Κυριακή
Σαββατοκύριακο / Καθημερινή	Σαββατοκύριακο, Καθημερινή
Ώρα της ημέρας	Περίοδοι 3 ^{ων} ωρών (τρέχουσα ώρα $\pm$ 1.5 ώρα)
Τμήμα της ημέρας	Πρωί (6-12), Απόγευμα (12-17), Βράδυ (17-22), Νύχτα (22-6)
Ωριαία διαστήματα (μιας ημέρας)	Κάθε ώρα



Εικόνα 12. Συγκρίσεις των μέσων ποσοστών επιτυχίας από την προσομοίωση των Lee, Seo & Lee (2010)

Για κάθε επαφή, ο αλγόριθμος υπολογίζει τις πιθανότητες των δοκιμών Bernoulli για κάθε διάσταση ανεξάρτητα και χρησιμοποιεί τη μεγαλύτερη από αυτές ως σκορ της επαφής για την ταξινόμηση. Οι ερευνητές πραγματοποίησαν αξιολόγηση της απόδοσης του αλγορίθμου μέσω προσομοίωσης χρήσης της δυναμικής λίστας 5 θέσεων για όλες τις εξερχόμενες κλήσεις ενός μήνα από πραγματικά αρχεία κλήσεων, αναλύοντας δεδομένα από τους προηγούμενους τρεις μήνες και συνέκριναν τα αποτελέσματα με τις λίστες πρόσφατων κλήσεων (RL) και πιο συχνά χρησιμοποιούμενων επαφών (FL) (εικόνα 12).

Το μέσο ποσοστό επιτυχούς πρόβλεψης της επαφής που κλήθηκε στην πρώτη οθόνη της λίστας ήταν 51,8%, 60,8% και 62,5% και το μέσο ποσοστό επιτυχίας για την πρώτη θέση της λίστας ήταν 30,3%, 36,6% και 39,3% για τη λίστα πρόσφατων κλήσεων, συχνών επαφών και την προσαρμοζόμενη με βάση τον αλγόριθμο λίστα, αντίστοιχα (όλες οι διαφορές ανάμεσα στους μέσους όρους ήταν στατιστικά σημαντικές). Από την ανάλυση φάνηκε πως υπάρχουν τρεις ομάδες χρηστών με διαφορετικά πρότυπα επικοινωνίας και ο αλγόριθμος παρήγαγε καλά αποτελέσματα για τη μια ομάδα μόνο (κόκκινη ομάδα στην εικόνα 12). Κάποιοι χρήστες συγκεντρώνουν το μεγαλύτερο μέρος των κλήσεων τους σε λίγες επαφές (το 70% των κλήσεων με τις 3 πιο συχνά χρησιμοποιούμενες επαφές για την κόκκινη ομάδα χρηστών), ενώ άλλοι τις κατανέμουν σε περισσότερες επαφές (λιγότερο από 20% και 40% των κλήσεων με την μια και τρεις πιο συχνά χρησιμοποιούμενες επαφές για τη μπλε ομάδα χρηστών) και επομένως η πρόβλεψη της συμπεριφοράς τους είναι δυσκολότερη.

Στη συνέχεια, πραγματοποιήθηκε αξιολόγηση στο πεδίο μιας διεπαφής που υλοποιούσε την προσαρμοζόμενη λίστα επαφών, στην οποία συμμετείχαν 10 χρήστες για τρεις μήνες. Συνολικά πραγματοποιήθηκαν 427 κλήσεις χρησιμοποιώντας τη διεπαφή, ενώ η μέση κατάταξη της καλούμενης επαφής στη λίστα ήταν 2,35. Η διεπαφή απέτυχε να προτείνει επιτυχώς την καλούμενη επαφή σε 161 περιπτώσεις, στις 101 από τις οποίες ο καλούμενος αριθμός δε βρισκόταν στο αρχείο πρόσφατων κλήσεων. Η μέση κατάταξη για τις επαφές στις περιπτώσεις αποτυχίας του αλγορίθμου (εξαιρώντας επαφές που κλήθηκαν για πρώτη φορά) ήταν 9,27 (διάμεσος 3). Οι χρήστες μετά από κάθε κλήση παρείχαν ανατροφοδότηση: στο 63,5% των περιπτώσεων βρήκαν τις προτάσεις χρήσιμες, στο 29,1% όχι χρήσιμες και στο 7,4% των περιπτώσεων δεν απάντησαν. Μια σημαντική παράμετρος που αποκαλύφθηκε από την ποιοτική έρευνα μετά το τέλος του πειράματος ήταν πως οι χρήστες αξιολόγησαν ως σημαντικότερη απαίτηση για μια τέτοια διεπαφή την αποτελεσματική μετάβαση σε έναν εναλλακτικό τρόπο ανάκτησης (π.χ. αναζήτηση ή λίστα πρόσφατων κλήσεων) σε περίπτωση αποτυχίας του αλγορίθμου πρόβλεψης από την ακρίβεια της πρόβλεψης.

Ακόμα μια προσέγγιση για την πρόβλεψη των εξερχόμενων κλήσεων που πραγματοποιούν οι χρήστες των έξυπνων κινητών τηλεφώνων με βάση την ανάλυση των προτύπων συμπεριφοράς του παρελθόντος παρουσιάστηκε στο (Kang, Won, Jeon & Lee, 2012). Σε αυτή την περίπτωση, προτάθηκε η στατιστική ανάλυση των δεδομένων κλήσεων, καθώς και η μελέτη των προτύπων επικοινωνίας. Η στατιστική ανάλυση περιλαμβάνει την ανάλυση συχνότητας των κλήσεων συνολικά, αλλά και με

βάση διάφορους παράγοντες όπως η ημέρα του μήνα, η ημέρα της εβδομάδας, η ώρα, το λεπτό κ.α. Η μελέτη των προτύπων αναγνωρίζει πρότυπα όπως ενδεχομένως το γεγονός ότι η επικοινωνία με μια επαφή προηγείται ή έπεται της επικοινωνίας με συγκεκριμένες άλλες επαφές. Το μοντέλο δεν περιγράφεται επακριβώς με όλες τις λεπτομέρειες υλοποίησης, αλλά μόνο σχετικά αφαιρετικά. Επιπλέον, αναφέρεται ότι εξήχθησαν τα δεδομένα κλήσεων από ένα σύνολο χρηστών (1.106 εγγραφές από τις οποίες οι τελευταίες 100 χρησιμοποιήθηκαν για τις προβλέψεις), αλλά παρουσιάζονται τα αποτελέσματα των προβλέψεων για 5 χρήστες (κατά μέσο όρο 85% περίπου επιτυχία, χωρίς όμως να διευκρινίζεται πόσες θέσεις πρόβλεψης χρησιμοποιήθηκαν), ενώ αφήνεται να εννοηθεί ότι αυτός δεν είναι ο συνολικός αριθμός συμμετεχόντων, χωρίς όμως να παρέχεται εξήγηση για τον τρόπο που επιλέχθηκαν.

Μια διαφορετική προσέγγιση παρουσιάστηκε στο (Min, Kim & Cho, 2008), όπου μοντελοποιήθηκε το κοινωνικό (φιλία) και προσωπικό πλαίσιο (συναισθηματική κατάσταση, βαθμός δραστηριότητας) του χρήστη και χρησιμοποιήθηκε για την παροχή προτάσεων χρησιμοποίησης υπηρεσιών, ενώ οι ερευνητές παρουσίασαν εφαρμογή του μοντέλου τους για τη δημιουργία μιας λίστας προτεινόμενων επαφών προς επικοινωνία. Η προσέγγιση βασίστηκε στην ανάλυση δεδομένων από τα αρχεία που καταγράφονται στις συσκευές κινητού υπολογισμού (ώρα, ημέρα της εβδομάδας, τοποθεσία, τύπος προγράμματος, ομάδα βιβλίου επαφών, γενέθλια, πλήθος πρόσφατων κλήσεων, πλήθος πρόσφατων μηνυμάτων) και χρησιμοποιεί δίκτυα Bayes (οι παράμετροι των οποίων ορίστηκαν από τους συγγραφείς με βάση τη γνώση του πεδίου που έχουν) για την εξαγωγή του υψηλού επιπέδου πλαισίου του χρήστη. Η ώρα, η ημέρα της εβδομάδας, η τοποθεσία και το πρόγραμμα χρησιμοποιήθηκαν για τη μοντελοποίηση της τάσης χρησιμοποίησης της υπηρεσίας από τον χρήστη. Το πλαίσιο επεκτείνεται σε ένα δέντρο πλαισίου (context tree), ενώ για την παραγωγή προτάσεων συγκρίνεται το τρέχον πλαίσιο με το ιστορικό των πλαισίων που έχει καταγραφεί. Οι ερευνητές παρουσίασαν μια αξιολόγηση του μοντέλου τους με δεδομένα που εξήγαγαν από έναν μόνο χρήστη για διάστημα 12 ημερών (συνολικά καταγράφηκαν μόνο 42 εισερχόμενες, 23 εξερχόμενες και 2 αναπάντητες κλήσεις). Αν και δεν μπορούν να βγουν ασφαλή συμπεράσματα, το μοντέλο είχε ακρίβεια πρόβλεψης περίπου 26% για μια θέση πρόβλεψης, περίπου 50% για 3 θέσεις και 70% για 5 θέσεις, ενώ απέτυχε να προβλέψει το 22% των κλήσεων επειδή πραγματοποιήθηκαν για πρώτη φορά με αυτές τις επαφές και δεν υπήρχαν ιστορικά στοιχεία.

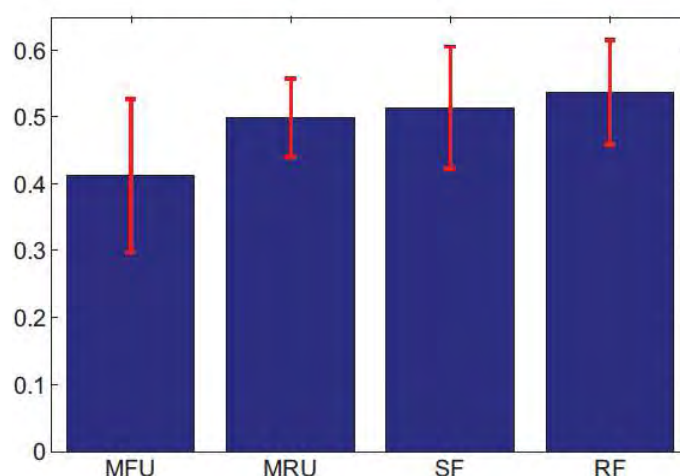
Οι Sun, Wang, Zheng και Hsu (2013) εστίασαν στην μίξη των χαρακτηριστικών (feature fusion) που αναπαριστούν το πλαίσιο του χρήστη, διακρίνοντας δύο διαφορετικές τεχνικές: μίξη με βάση το σκορ (SF – score based fusion) και μίξη με βάση την κατάταξη (RF – rank based fusion). Σε αυτή την περίπτωση, το πρόβλημα της δημιουργίας λιστών με προτάσεις αντιμετωπίζεται ως πρόβλημα κατάταξης και επιλογής. Έστω ένα σύνολο $I=\{I_i | i=(1,2,...,n)\}$ αντικειμένων προς ταξινόμηση. Ο χώρος των χαρακτηριστικών ορίζεται για το σύνολο των αντικειμένων ως $F=\{f_i | i=(1,2,...,m)\}$. Η συνάρτηση σκορ ενός χαρακτηριστικού f_i για το I μπορεί να οριστεί ως $S_{fi}(I):I \rightarrow R$. Τα αντικείμενα μπορούν να ταξινομηθούν με τη χρήση μιας συνάρτησης ταξινόμησης με βάση τα σκορ για το χαρακτηριστικό f_i , που μπορεί να οριστεί ως $R(S_{fi}(I)):I \rightarrow N$ όπου $N=(1,2,...,n)$. Καθώς τα σκορ των χαρακτηριστικών είναι σε διαφορετικές κλίμακες, η εφαρμογή κανονικοποίησης είναι απαραίτητη. Στη μίξη με βάση το σκορ, το τελικό σκορ κάθε αντικειμένου προκύπτει από το μέσο όρο των κανονικοποιημένων σκορ των επιμέρους m χαρακτηριστικών. Στη μίξη με βάση την κατάταξη, το τελικό σκορ κάθε αντικειμένου προκύπτει από το μέσο όρο των τιμών της συνάρτησης κατάταξης για τα m χαρακτηριστικά. Και στις δύο περιπτώσεις τα αντικείμενα κατατάσσονται σύμφωνα με τα συνολικά τους σκορ και επιλέγονται αυτά που θα προταθούν από την προσαρμοζόμενη διεπαφή.

Οι ερευνητές εφάρμοσαν την τεχνική της μίξης των χαρακτηριστικών για την πρόβλεψη των εφαρμογών που είναι πιο πιθανό να χρησιμοποιηθούν και των επαφών που είναι πιο πιθανό να κληθούν σε ένα έξυπνο κινητό τηλέφωνο. Για την περίπτωση των επαφών βασίστηκαν στα αποτελέσματα (Stefanis, Plessas, Komninos & Garofalakis, 2012) που παρουσιάζονται στην ενότητα 3.2 της διατριβής και πρότειναν ως χαρακτηριστικά τη συχνότητα και προσφατότητα της επικοινωνίας, αναλύοντας τις εξερχόμενες μόνο κλήσεις. Η μίξη με βάση την κατάταξη προσφέρει καλύτερα αποτελέσματα όταν τα δύο χαρακτηριστικά έχουν υψηλή απόδοση και παρουσιάζουν έντονη ποικιλομορφία (Hsu & Taksu, 2005), κάτι που εντόπισαν να συμβαίνει για τη συχνότητα και προσφατότητα της επικοινωνίας και επομένως θεώρησαν ότι αυτή η τεχνική είναι πιο κατάλληλη για την περίπτωση της δυναμικής λίστας επαφών. Στο πείραμα αξιολόγησης που πραγματοποιήθηκε, χρησιμοποιήθηκε ένα σύνολο δεδομένων που προέκυψε από τη συγκέντρωση δεδομένων από 25 χρήστες iPhone συσκευών για ένα έτος (πίνακας 3).

Πίνακας 3. Στατιστικά του συνόλου δεδομένων που χρησιμοποιήθηκε στο (Sun, Wang, Zheng & Hsu, 2013) για τη λίστα επαφών

	Συνολικός αριθμός επαφών	Συνολικός αριθμός εξερχόμενων κλήσεων
Μέγιστη τιμή	452	2601
Ελάχιστη τιμή	73	321
Μέσος όρος	196.4	1414.7

Οι πρώτες 100 κλήσεις χρησιμοποιήθηκαν ως ιστορικό και για τις υπόλοιπες πραγματοποιήθηκαν προβλέψεις, προσομοιώνοντας μια δυναμική λίστα επαφών με 4 θέσεις. Τα αποτελέσματα παρουσιάζονται στην εικόνα 13.



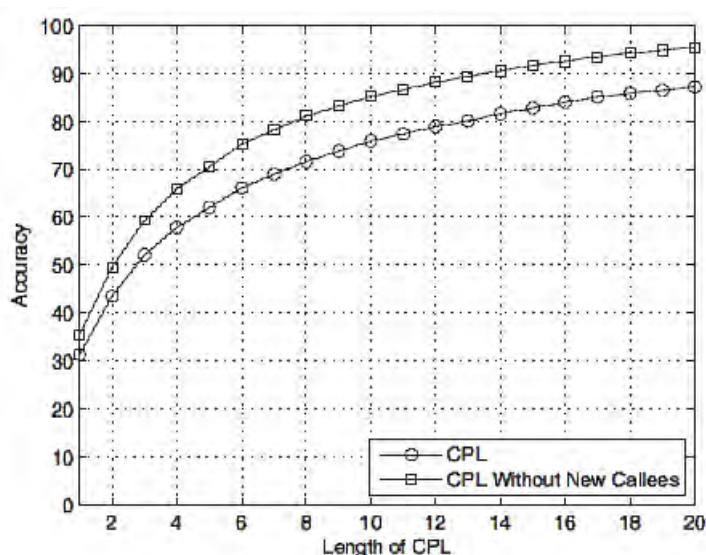
Εικόνα 13. Απόδοση μίξης χαρακτηριστικών ως προς την επιτυχία πρόβλεψης. MFU=συχνά χρησιμοποιούμενες επαφές, MRU=πιο πρόσφατα χρησιμοποιούμενες επαφές, SF=μίξη με βάση το σκορ, RF=μίξη με βάση την κατάταξη

Όταν η πρόβλεψη έγινε με τη χρήση των παραδοσιακών τρόπων ανάκτησης επαφών, όπως οι πιο συχνά χρησιμοποιούμενες επαφές (MFU – συχνότητα επικοινωνίας) ή οι πιο πρόσφατες επαφές (MRU – προσφατότητα επικοινωνίας), τα ποσοστά επιτυχίας των προβλέψεων ήταν σχετικά χαμηλά (41,2% και 49,9% αντίστοιχα). Η μίξη των δύο χαρακτηριστικών είχε ως αποτέλεσμα τη βελτίωση των αποτελεσμάτων πρόβλεψης. Όταν χρησιμοποιήθηκε η μίξη με βάση την κατάταξη (53,7%) η ακρίβεια πρόβλεψης ήταν κατά 2% υψηλότερη σε σχέση με τη μίξη με βάση το σκορ (51,5%), αν και δεν παρουσιάστηκε στατιστική ανάλυση για τη σημαντικότητα των διαφορών αυτών.

Η ιδέα ότι η συμπεριφορά των χρηστών ακολουθεί επαναλαμβανόμενα πρότυπα και επομένως από την ανάλυση της συμπεριφοράς του παρελθόντος μπορεί να προβλεφθεί η μελλοντική, παρουσιάστηκε στο (Phithakkitnukoon & Dantu, 2011). Πιο συγκεκριμένα, παρουσιάστηκε ένα μοντέλο ανάλυσης των αρχείων κλήσεων των κινητών συσκευών και πρόβλεψης των εισερχόμενων και εξερχόμενων κλήσεων. Οι ερευνητές υποστήριξαν ότι ο τρόπος με τον οποίο πραγματοποιούνται οι κλήσεις

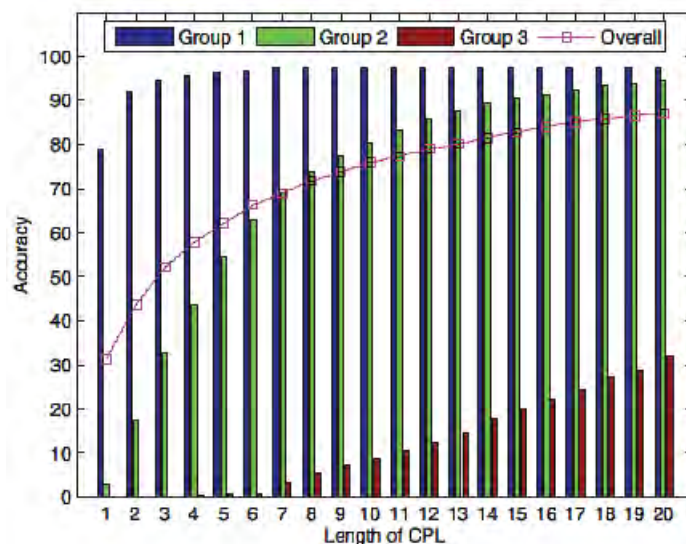
παρουσιάζει ένα μοναδικό πρότυπο, το οποίο μπορεί να παρατηρηθεί μέσα από διάφορους παράγοντες:

- η ώρα της κλήσης: υπάρχουν επαφές με τις οποίες επικοινωνούμε συνήθως συγκεκριμένες χρονικές περιόδους της ημέρας.
- η ημέρα της εβδομάδας: ενδεχομένως με κάποιες επαφές να επικοινωνούμε συγκεκριμένες μέρες της εβδομάδας, όπως για παράδειγμα με κάποιον φίλο που έχει το ρεπό του κάθε Τρίτη.
- η συχνότητα της επικοινωνίας: επαφές με τις οποίες επικοινωνούμε πιο συχνά, ανεξάρτητα από την ώρα και την ημέρα, είναι πιο πιθανό να χρειαστεί να ξαναεπικοινωνήσουμε.
- η αμοιβαιότητα: μερικές φορές οι χρήστες αναμένουν ή πραγματοποιούν κλήσεις με βάση τις τελευταίες επικοινωνίες με κάποια επαφή. Για παράδειγμα, μπορεί η τελευταία συζήτηση να ήταν «κάλεσέ με σε μια ώρα» ή «θα σε πάρω αύριο το πρωί να τα πούμε».



Εικόνα 14. Απόδοση του συστήματος πρόβλεψης εξερχόμενων κλήσεων (CPL) με και χωρίς τους για πρώτη φορά κληθέντες (Phithakkitnukoon & Dantu, 2011)

Με βάση τις παραπάνω παρατηρήσεις ανέπτυξαν ένα μοντέλο που βασίζεται σε έναν απλοϊκό ταξινομητή Bayes και χρησιμοποιεί την ώρα της ημέρας, την ημέρα της εβδομάδας, τη συχνότητα της επικοινωνίας και μια αναπαράσταση της αμοιβαιότητας για να προβλέψει τις εισερχόμενες ή τις εξερχόμενες κλήσεις (αναλύοντας το αντίστοιχο τμήμα του αρχείου κλήσεων σε κάθε περίπτωση).



Εικόνα 15. Ακρίβεια του συστήματος πρόβλεψης εξερχόμενων κλήσεων (CPL) για διαφορετικά μήκη της λίστας προτάσεων και διαφορετικές κοινωνικές ομάδες (Phithakkitnukoon & Dantu, 2011)

Η ακρίβεια του μοντέλου αξιολογήθηκε μέσω προσομοίωσης προβλέψεων σε πραγματικά δεδομένα κλήσεων που εξήχθησαν από 30 χρήστες. Στην εικόνα 14 εμφανίζονται τα ποσοστά επιτυχίας του μοντέλου για όλες τις επαφές και για επαφές με τις οποίες έχει γίνει έστω μια κλήση (στην αντίθετη περίπτωση δεν μπορεί να γίνει επιτυχής πρόβλεψη μόνο από το πλαίσιο των αρχείων κλήσεων), για διάφορα μεγέθη της λίστας προτάσεων. Όταν δε λαμβάνονται υπόψη οι κλήσεις προς επαφές με τις οποίες δεν υπάρχει προηγούμενη επικοινωνία, παρατηρήθηκε μια βελτίωση περίπου 10% στις προβλέψεις. Επιπλέον, οι ερευνητές μελέτησαν την επίπτωση που έχουν οι κοινωνικοί δεσμοί στην απόδοση του μοντέλου. Καθώς η κοινωνική απόσταση ανάμεσα στους ανθρώπους καθορίζεται από δείκτες όπως ο χρόνος και η ένταση των αλληλεπιδράσεων της σχέσης, χρησιμοποίησαν τη διάρκεια κλήσης και τη συχνότητα της επικοινωνίας για να διακρίνουν τρεις ομάδες επαφών:

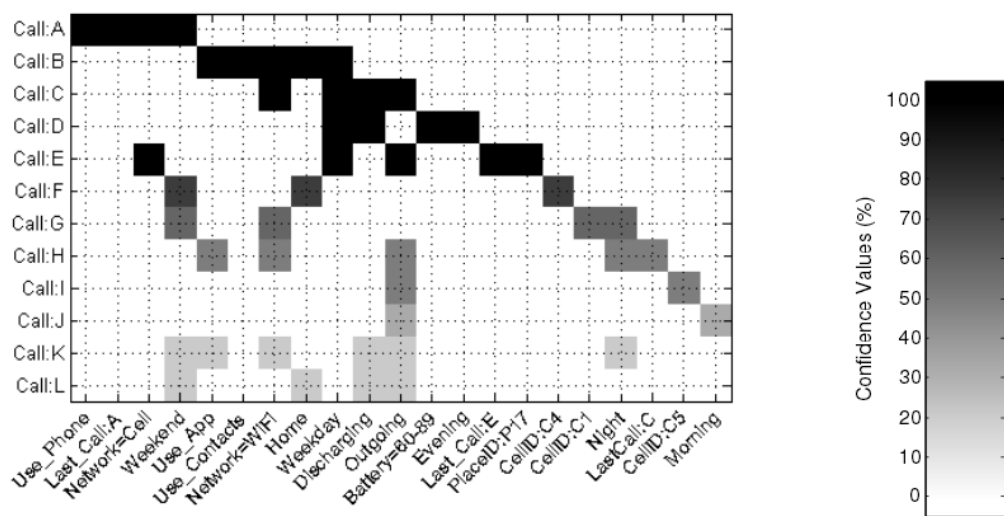
- Κοινωνικά κοντινότερες επαφές (Group 1): άτομα με τα οποία οι χρήστες διατηρούν την κοντινότερη κοινωνικά διασύνδεση. Συνήθως πρόκειται για μέλη της οικογένειας, φίλους και συναδέλφους. Με αυτά τα άτομα ανταλλάσσουν τις περισσότερες κλήσεις και μιλάνε για μεγαλύτερη διάρκεια.
- Κοινωνικά κοντινές επαφές (Group 2): Άτομα με τα οποία οι χρήστες δεν είναι τόσο κοντά όσο αυτά της προηγούμενης κατηγορίας, αλλά όταν συνδέονται μαζί τους επικοινωνούν για σχετικά μεγάλες χρονικές περιόδους. Πρόκειται συνήθως για μακρινούς συγγενείς και γείτονες.
- Κοινωνικά μακρινές επαφές (Group 3): Άτομα με τα οποία οι χρήστες επικοινωνούν πολύ λιγότερο και έχουν μικρότερη κοινωνική σύνδεση.

Στην εικόνα 15 παρουσιάζονται τα αποτελέσματα των προβλέψεων του μοντέλου ανά ομάδα. Όπως αναμενόταν, η πρόβλεψη ήταν πολύ πιο ακριβής για τις επαφές που είναι κοινωνικά πιο κοντά στη ζωή των χρηστών.

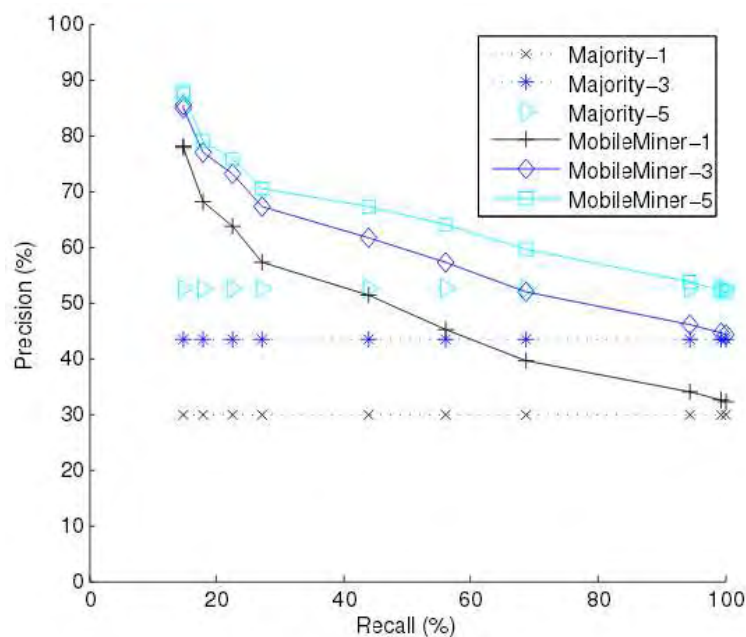
Για την περίπτωση των εξερχόμενων κλήσεων, το μοντέλο θα μπορούσε να χρησιμοποιηθεί εκτός από τη διευκόλυνση της ανάκτησης των επαφών, ως εφαρμογή υπενθύμισης κλήσεων. Οι ερευνητές πρότειναν τη διασύνδεσή του με το ημερολόγιο των χρηστών για την υπενθύμιση πραγματοποίησης κλήσεων ανάλογα με το ιστορικό κλήσεων και τα γεγονότα που έχουν εισαχθεί στο ημερολόγιο.

Η ίδια ερευνητική ομάδα ανέπτυξε ένα πιθανοτικό μοντέλο που αναλύει το ιστορικό κλήσεων των χρηστών για να προβλέψει τη μελλοντική τους συμπεριφορά (Phithakkitnukoon, Dantu, Claxton & Eagle, 2011) και το οποίο μπορεί να χρησιμοποιηθεί για την πρόβλεψη εισερχόμενων και εξερχόμενων κλήσεων. Στο μοντέλο ενσωματώθηκαν ως παράγοντες που επηρεάζουν την επικοινωνία ο χρόνος πραγματοποίησης κλήσεων, η περιοδικότητα τους και η αμοιβαιότητα ανάμεσα στους χρήστες. Οι ερευνητές παρατήρησαν ότι τα πρότυπα κλήσης των χρηστών προς τις επαφές τους διακρίνονται σε δύο κατηγορίες: κλήσεις που πραγματοποιούνται μέσα σε μια συγκεκριμένη περίοδο της ημέρας και η συχνότητά τους ελαττώνεται όσο ο χρόνος αποκλίνει από την περίοδο αυτή (single-hop) και κλήσεις που πραγματοποιούνται σε διαφορετικές χρονικές περιόδους μέσα στην ημέρα (multi-hop). Ένα από τα πιο σημαντικά ευρήματα ήταν πως το πιο πρόσφατο ιστορικό σχετίζεται περισσότερο με τη μελλοντική συμπεριφορά σε σύγκριση με το παλαιότερο, επομένως η χρησιμοποίηση ενός πρόσφατου τμήματος του ιστορικού των χρηστών δίνει καλύτερα αποτελέσματα απ' ό,τι η χρήση ολόκληρου του ιστορικού. Κάτι τέτοιο είναι λογικό, εφόσον τα πρότυπα επικοινωνίας των χρηστών αλλάζουν, οπότε η χρήση ολόκληρου του ιστορικού κλήσεων περιορίζει τη δυνατότητα άμεσου εντοπισμού αυτών των αλλαγών. Μάλιστα, υπολόγισαν ότι για την πρώτη περίπτωση κλήσεων (single-hop) το ιδανικό μέγεθος ιστορικού είναι 30 ημέρες, ενώ για τη δεύτερη περίπτωση κλήσεων (multi-hop) το ιδανικό μέγεθος ιστορικού είναι 60 ημέρες. Εντούτοις, στην αξιολόγηση που πραγματοποιήθηκε με πραγματικά σύνολα δεδομένων από χρήστες κινητών τηλεφώνων, χρησιμοποίησαν δεδομένα 7 μηνών ως περίοδο εκπαίδευσης του μοντέλου και τις κλήσεις των δύο τελευταίων μηνών για προβλέψεις. Όταν η λίστα πρόβλεψης περιελάμβανε μια θέση το μοντέλο προέβλεψε σωστά το 41% των κλήσεων. Όταν η λίστα είχε 5 θέσεις το ποσοστό αυτό αυξήθηκε στο 70%, ενώ όταν περιελάμβανε 15 προτάσεις η επιτυχία ήταν 100%, αν και δε διευκρινίζεται αν οι κλήσεις προς επαφές για πρώτη φορά λήφθηκαν υπόψη. Τέλος, και πάλι βρέθηκε ότι το ποσοστό επιτυχίας των

προβλέψεων εξαρτάται από τον πληθυσμό των επαφών με τις οποίες επικοινωνούν οι χρήστες, όντας πιο χαμηλά τα ποσοστά όταν ο πληθυσμός αυτός είναι μεγαλύτερος.



Εικόνα 16. Παράδειγμα προτύπων που συμπίπτουν (δεξιά) και υπόμνημα για τις τιμές εμπιστοσύνης (αριστερά) (Srinivasan, Moghaddam, Mukherji, Rachuri, Xu & Tapia, 2014)



Εικόνα 17. Απόδοση πρόβλεψης του MobileMiner για τους συνδυασμούς ανάκλησης και ακρίβειας και για 1,3 και 5 θέσεις πρόβλεψης (Srinivasan, Moghaddam, Mukherji, Rachuri, Xu & Tapia, 2014)

Μια άλλη προσέγγιση για την πρόβλεψη της συμπεριφοράς των χρηστών που μπορεί να βρει εφαρμογή στην πρόβλεψη των εξερχόμενων κλήσεων ή στο άνοιγμα των εφαρμογών, βασίζεται στην αναγνώριση προτύπων των οποίων η παρουσία συμπίπτει την ίδια χρονική περίοδο (Srinivasan, Moghaddam, Mukherji, Rachuri, Xu & Tapia, 2014). Η προσέγγιση χρησιμοποιεί κανόνες συσχέτισης (π.χ. εάν είναι βράδυ και το

επίπεδο φόρτισης της μπαταρίας είναι κάτω από τη μέση ειδοποίησε τον χρήστη να τη φορτίσει) που έχουν το πλεονέκτημα ότι είναι εύκολα κατανοητοί από τους ανθρώπους. Ένα παράδειγμα προτύπων που συμπίπτουν και μπορούν να αξιοποιηθούν για την παραγωγή προβλέψεων, οι οποίες χρησιμοποιούνται για τη δημιουργία συντομεύσεων, παρουσιάζεται στην εικόνα 16. Μπορεί να παρατηρήσει κανείς ότι ο συγκεκριμένος χρήστης καλεί τις επαφές Α και Ε όταν είναι οι τελευταίες επαφές που έχει καλέσει. Επιπλέον, καλεί την επαφή Α τις καθημερινές όταν η συσκευή του είναι συνδεδεμένη στο δίκτυο κινητής τηλεφωνίας και την επαφή Β τα Σαββατοκύριακα όταν η συσκευή του είναι συνδεδεμένη σε Wi-Fi. Ο αλγόριθμος (MobileMiner) παράγει προτάσεις μόνο όταν το επίπεδο εμπιστοσύνης των προβλέψεων είναι πάνω από κάποιο προαποφασισμένο όριο. Η τιμή του ορίου επηρεάζει την ανάκληση (ποσοστό των εξερχόμενων κλήσεων για τις οποίες παράγεται πρόβλεψη) και την ακρίβεια (ποσοστό των φορών που ο χρήστης χρησιμοποιεί τη συντόμευση για την πραγματοποίηση κλήσης). Χαμηλότερο επίπεδο εμπιστοσύνης οδηγεί σε υψηλότερη ανάκληση και χαμηλότερη ακρίβεια, ενώ υψηλότερο επίπεδο εμπιστοσύνης έχει ως συνέπεια χαμηλότερη ανάκληση και υψηλότερη ακρίβεια. Τα αποτελέσματα αξιολόγησης της προσέγγισης με ένα σύνολο 25 χρηστών παρουσιάζονται στην εικόνα 17.

Επιπλέον, ένα μέρος της ερευνητικής δραστηριότητας μελετάει τρόπους πρόβλεψης της γενικής συμπεριφοράς των χρηστών συσκευών κινητού υπολογισμού. Για την περίπτωση των κλήσεων ή των μηνυμάτων, αυτές οι εργασίες προσπαθούν να προβλέψουν αν θα υπάρξει δραστηριότητα στα διακριτά χρονικά διαστήματα της ημέρας, χωρίς να ενδιαφέρονται για τις επαφές με τις οποίες πραγματοποιείται η επικοινωνία. Μια ενδεικτική εργασία είναι η (Dai, Yang & Ho, 2013), όπου προβλέπεται αν ο χρήστης θα πραγματοποιήσει κλήση ή όχι για κάθε ώρα της ημέρας. Ένα ενδιαφέρον εύρημα ήταν πως η συμπεριφορά κατά τη διάρκεια του Σαββατοκύριακου είναι πιο δύσκολο να προβλεφθεί σε σχέση με τις καθημερινές. Αντίστοιχη προσέγγιση περιγράφεται και στο (Dai & Ho, 2014).

Ακόμα μια ερευνητική εργασία όπου δόθηκε έμφαση στην ελαχιστοποίηση των βημάτων που απαιτούνται από τους χρήστες για την πραγματοποίηση εξερχόμενων επικοινωνιών (κλήσεις, μηνύματα) με τη χρήση συντομεύσεων παρουσιάζεται στο (Bridle & McCreath, 2006). Σε αυτή την περίπτωση, οι ερευνητές μελέτησαν διάφορες μεθόδους για την παραγωγή της συντόμευσης (αντιστοιχεί στην προσαρμοζόμενη λίστα μιας θέσης), η οποία εκτός από την επαφή συνδυάζει και τον προτεινόμενο τρόπο επικοινωνίας:

- Τελευταία επικοινωνία: η συντόμευση προκύπτει από την τελευταία εξερχόμενη επικοινωνία.
- Πιο συχνή επικοινωνία: η δημιουργία της συντόμευσης γίνεται λαμβάνοντας υπόψη την πιο συχνή επικοινωνία έως εκείνη τη στιγμή.
- C4.5: όλες οι προηγούμενες εξερχόμενες επικοινωνίες αναλύονται με βάση τον αλγόριθμο C4.5 και προκύπτει ένα δέντρο απόφασης. Οι ιδιότητες που χρησιμοποιούνται για την πρόβλεψη είναι η τρέχουσα ημέρα, η ώρα και η τοποθεσία (Cell ID). Η επαφή και ο τύπος επικοινωνίας αποτελούν το ζεύγος από το οποίο καθορίζεται η συντόμευση.
- Απλοϊκός ταξινομητής Bayes: η προηγούμενη εξερχόμενη επικοινωνία χρησιμοποιείται ως σύνολο εκπαίδευσης με τις ίδιες ιδιότητες πρόβλεψης με την προηγούμενη περίπτωση. Η συντόμευση προκύπτει από την επαφή και τον τύπο επικοινωνίας με τη μεγαλύτερη πιθανότητα που υπολογίζει ο ταξινομητής.
- Με βάση τα ενδεχόμενα (contingency based): αυτή η μέθοδος υπολογίζει εάν μια επαφή και ένας τύπος επικοινωνίας εξαρτώνται μεταξύ τους. Η επαφή και ο τύπος επικοινωνίας με τη μεγαλύτερη στατιστική σημαντικότητα εξάρτησης χρησιμοποιούνται για τη συντόμευση.
- Υβριδική προσέγγιση: Χρησιμοποιείται ο απλοϊκός ταξινομητής Bayes, αλλά σε περίπτωση που η πιθανότητα πρόβλεψης είναι μικρότερη από 90% χρησιμοποιείται η μέθοδος πιο συχνής επικοινωνίας.

Από την αξιολόγηση των μεθόδων προέκυψε ότι η μέθοδος της τελευταίας επικοινωνίας μειώνει περισσότερο από κάθε άλλη το πλήθος των κλικ που πρέπει να πραγματοποιήσουν οι χρήστες (24% εξοικονόμηση), ενώ η C4.5 μέθοδος και ο απλοϊκός ταξινομητής Bayes ακολουθούν. Η σχετικά χαμηλή απόδοση των αλγορίθμων που βασίζονται στη μάθηση από τα ιστορικά δεδομένα αποδίδεται σε άσχετες ιδιότητες που περιλαμβάνονταν στα παραδείγματα εκπαίδευσης. Εντούτοις, η μέθοδος της τελευταίας επικοινωνίας έχει ως αποτέλεσμα συντομεύσεις που αλλάζουν πολύ συχνά (όπως και οι μέθοδοι C4.5 και απλοϊκού ταξινομητή Bayes). Από την άλλη οι μέθοδοι της πιο συχνής επικοινωνίας και με βάση τα ενδεχόμενα προτείνουν συντομεύσεις που είναι δραστικά πιο σταθερές. Εντέλει, η υβριδική προσέγγιση προτάθηκε ως η πιο κατάλληλη για την παραγωγή συντομεύσεων. Σύμφωνα με τους ερευνητές, περαιτέρω έρευνα χρειάζεται για τον απαιτούμενο συμβιβασμό ανάμεσα στην αποτελεσματικότητα και την ακρίβεια πρόβλεψης της μεθόδου.

2.5 Η τοποθεσία ως παράγοντας πλαισίου για την πρόβλεψη κλήσεων

Είναι ομόφωνα αποδεκτό στη βιβλιογραφία πως η τοποθεσία φαίνεται να είναι μια από τις πιο σημαντικές πτυχές του πλαισίου στην κινητή επικοινωνία (Arminen, 2006). Μάλιστα, οι χρήστες κατατάσσουν την τοποθεσία ως τη δεύτερη σημαντικότερη πτυχή του πλαισίου (Barkhuus, 2003). Η τοποθεσία, εντούτοις, είναι ένα πολύπλοκο είδος πληροφορίας, καθώς οι χρήστες αναφέρονται σε αυτή σε διαφορετικά επίπεδα λεπτομέρειας (πόλη, συνοικία, κτίριο, δωμάτιο μέσα στο κτίριο κ.λπ.)

Όπως είναι φυσικό, η τοποθεσία είναι ανάμεσα στους παράγοντες πλαισίου που έχουν προταθεί ως κατάλληλοι για την πρόβλεψη κλήσεων και τη δημιουργία προσαρμοζόμενων λιστών επαφών. Παρόλα αυτά, καμία από τις ερευνητικές προσπάθειες που έχουν δημοσιευθεί δεν έχουν εξετάσει την επίπτωση που έχει η τοποθεσία του χρήστη στην επιλογή της επαφής για την έναρξη μιας νέας επικοινωνίας.

Στο (Barzaiq & Loke, 2011) η τοποθεσία δε χρησιμοποιήθηκε για τη δημιουργία της δυναμικής λίστας επαφών, αλλά παρουσιάστηκε ως ένας ενδεχομένως σημαντικός παράγοντας πλαισίου που θα μπορούσε να χρησιμοποιηθεί στον αλγόριθμο πρόβλεψης για να βελτιώσει την ακρίβειά του. Μάλιστα, παρατέθηκε το παρακάτω σενάριο που σηματοδοτεί το είδος των χρηστών που θα μπορούσαν να επωφεληθούν από αυτόν τον παράγοντα πλαισίου:

«Ένας χρήστης ονόματι Παύλος ζει για 9 μήνες στην Αυστραλία για σπουδές και στη συνέχεια επιστρέφει στην πατρίδα του για τις καλοκαιρινές διακοπές διάρκειας 3 μηνών. Είναι αναμενόμενο την χρονική περίοδο που βρίσκεται στην Αυστραλία οι περισσότερες κλήσεις του να πραγματοποιούνται με άτομα που ζουν στην Αυστραλία. Από την άλλη, όταν επιστρέφει στην πατρίδα του, οι περισσότερες κλήσεις που ανταλλάσσει είναι με επαφές από την πατρίδα του».

Στο (Sun, Wang, Zheng & Hsu, 2013) και στο (Phithakkitnukoon, Dantu, Claxton & Eagle, 2011) η τοποθεσία αναφέρεται ως ένας δυνητικά σημαντικός παράγοντας που θα μπορούσε να υιοθετηθεί από τους αλγορίθμους που παρουσιάστηκαν. Στο (Min, Kim & Cho, 2008) η σημασιολογική τοποθεσία είναι μέρος του μοντέλου αναπαράστασης του πλαισίου που χρησιμοποιείται για την πρόβλεψη. Οι ερευνητές πραγματοποίησαν μια αξιολόγηση της εφαρμογής τους μόνο με έναν συμμετέχοντα και για μια πολύ σύντομη περίοδο (12 ημέρες), χωρίς επί της ουσίας να εξετάσουν την

επίπτωση οποιασδήποτε από τις διαστάσεις του πλαισίου στην ακρίβεια πρόβλεψης. Παρότι μια τόσο περιορισμένη αξιολόγηση δεν μπορεί να προσφέρει αξιόπιστα αποτελέσματα, πρέπει να σημειωθεί ότι η συμπερίληψη της τοποθεσίας στο μοντέλο δε φαίνεται να βελτίωσε τα αποτελέσματα, καθώς η ακρίβεια που αναφέρθηκε ήταν κοντά σε αυτή που παρουσιάστηκε σε άλλες σχετικές εργασίες (περίπου 70% για λίστα με 5 προτάσεις) όπου η τοποθεσία δεν είχε χρησιμοποιηθεί, όπως στο (Phithakkitnukoon, Dantu, Claxton & Eagle, 2011). Ακόμα μια περίπτωση όπου η σημασιολογική τοποθεσία (π.χ. σπίτι, εργασία, έξω) θεωρείται μέρος του πλαισίου του χρήστη είναι αυτή που παρουσιάστηκε στο (Srinivasan, Moghaddam, Mukherji, Rachuri, Xu & Taria, 2014). Καθώς ο αλγόριθμος πρόβλεψης παράγει προτάσεις ανάλογα με ένα επίπεδο εμπιστοσύνης και όχι όταν η τιμή του είναι κάτω από ένα συγκεκριμένο κατώφλι (σε αυτή την περίπτωση δεν παράγεται πρόβλεψη), οι ερευνητές παρουσιάζουν το συμβιβασμό ανάμεσα σε ανάκληση και ακρίβεια. Για την απευθείας σύγκριση με άλλες ερευνητικές εργασίες, πρέπει να λάβει κανείς υπόψη του την περίπτωση που η ανάκληση είναι 100%, για την οποία η απόδοση του αλγορίθμου είναι κοντά ή ακόμα και χαμηλότερη σε σχέση με άλλους αλγόριθμους πρόβλεψης εξερχόμενων κλήσεων (περίπου 30% για 1 πρόταση, 45% για 4 προτάσεις και 55% για 5 προτάσεις).

Ενώ, όπως είναι εμφανές, η τοποθεσία θεωρείται ευρέως ως μια κατάλληλη διάσταση πλαισίου για προσαρμοζόμενες διεπαφές πρόσβασης στις επαφές, η έρευνα για τις επιπτώσεις της στις κινητές επικοινωνίες και ειδικότερα για την περίπτωση των εξερχόμενων κλήσεων είναι περιορισμένη. Σε μια εργασία (Calabrese, Smoreda, Blondel & Ratti, 2011) οι συγγραφείς ανέλυσαν δεδομένα κινητών τηλεπικοινωνιών διάρκειας ενός έτους και εξερεύνησαν τη σχέση μεταξύ των κλήσεων των συνδρομητών και της φυσικής τους τοποθεσίας. Η ανάλυση περιελάμβανε ζεύγη χρηστών που επικοινωνούσαν μεταξύ τους και αποκάλυψε ότι περισσότερο από 90% των χρηστών που καλούσαν ο ένας τον άλλο είχαν μοιραστεί την ίδια τοποθεσία (ίδια κεραία κινητής τηλεφωνίας), ακόμα και εάν ζούσαν μακριά ο ένας από τον άλλο. Επιπροσθέτως, βρήκαν ότι το 69% των χρηστών που ανταλλάσσουν επικοινωνία συχνά (τουλάχιστον μια φορά το μήνα κατά μέσο όρο) είχαν μοιραστεί τον ίδιο χώρο την ίδια χρονική στιγμή. Ακόμα ένα ενδιαφέρον εύρημα ήταν πως το πλήθος των κλήσεων ανάμεσα στους χρήστες αυξανόταν ακριβώς πριν και μετά τη συνύπαρξή τους στην ίδια τοποθεσία, γεγονός ενδεικτικό κατ' ιδίαν συναντήσεων. Οι κλήσεις πριν τη συνύπαρξη θα μπορούσαν να αφορούν τη διευθέτηση κάποιας συνάντησης, ενώ οι κλήσεις μετά την συνύπαρξη θα μπορούσαν να αφορούν συζητήσεις μετά τις συναντήσεις. Επιπροσθέτως, σε μια άλλη έρευνα (Karikoski & Soikkeli, 2013)

μελετήθηκε η επίπτωση της σημασιολογικής τοποθεσίας (σπίτι, γραφείο, αλλού, εξωτερικό) στις υπηρεσίες επικοινωνίας από έξυπνες συσκευές κινητού υπολογισμού. Σχετικά με τις εξερχόμενες κλήσεις, αναφέρεται ότι πραγματοποιούνται λιγότερο έντονα στο πλαίσιο του σπιτιού (ενδεχομένως επειδή προτιμούνται άλλα μέσα επικοινωνίας σε αυτό το περιβάλλον), όπου όμως η διάρκειά τους είναι η μεγαλύτερη, ενώ όταν πραγματοποιούνται στο γραφείο η διάρκειά τους είναι η μικρότερη. Επικοινωνία μέσω μηνυμάτων βρέθηκε πως πραγματοποιείται εντατικά σε όλες τις τοποθεσίες, με τη μεγαλύτερη συχνότητα να εμφανίζεται στο γραφείο, όπως συμβαίνει και για την εφαρμογή ηλεκτρονικής αλληλογραφίας. Επιπλέον, μια έρευνα (LaRue, Mitchell, Terhorst & Karimi, 2010), στην οποία οι χρήστες κινητών συσκευών συμπλήρωσαν ένα ερωτηματολόγιο, μελέτησε τη χρήση χαρακτηριστικών των κινητών τηλεφώνων που βασίζονται στην τοποθεσία, χωρίς όμως να προσδιορίζει τα χαρακτηριστικά που χρησιμοποιήθηκαν. Σύμφωνα με τα αποτελέσματα που αναφέρθηκαν, οι περισσότεροι χρήστες χρησιμοποιούσαν το τηλέφωνό τους στο γραφείο κατά τη διάρκεια των εργάσιμων ημερών (ενώ το σπίτι ήταν η δεύτερη πιο δημοφιλής απάντηση) και καθώς οδηγούσαν ή στο σπίτι κατά τη διάρκεια του Σαββατοκύριακου. Τέλος, παρόμοια αποτελέσματα προέκυψαν από ακόμα μια σχετική μελέτη (Verkasalo, 2009), η οποία μελέτησε τον τρόπο που χρησιμοποιούνται οι κινητές υπηρεσίες σε διαφορετικά πλαίσια. Όπως αποκαλύφθηκε, οι χρήστες αφιερώνουν το μεγαλύτερο μέρος του χρόνου τους σε υπηρεσίες επικοινωνίας (μηνύματα, κλήσεις), η χρήση των οποίων κατανέμεται σχετικά ομοιόμορφα στο σπίτι, το γραφείο και άλλες τοποθεσίες. Και σε αυτή την έρευνα αναφέρθηκε πως η διάρκεια των κλήσεων είναι μεγαλύτερη στο σπίτι. Σημαντικό μέρος του χρόνου (12-15%) βρέθηκε πως αφιερώνεται στη χρήση εφαρμογών διαχείρισης προσωπικής πληροφορίας. Τέλος, στη μελέτη του Lugano (2008) βρέθηκε πως το 64% των επικοινωνιών κατευθύνθηκε προς ή λήφθηκε από άτομα που ζούσαν στην ίδια πόλη, ενώ το 6% της επικοινωνίας κατευθύνθηκε σε άλλες πόλεις του εσωτερικού (Φινλανδία) και το 30% σε προορισμούς του εξωτερικού. Το μεγάλο ποσοστό επικοινωνιών με το εξωτερικό οφείλεται στο μεγάλο αριθμό συμμετεχόντων που είχαν διαφορετική εθνικότητα.

2.6 Σύνοψη

Η διαχείριση της προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού είναι ένα πρόβλημα που έχει απασχολήσει σε μεγάλο βαθμό την επιστημονική κοινότητα. Το πλαίσιο των χρηστών μπορεί να βελτιώσει σημαντικά τη διαχείριση προσωπικής πληροφορίας σε αυτές τις συσκευές και ειδικότερα την αξιοποίηση της πληροφορίας.

Διάφορα μοντέλα αναπαράστασης του πλαισίου έχουν προταθεί, ανάλογα με τις απαιτήσεις της κάθε περίπτωσης.

Το πλήθος των επαφών (από εκατοντάδες έως και χιλιάδες) που αποθηκεύεται στις συσκευές κινητού υπολογισμού αυξάνει διαρκώς και δυσκολεύει την ανάκτηση τους. Από την άλλη το πρόβλημα γίνεται ακόμα πιο έντονο, καθώς οι χρήστες δε χρησιμοποιούν σχεδόν τις μισές επαφές τους ή δεν αναγνωρίζουν πολλές από αυτές, αλλά ταυτόχρονα είναι απρόθυμοι να καταναλώσουν χρόνο και προσπάθεια για να τις διαγράψουν. Μελέτες για τα πρότυπα επικοινωνίας των χρηστών δείχνουν πως οι παραδοσιακές διεπαφές (π.χ. λίστα πρόσφατων κλήσεων, λίστα συχνών κλήσεων κ.λπ.) αποτυγχάνουν να τους εξυπηρετήσουν αποτελεσματικά όσον αφορά την ανάκτηση.

Η ανάπτυξη προσαρμοζόμενων διεπαφών με τη χρήση αλγορίθμων πρόβλεψης κλήσεων είναι ο πιο συχνά προτεινόμενος τρόπος αντιμετώπισης των προβλημάτων ανάκτησης των επαφών. Εντούτοις, στις περισσότερες περιπτώσεις χρησιμοποιούνται στο μοντέλο πρόβλεψης στοιχεία του πλαισίου χωρίς να αιτιολογείται η επιλογή τους. Χαρακτηριστικά, η τοποθεσία προτείνεται σχεδόν από το σύνολο της βιβλιογραφίας ως ένας σημαντικός παράγοντας που μπορεί να βελτιώσει τις προσαρμοζόμενες διεπαφές, χωρίς ωστόσο να έχει μελετηθεί η επίπτωσή της στην ακρίβεια των προβλέψεων. Τέλος, στην πλειοψηφία των περιπτώσεων, η αξιολόγηση των αλγορίθμων πρόβλεψης έγινε μέσω προσομοίωσης, ενώ όπου πραγματοποιήθηκε πείραμα πεδίου, αυτό ήταν μικρής διάρκειας ή περιελάμβανε πολύ μικρό αριθμό από συμμετέχοντες, οπότε δεν μπορούν να εξαχθούν ασφαλή συμπεράσματα.

3. Μοντελοποίηση Πλαισίου Συσκευών Κινητού Υπολογισμού

3.1 Αναπαράσταση πλαισίου ως διάνυσμα

Όπως ήδη αναφέρθηκε στην εισαγωγή, οι συσκευές κινητού υπολογισμού συγκεντρώνουν ένα μεγάλο πλήθος πληροφοριών για το περιβάλλον και τις κοινωνικές αλληλεπιδράσεις των χρηστών τους. Από το σύνολο αυτών των πληροφοριών εξάγεται το κινητό κοινωνικό πλαίσιο του χρήστη. Η αξιοποίηση του κινητού κοινωνικού πλαισίου προϋποθέτει τη μαθηματική μοντελοποίηση και αναπαράστασή του. Στην επισκόπηση της σχετικής ερευνητικής βιβλιογραφίας παρουσιάστηκε ένα πλήθος διαφορετικών μοντέλων αναπαράστασης του πλαισίου. Τα δύο πιο συχνά χρησιμοποιούμενα μοντέλα είναι αυτά που βασίζονται σε διανύσματα (π.χ. Du & Wang, 2008) και αυτά που βασίζονται σε οντολογίες (π.χ. Korpiä, Häkkinen, Kela, Ronkainen & Käsälä, 2004). Οι προσεγγίσεις που βασίζονται σε διανύσματα έχουν το μειονέκτημα της υπολογιστικής πολυπλοκότητας σε αναζητήσεις ομοιότητας (similarity searches), καθώς κάθε διάσταση χρειάζεται να συγκριθεί ξεχωριστά. Επιπλέον, τα ζυγισμένα διανύσματα απαιτούν συνήθως μια εμπειρική (και επομένως επιρρεπή σε λάθη) εκτίμηση των βαρών. Από την άλλη, οι οντολογίες τείνουν να είναι μη ευέλικτες και περιοριστικές όσον αφορά την εφαρμογή τους σε μεγαλύτερα εύρη προβλημάτων, απαιτώντας προσεκτικές, χειρωνακτικές προσεγγίσεις για την κατασκευή τους.

Στο (Bergman, Komninos, Liarokapis & Clarke, 2012), όπως ήδη παρουσιάστηκε, βρέθηκε ότι ένα μεγάλο μέρος των επαφών (σχεδόν οι μισές στο σύνολο) από τα αποθετήρια των χρηστών δε χρησιμοποιούνται. Οι ερευνητές πραγματοποίησαν ένα πείραμα με μια διεπαφή όπου χρησιμοποιήθηκε μόνο μια διάσταση του κοινωνικού πλαισίου (συχνότητα επικοινωνίας) για τον εντοπισμό των σημαντικών επαφών. Σε αυτή την περίπτωση η «σημαντικότητα» των επαφών ορίστηκε ως ένα διάνυσμα μονής τιμής (single-value vector) της συχνότητας χρήσης. Όπως αναφέρθηκε, οι χρήστες παρείχαν πολύ θετική ανατροφοδότηση γι' αυτή την προσέγγιση (86% τη θεώρησαν πιο εύκολη στη χρήση από την παραδοσιακή διεπαφή και 64% θα ήθελαν να ήταν διαθέσιμη στην επόμενη συσκευή τους), κάτι που υποστηρίχθηκε και από ποσοτικά δεδομένα (λιγότερα πατήματα πλήκτρων και λιγότερος απαιτούμενος χρόνος για εύρεση των επαφών και έναρξη επικοινωνίας). Παρότι σε αυτή την προσέγγιση υιοθετήθηκε μια απλοϊκή ταξινόμηση της σημαντικότητας των επαφών (απλά η συχνότητα επικοινωνίας), τα ευρήματα καταδεικνύουν αρκετά

αποτελεσματικά πως η αξιοποίηση του πλαισίου ακόμα και με μια μόνο διάσταση μπορεί να έχει θετική επίπτωση στην ανάκτηση προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού.

Στα πλαίσια της παρούσας διδακτορικής έρευνας, και θεωρώντας ότι η αξιοποίηση περισσότερων διαστάσεων του κινητού και κοινωνικού πλαισίου σε σχέση με την προσέγγιση των Bergman, Komninos, Liarakis και Clarke (2012) μπορεί να οδηγήσει σε ακόμα καλύτερα αποτελέσματα ως προς την ανάκτηση των επαφών, ήταν εμφανές ότι απαιτείται ένας k-πολυδιάστατος χώρος (k-dimensional context space) για την αναπαράσταση του πλαισίου (Komninos, Plessas, Stefanis & Garofalakis, 2011a; 2011b). Η υιοθέτηση μιας λύσης που βασίζεται σε διανύσματα, παρά τα μειονεκτήματα, παρέχει σημαντική ευελιξία όσον αφορά την αναπαράσταση του πλαισίου (Komninos, Plessas, Stefanis & Garofalakis, 2011a). Όπως φαίνεται από τις εξισώσεις (1), (2) και (3), προτείνεται ένα διανυσματικό μοντέλο για την αναπαράσταση του πλαισίου ενός αντικειμένου προσωπικής πληροφορίας i στο k-πολυδιάστατο χώρο ως μια k-πλειάδα, όπου d_n είναι η ατομική τιμή για τη διάσταση n του πλαισίου. Η σημαντικότητα του I_i ορίζεται ως οι σταθμισμένες αποστάσεις ανάμεσα στις ατομικές διανυσματικές τιμές που περιγράφουν το τρέχον πλαίσιο και τις ατομικές διανυσματικές τιμές του αντικειμένου i ($\Delta\vec{C}(i)$).

$$\vec{C}(i) = (d_1, d_2, \dots, d_n) \quad (1)$$

$$I_i = \sum_{n=1}^n (w_{d_n} \times (1 - \Delta\vec{C}(i))) \quad (2)$$

$$\Delta\vec{C}(i) = |\vec{C}(n) - \vec{C}(i)| \quad (3)$$

Από αυτόν τον ορισμό είναι εμφανές ότι η σημαντικότητα ενός αντικειμένου προσωπικής πληροφορίας είναι απόλυτα συνυφασμένη με το τρέχον πλαίσιο του χρήστη, σε αντίθεση με μια γενικότερη θεώρηση της. Αυτό σημαίνει πως ένα αντικείμενο πληροφορίας που μπορεί ένας χρήστης να χρειάζεται σε ένα συγκεκριμένο πλαίσιο, ενδεχομένως να μην έχει την ίδια αξία σε κάποιες άλλες συνθήκες. Θα μπορούσε να πει κανείς ότι η πληροφορία είναι σημαντική για τον χρήστη όταν τη χρειάζεται στο πλαίσιο στο οποίο βρίσκεται. Ενδεικτικό για τη διαφοροποίηση της σημαντικότητας ανάλογα με το πλαίσιο είναι το παρακάτω σενάριο:

«Ένας χρήστης έχει στο αποθετήριο των επαφών του την επαφή «Γιάννης», την οποία καλεί σε καθημερινή βάση. Επίσης, στις επαφές του περιλαμβάνεται και η επαφή

«Ιωάννα», με την οποία επικοινωνεί μόνο μια φορά το χρόνο. Η επαφή «Γιάννης» μπορεί να θεωρηθεί γενικά σημαντική. Στην περίπτωση όμως που ο χρήστης έχει καθυστερήσει σε ένα ραντεβού με την Ιωάννα που είναι προγραμματισμένο να λάβει χώρα σε 10 λεπτά, τότε η επαφή «Ιωάννα» γίνεται αναμφισβήτητα πιο σημαντική από οποιαδήποτε άλλη και η σημαντικότητά της θα πρέπει να αυξηθεί και να φθίνει σταδιακά καθώς ο χρόνος πλησιάζει και απομακρύνεται από το προγραμματισμένο γεγονός».

Αν λάβει κανείς υπόψη τα κριτήρια που σύμφωνα με τη βιβλιογραφία (Komninos & Liaroikaris, 2009) πρέπει να διερευνηθούν ως προς τη σχετικότητά τους με τη σημαντικότητα των αντικειμένων προσωπικής πληροφορίας και ιδιαίτερα των επαφών, προκύπτουν οι παρακάτω διαστάσεις πλαισίου:

- Συχνότητα (π.χ. συχνότητα χρήσης τους προηγούμενους n μήνες)
- Προσφατότητα (π.χ. ημέρες που έχουν περάσει από την τελευταία χρήση)
- Τοποθεσία (π.χ. γεωγραφικές περιοχές από τις οποίες τουλάχιστον n% των χρήσεων έχουν πραγματοποιηθεί)
- Ώρα της ημέρας (π.χ. χρονικό τμήμα της ημέρας κατά το οποίο έχουν συμβεί τουλάχιστον n% των χρήσεων)
- Προγραμματισμένη εργασία (π.χ. λογικό μέτρο της ύπαρξης μιας προγραμματισμένης εργασίας που περιλαμβάνει την επαφή εντός ενός χρονικού παραθύρου ή η χρονική απόσταση από μια τέτοια εργασία)
- Προσωπική προτίμηση (π.χ. κλίμακα 1-5 της αξιολόγησης του χρήστη για την σημαντικότητα μιας επαφής)

Για να υπολογιστεί η σημαντικότητα μιας επαφής με βάση τους παραπάνω τύπους πληροφοριών πλαισίου και το τρέχον πλαίσιο του χρήστη, μια τυπική προσέγγιση είναι ο υπολογισμός της απόστασης ανάμεσα στο τρέχον πλαίσιο και τα δεδομένα πλαισίου που αντιστοιχούν στην επαφή (π.χ. η τρέχουσα ώρα της ημέρας σε σχέση με τη συνηθισμένη ώρα της ημέρας που χρησιμοποιείται κυρίως η επαφή), σε συνδυασμό με το στατικό πλαίσιο (όπως η προσωπική προτίμηση των χρηστών που μπορεί να εκφράζεται με μια μορφή γενικής αξιολόγησης). Οι τιμές που προκύπτουν πρέπει να ζυγιστούν και το άθροισμα των ζυγισμένων τιμών μπορεί να χρησιμοποιηθεί για να συμπεράνει κανείς τη σημαντικότητα της επαφής στο τρέχον πλαίσιο (εξισώσεις 1, 2 και 3).

Ένα σημείο που πρέπει να τονιστεί είναι πως η μοντελοποίηση που παρουσιάστηκε εκτείνεται πέρα από την έννοια της σημαντικότητας των επαφών. Στην ουσία, η χρήση ενός πολυδιάστατου διανύσματος για την αναπαράσταση του πλαισίου και τον

υπολογισμό της σημαντικότητας μπορεί να εφαρμοστεί για οποιοδήποτε τύπο προσωπικής πληροφορίας αποθηκεύεται στις συσκευές των χρηστών (π.χ. φωτογραφίες, μηνύματα, emails κ.λπ.). Επομένως, το μοντέλο μπορεί να επεκταθεί και να χρησιμοποιηθεί για οποιοδήποτε πρόβλημα ανάκτησης προσωπικής πληροφορίας από συσκευές κινητού υπολογισμού, λαμβάνοντας υπόψη τις διαστάσεις εκείνες του πλαισίου που είναι σχετικές με το πρόβλημα.

3.1.1 Μείωση διαστάσεων

Μια λύση που θα μπορούσε να αντιμετωπίσει το πρόβλημα της υπολογιστικής πολυπλοκότητας της ανάκτησης στην περίπτωση των διανυσματικών μοντέλων, αλλά και τον υπολογισμό των βαρών για κάθε διάσταση (που είναι πολύ πιθανό να χρειάζεται να οριστούν δυναμικά με βάση το πλαίσιο και όχι στατικά) είναι η μείωση διαστάσεων (dimensionality reduction). Όπως μπορεί να συμπεράνει κανείς από το όνομά της, πρόκειται για μια αλγοριθμική τεχνική που μειώνει τις διαστάσεις των δεδομένων και βρίσκει εφαρμογή σε πεδία όπως οι βάσεις δεδομένων, η ανάκτηση πληροφορίας, η εξόρυξη δεδομένων, τα συστήματα παραγωγής προτάσεων, η επεξεργασία σημάτων κ.α. Τα δεδομένα που συλλέγονται από τον πραγματικό κόσμο χαρακτηρίζονται από υψηλό βαθμό διαστάσεων, ένα γεγονός που επηρεάζει την απόδοση της επεξεργασίας τους και χαρακτηρίστηκε από τον Bellman (1957) με τη φράση “curse of dimensionality”, η οποία δηλώνει εκθετική εξάρτηση των αλγορίθμων από το πλήθος των διαστάσεων των δεδομένων εισόδου.

Η ιδέα πίσω από την αλγοριθμική αυτή τεχνική είναι ο μετασχηματισμός των δεδομένων από ένα χώρο με υψηλό αριθμό διαστάσεων (high-dimensional space) σε ένα χώρο με χαμηλό αριθμό διαστάσεων (low-dimensional space), διατηρώντας μερικές κρίσιμες σχέσεις ανάμεσα στα στοιχεία του συνόλου δεδομένων. Με μαθηματικούς όρους δεδομένου ενός αντικειμένου p διαστάσεων $x=(x_1, \dots, x_p)^T$, το πρόβλημα ανάγεται στην εύρεση μιας αναπαράστασης του με μικρότερο αριθμό διαστάσεων $s=(s_1, \dots, s_k)^T$ όπου $k < p$, η οποία περιλαμβάνει το περιεχόμενο του πρωτότυπου αντικειμένου σύμφωνα με κάποιο κριτήριο (Fodor, 2002). Υπάρχουν δυο κατηγορίες μεθόδων που επιλύουν το πρόβλημα:

- Η επιλογή χαρακτηριστικών (feature selection), όπου ένα βέλτιστο υποσύνολο χαρακτηριστικών (διαστάσεων) επιλέγεται.
- Η εξαγωγή χαρακτηριστικών (feature extraction), όπου τα υπάρχοντα χαρακτηριστικά συνδυάζονται και μετασχηματίζονται σε νέα.

Η ερευνητική ιδέα που παρουσιάζεται σε αυτή την ενότητα αφορά την εφαρμογή της τεχνικής της μείωσης διαστάσεων σε αντικείμενα προσωπικής πληροφορίας που

έχουν επαυξηθεί με δεδομένα πλαισίου, όπως οι εγγραφές στο αποθετήριο των επαφών. Η επιλογή χαρακτηριστικών δε φαίνεται να είναι μια πρακτική λύση σε αυτή την περίπτωση, καθώς δεν είναι δυνατό να γνωρίζει κανείς εκ των προτέρων τα χαρακτηριστικά που περιγράφουν με τον βέλτιστο τρόπο το διάνυσμα του πλαισίου, ούτε και τα βάρη που πρέπει να τους αποδοθούν. Από την άλλη, καθώς τα αντικείμενα προσωπικής πληροφορίας μπορούν να αναπαρασταθούν σύμφωνα με το μοντέλο που παρουσιάστηκε ως πολυδιάστατα διανύσματα, η εξαγωγή ενός μικρού συνόλου χαρακτηριστικών που αναπαριστούν με ακρίβεια τα πρωτότυπα αντικείμενα και τις σχέσεις μεταξύ τους θα επέτρεπε γρήγορες και ακριβείς αναζητήσεις ομοιότητας για σχετικά αντικείμενα προσωπικής πληροφορίας. Επιπροσθέτως, μετά τη μείωση των διαστάσεων των αντικειμένων σε μια, δύο ή τρεις, είναι εφικτή η αντιστοίχισή τους σε χώρους αντίστοιχου αριθμού διαστάσεων και επομένως η οπτικοποίησή τους, η οποία μπορεί να αποκαλύψει υπάρχουσες ομάδες αντικειμένων.



Εικόνα 18. Υβριδική, μονοδιάστατη αντιστοίχιση της σημαντικότητας



Εικόνα 19. Διεπαφές ανάκτησης που χρησιμοποιούν 2D (αριστερά) και 3D (δεξιά) προβολές της σημαντικότητας των αντικειμένων (επαφές, μηνύματα, emails) σε συνδυασμό με διατηρημένες μη προβολόμενες διαστάσεις (χρόνος, απόσταση από την τρέχουσα τοποθεσία)

Παραδείγματα τέτοιων αναπαραστάσεων παρουσιάζονται στις εικόνες 18 και 19. Στην εικόνα 18 παρουσιάζεται μια υβριδική διεπαφή, όπου οι επαφές ταξινομούνται

αλφαβητικά (για να διατηρηθεί η εξοικείωση των χρηστών με τις υπάρχουσες διεπαφές), αλλά για κάθε γράμμα οι σημαντικές επαφές τοποθετούνται στην κορυφή της λίστας με έντονα γράμματα. Στην εικόνα 19 παρουσιάζονται διεπαφές που έχουν προκύψει από τη μείωση του διανύσματος του πλαισίου σε δύο ή τρεις διαστάσεις. Στην περίπτωση των δύο διαστάσεων, η διάσταση του χρόνου μπορεί να διατηρηθεί ούτως ώστε ο χρήστης να μπορεί να ανακτήσει την πιο σημαντική επαφή ανά χρονική περίοδο. Στην περίπτωση των τριών διαστάσεων παρουσιάζεται ένα παράδειγμα για το πώς αυτή η τεχνική επεκτείνεται πέρα από τον τομέα των επαφών και συμπεριλαμβάνει και άλλους τύπους αντικειμένων προσωπικής πληροφορίας (π.χ. μηνύματα και emails). Οι άξονες που χρησιμοποιούνται για την αναπαράσταση των αντικειμένων είναι η σημαντικότητα και ο χρόνος με την απόσταση από την τρέχουσα τοποθεσία, οι οποίοι έχουν διατηρηθεί από τα αρχικά δεδομένα.

Ένα μεγάλο πλήθος από αλγορίθμους μείωσης διαστάσεων, με διαφορετικά χαρακτηριστικά, υπάρχει διαθέσιμο στη βιβλιογραφία. Μια ενδιαφέρουσα προσέγγιση που θα μπορούσε να βρει εφαρμογή στις συσκευές κινητού υπολογισμού λόγω της απλότητας της και της υπολογιστικής αποδοτικότητας είναι ο αλγόριθμος Fastmap (Faloutsos και Lin, 1995). Ο Fastmap είναι ένας γρήγορος αλγόριθμος που χρησιμοποιεί γεωμετρικούς μετασχηματισμούς για να αντιστοιχίσει αντικείμενα ενός χώρου με υψηλό αριθμό διαστάσεων σε ένα χώρο με χαμηλό αριθμό διαστάσεων, διατηρώντας ικανοποιητικά τις αποστάσεις ανάμεσα στα αντικείμενα και τη δομή του συνόλου δεδομένων και επομένως διατηρώντας και τις ανομοιοότητες ανάμεσα στα αντικείμενα. Τα πειράματα των ερευνητών (Faloutsos και Lin, 1995) έδειξαν ότι ο αλγόριθμος αποδίδει καλά για συσταδοποιήσεις και οπτικοποιήσεις. Η υπολογιστική του πολυπλοκότητα είναι $O(N \cdot k^2)$, όπου k το πλήθος των διαστάσεων του χώρου – στόχου και τα βήματα του αλγορίθμου είναι τα παρακάτω:

1. Βρες δύο αντικείμενα που βρίσκονται μακριά το ένα από το άλλο.
2. Πρόβαλε όλα τα αντικείμενα πάνω στη γραμμή που ορίζουν τα δύο αντικείμενα για να υπολογιστεί η πρώτη συντεταγμένη.
3. Πρόβαλε όλα τα αντικείμενα σε ένα υπερεπίπεδο κάθετο στη γραμμή που τα δύο αντικείμενα ορίζουν.
4. Επανάλαβε $k-1$ φορές.

3.1.2 Απόδοση του αλγορίθμου Fastmap σε συσκευές κινητού υπολογισμού

Για τη διερεύνηση της καταλληλότητας της εφαρμογής του αλγορίθμου σε συσκευές κινητού υπολογισμού, δεδομένων των περιορισμών που αυτές αντιμετωπίζουν,

πραγματοποιήθηκε πείραμα σε έξυπνα κινητά τηλέφωνα Android με σκοπό να μελετηθεί αν ο χρόνος εκτέλεσης του αλγορίθμου είναι τέτοιος που επιτρέπει την εκτέλεσή του σε πραγματικό χρόνο (Stefanis, Plessas, Komninos & Garofalakis, 2012a). Για το σκοπό αυτό υλοποιήθηκε ο αλγόριθμος Fastmap για το λειτουργικό σύστημα Android (SDK API Level 8 – Android 2.2.x) με βάση την πρωτότυπη υλοποίηση σε C των Faloutsos και Lin (1995). Τα πειράματα πραγματοποιήθηκαν με 7 συνθετικά σύνολα δεδομένων που περιελάμβαναν τιμές που παρήχθησαν με γεννήτρια ψευδοτυχαίων αριθμών (πίνακας 4). Τα τυχαία αυτά αντικείμενα αναπαριστούν αντικείμενα προσωπικής πληροφορίας (όπως οι επαφές). Σε κάθε αντικείμενο αντιστοιχεί ένα διάνυσμα με τις τιμές των διαστάσεων του κινητού κοινωνικού πλαισίου.

Πίνακας 4. Σύνολα δεδομένων που χρησιμοποιήθηκαν στα πειράματα

Σύνολο	Αντικείμενα	Διαστάσεις	Συνολικά Στοιχεία
1	500	5	2.500
2	1.000	5	5.000
3	2.000	5	10.000
4	500	8	4.000
5	1.000	8	8.000
6	2.000	8	16.000
7	2.000	10	20.000

Πίνακας 5. Χαρακτηριστικά συσκευών που χρησιμοποιήθηκαν στα πειράματα

#	Model	OS Version	CPU	RAM
Σ1	HTC Magic	2.3.7	Qualcomm MSM7200 A ARM 11 - 528 MHz	288 MB
Σ2	Huawei U8150 IDEOS	2.2	Qualcomm MSM7225 - 528 MHz	256 MB
Σ3	Samsung Galaxy 3	2.2	Samsung S5P6422 667 MHz	256 MB
Σ4	Samsung Nexus S	2.3.6	Cortex-A8 - 1 GHz	512 MB
Σ5	Sony Ericsson Xperia Arc S	2.3.4	Qualcomm MSM8255 T Scorpion - 1.4 GHz	512 MB
Σ6	Sony Ericsson Xperia mini pro	2.3.4	Qualcomm MSM8255 Scorpion - 1 GHz	512 MB

Για τα πειράματα χρησιμοποιήθηκε μια ποικιλία από διαφορετικές συσκευές που καλύπτουν μεγάλο εύρος τεχνολογικών δυνατοτήτων. Ο πίνακας 5 παρουσιάζει τα χαρακτηριστικά αυτών των συσκευών. Για κάθε μια συσκευή Σ1 – Σ6, ο αλγόριθμος εκτελέστηκε για όλα τα σύνολα δεδομένων για $k=2$ και $k=3$, όπου k το πλήθος των διαστάσεων του τελικού διανυσματικού χώρου. Για κάθε σύνολο δεδομένων και κάθε συσκευή πραγματοποιήθηκαν 10 εκτελέσεις του αλγορίθμου, ώστε να αποκτηθεί μια ρεαλιστική εικόνα, καθώς οι συσκευές Android εκτελούν στο παρασκήνιο και άλλες διεργασίες που επηρεάζουν την απόδοση και υπολογίστηκαν οι μέσοι όροι των χρόνων εκτέλεσης. Επιπλέον, οι υπολογισμοί αφορούσαν τον πραγματικό χρόνο εκτέλεσης του αλγορίθμου, χωρίς να λαμβάνουν υπόψη το χρόνο που απαιτείται για τη δημιουργία των γραφικών της οθόνης ή την ανάγνωση των δεδομένων από τα αρχεία και τη μεταφορά τους στη μνήμη της συσκευής. Οι μέσοι χρόνοι εκτέλεσης για κάθε περίπτωση παρουσιάζονται στους πίνακες 6 και 7.

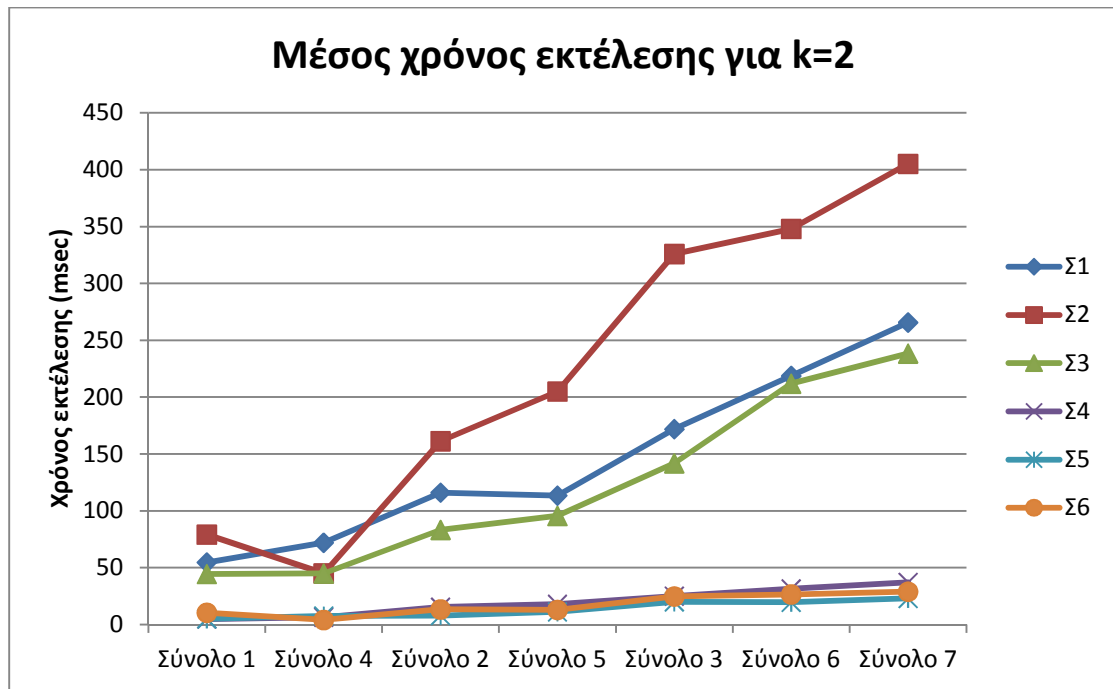
Πίνακας 6. Μέσος χρόνος εκτέλεσης πειράματος ανά συσκευή και σύνολο δεδομένων σε msec για $k=2$

	Σύνολο 1	Σύνολο 2	Σύνολο 3	Σύνολο 4	Σύνολο 5	Σύνολο 6	Σύνολο 7
Σ1	54,7	116,1	171,9	72,1	113,5	218,7	265,7
Σ2	79,2	161,3	325,9	45,1	205	348	405,2
Σ3	44,6	83,3	141,9	45,1	95,6	212	238,4
Σ4	4,9	15,4	25,1	6,5	18	31,5	37,2
Σ5	5,1	8	20	7,7	11,2	19,6	23
Σ6	10,5	13,4	24,8	4,1	13	26,6	29

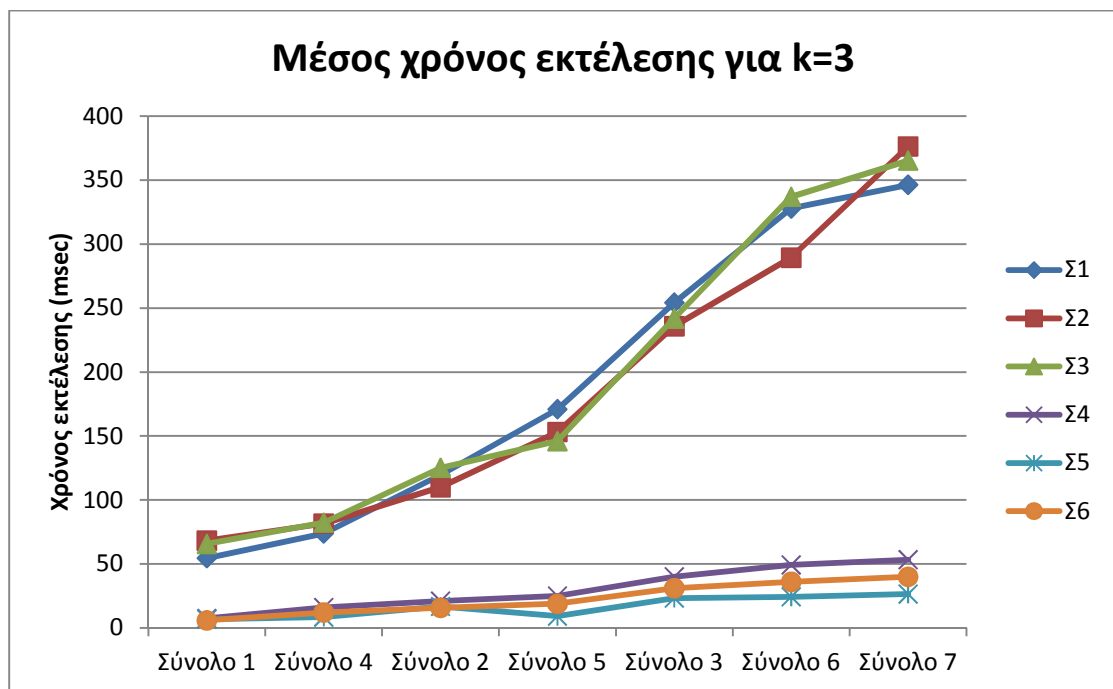
Πίνακας 7. Μέσος χρόνος εκτέλεσης πειράματος ανά συσκευή και σύνολο δεδομένων σε msec για $k=3$

	Σύνολο 1	Σύνολο 2	Σύνολο 3	Σύνολο 4	Σύνολο 5	Σύνολο 6	Σύνολο 7
Σ1	54,6	119,5	254,3	73,9	171	328	346,4
Σ2	68,2	110	235,9	81,6	153,2	289,4	376,4
Σ3	65,8	125,2	241,9	82,4	146,1	337,1	365,4
Σ4	7,4	21	40	16	25	49,3	53,4
Σ5	6,8	16,8	23,4	8,5	9,2	24,4	26,5
Σ6	5,9	15,7	31,1	12,1	18,9	36,1	40

Τα αποτελέσματα ήταν ενθαρρυντικά όσον αφορά την απόδοση του αλγορίθμου σε συσκευές κινητού υπολογισμού: ο χειρότερος χρόνος εκτέλεσης ($k=3$, Σύνολο 7, Σ2) είναι κάτω από 0,4 δευτερόλεπτα για μια συσκευή χαμηλών χαρακτηριστικών και για ένα σύνολο το οποίο είναι πολύ μεγαλύτερο από το μέσο αναμενόμενο σε μια συσκευή κινητού υπολογισμού. Στα γραφήματα 2 και 3 εμφανίζονται τα αποτελέσματα οπτικοποιημένα, όπου φαίνεται ότι όλες οι συσκευές ακολουθούν παρόμοιες γραμμές για το χρόνο εκτέλεσης, με διακυμάνσεις που είναι συνεπείς με τις διαφορές στα χαρακτηριστικά τους (υπολογιστική ισχύς και μνήμη).



Γράφημα 2. Μέσος χρόνος εκτέλεσης πειράματος για k=2



Γράφημα 3. Μέσος χρόνος εκτέλεσης πειράματος για k=3

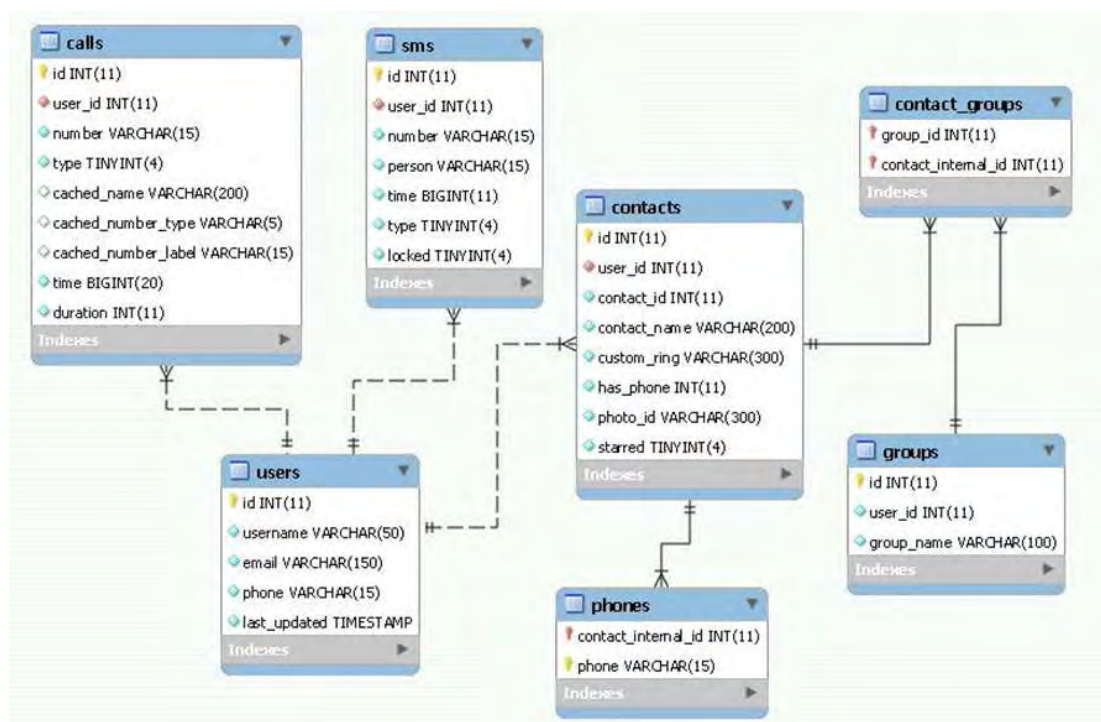
3.2 Πείραμα μελέτης αποθετηρίου επαφών και προτύπων επικοινωνίας

Το επόμενο βήμα μετά τον ορισμό ενός μοντέλου για την αναπαράσταση του πλαισίου, είναι ο εντοπισμός εκείνου του τμήματος του που παίζει σημαντικό ρόλο για τη διαδικασία ανάκτησης των επαφών. Σύμφωνα με τους Korpirää, Häkklä, Kela,

Ronkainen και Käsälä (2004), όταν σχεδιάζεται ένα ενήμερο για το πλαίσιο σύστημα, δεν είναι σχετικό με το πρόβλημα που προσπαθεί να αντιμετωπίσει όλο το πλαίσιο που καταγράφεται, αλλά μόνο ένα μέρος του, το οποίο ο σχεδιασμός πρέπει να λάβει υπόψη. Για την εύρεση και ενσωμάτωση στο μοντέλο εκείνων των διαστάσεων του πλαισίου που επηρεάζουν την αξιοποίηση των επαφών από τους χρήστες των συσκευών κινητού υπολογισμού, σχεδιάστηκε και εκτελέστηκε ένα πείραμα μελέτης του αποθετηρίου των επαφών και των προτύπων επικοινωνίας των χρηστών (Stefanis, Plessas, Komninos & Garofalakis, 2012b).

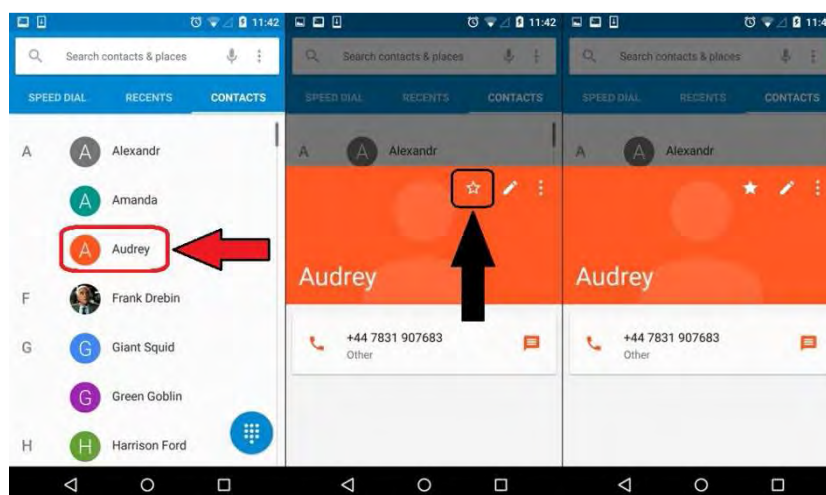
3.2.1 Σχεδιασμός πειράματος και δείγμα συμμετεχόντων

Στόχος του πειράματος ήταν να μελετηθεί η σύνθεση του αποθετηρίου επαφών των συμμετεχόντων και να αναλυθούν τα πρότυπα επικοινωνίας τους. Για την πραγματοποίηση του σχεδιάστηκε και υλοποιήθηκε μια Android εφαρμογή, η οποία εξήγαγε από τις συσκευές των συμμετεχόντων τη λίστα επαφών, το αρχείο κλήσεων, το αρχείο μηνυμάτων SMS (χωρίς το περιεχόμενο των μηνυμάτων) και τις υπάρχουσες ομάδες χρηστών. Η πληροφορία αυτή αποθηκευόταν σε αρχεία κειμένου, τα οποία στη συνέχεια επεξεργάζονταν και οι επιμέρους εγγραφές αποθηκεύονταν σε μια βάση δεδομένων (εικόνα 20). Για την προστασία της ιδιωτικότητας των συμμετεχόντων υπήρχε η δυνατότητα κρυπτογράφησης δεδομένων όπως τα ονόματα των επαφών και οι αριθμοί τηλεφώνου.

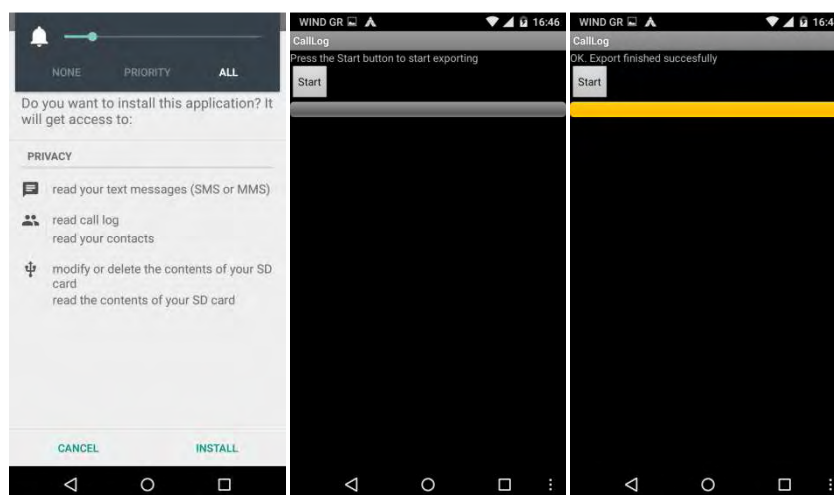


Εικόνα 20. Σχήμα ΒΔ για την αποθήκευση των δεδομένων από τις συσκευές των χρηστών

Στο παραπάνω σχήμα της ΒΔ περιλαμβάνονται πίνακες για την αποθήκευση των χρηστών που συμμετείχαν στο πείραμα (users), των επαφών (contacts και phones), των κλήσεων (calls), των μηνυμάτων (sms) και των ομάδων επαφών (groups και contact_groups). Τα ονόματα των στηλών σε κάθε πίνακα είναι περιγραφικά σχετικά με την πληροφορία που αποθηκεύουν. Ιδιαίτερη μνεία ίσως χρειάζεται να γίνει για τη στήλη type του πίνακα calls που καθορίζει αν μια κλήση είναι εισερχόμενη (τιμή 1), εξερχόμενη (τιμή 2) ή μη απαντημένη (τιμή 3). Επίσης, η στήλη starred (τιμή 1 ή 0) στον πίνακα contacts αναφέρεται στη δυνατότητα που έχουν οι χρήστες των συσκευών Android να δηλώσουν επαφές ως αγαπημένες (εικόνα 21).



Εικόνα 21. Δυνατότητα χαρακτηρισμού επαφής ως αγαπημένη σε συσκευές Android



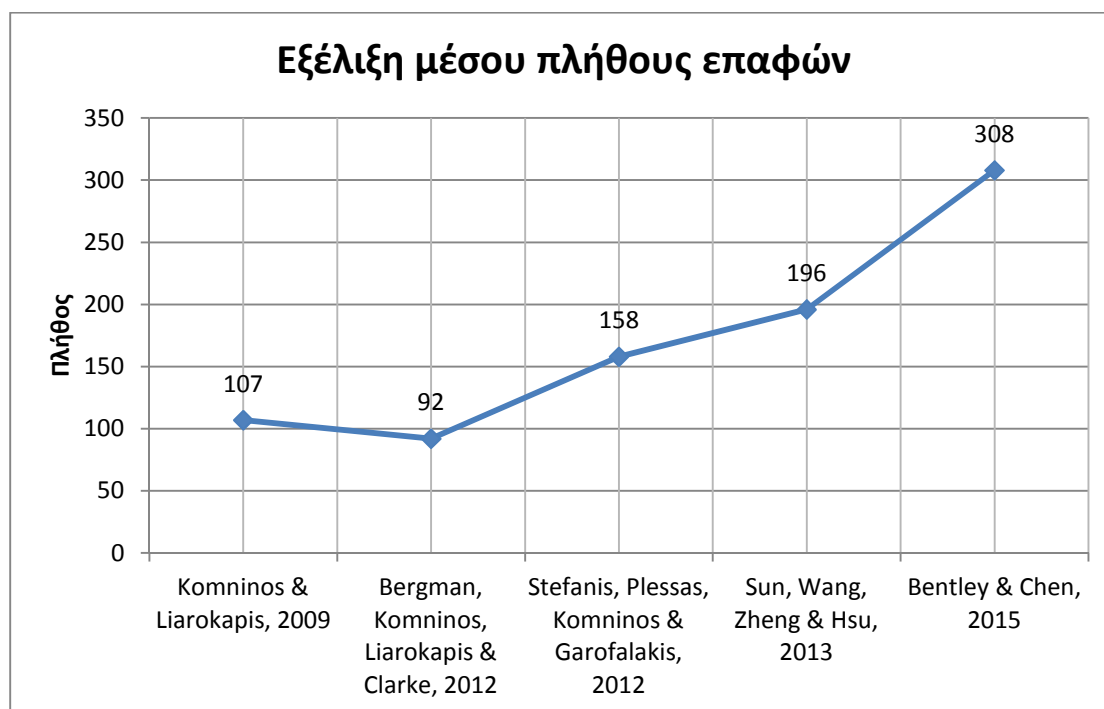
Εικόνα 22. Screenshots από τις οθόνες εγκατάστασης και εκτέλεσης της εφαρμογής

Η εφαρμογή (εικόνα 22) μοιράστηκε σε 42 χρήστες συσκευών Android που εκδήλωσαν ενδιαφέρον μετά από δημόσια ανακοίνωση στο φόρουμ της σχολής. Μόνο όμως 34 από τα σύνολα δεδομένων που συλλέχθηκαν θεωρήθηκαν έγκυρα, καθώς τα υπόλοιπα ήταν ελλιπή (π.χ. λόγω υπερβολικά μικρού αριθμού εγγραφών

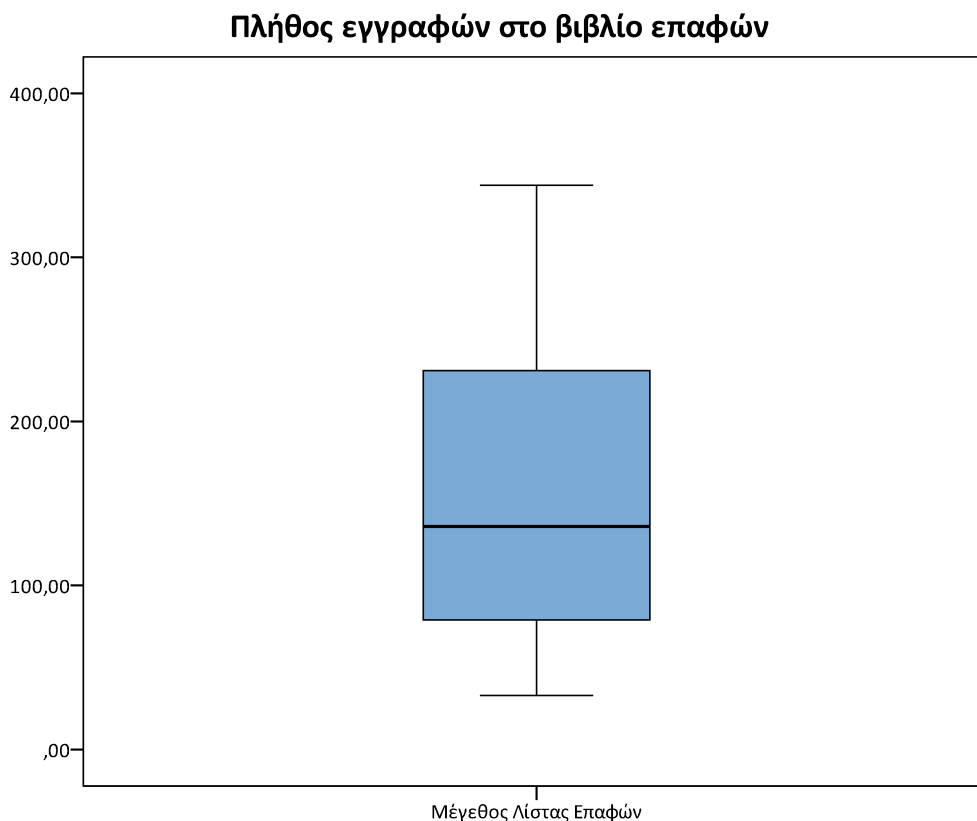
στο αρχείο κλήσεων). Από τους 34 χρήστες των οποίων τα δεδομένα αναλύθηκαν, 31 ήταν άνδρες και 3 γυναίκες. Η ηλικία τους κυμαινόταν από 19 έως 39 έτη, ενώ προέρχονταν από διάφορα υπόβαθρα, αν και οι περισσότεροι ήταν φοιτητές Μηχ. Η/Υ και Πληροφορικής.

3.2.2 Αποτελέσματα

Αρχικά, η ανάλυση των αρχείων που εξήχθησαν από τις συσκευές των χρηστών που συμμετείχαν στο πείραμα επικεντρώθηκε στο μέγεθος της λίστας επαφών. Συνολικά, τα αποθετήρια των συμμετεχόντων περιείχαν 5.384 επαφές, ενώ βρέθηκε ότι κατά μέσο όρο κάθε λίστα επαφών περιείχε 158 εγγραφές ($\text{mean}=158,38$, $\text{sd}=84,74$, $\text{min}=33$, $\text{max}=344$). Συγκρίνοντας με αντίστοιχες προγενέστερες έρευνες, είναι εμφανές ότι το μέσο μέγεθος του αποθετηρίου επαφών είχε αυξηθεί σχεδόν κατά 50%. Το γράφημα 4 αποτελεί επέκταση του γραφήματος 1 με τα αποτελέσματα του παρόντος πειράματος, ενώ στο γράφημα 5 παρουσιάζεται η κατανομή του αριθμού των επαφών ανά βιβλίο επαφών χρήστη. Αξίζει να σημειωθεί ότι στις μεταγενέστερες μελέτες παρατηρήθηκε ακόμα μεγαλύτερο μέσο μέγεθος λίστας επαφών, ενώ στην πιο πρόσφατη (Bentley & Chen, 2015) έχει σχεδόν διπλασιαστεί σε σχέση με τα ευρήματα του πειράματος που περιγράφεται σε αυτή την ενότητα. Είναι, επομένως, βάσιμοι οι προβληματισμοί που αναπτύχθηκαν στα προηγούμενα κεφάλαια σχετικά με τη δυσκολία αξιοποίησης των επαφών, λόγω του ολοένα αυξανόμενου μεγέθους του αποθετηρίου.



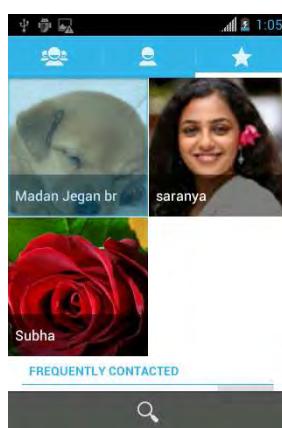
Γράφημα 4. Η εξέλιξη του μέσου πλήθους επαφών στο αποθετήριο μετά το πείραμα



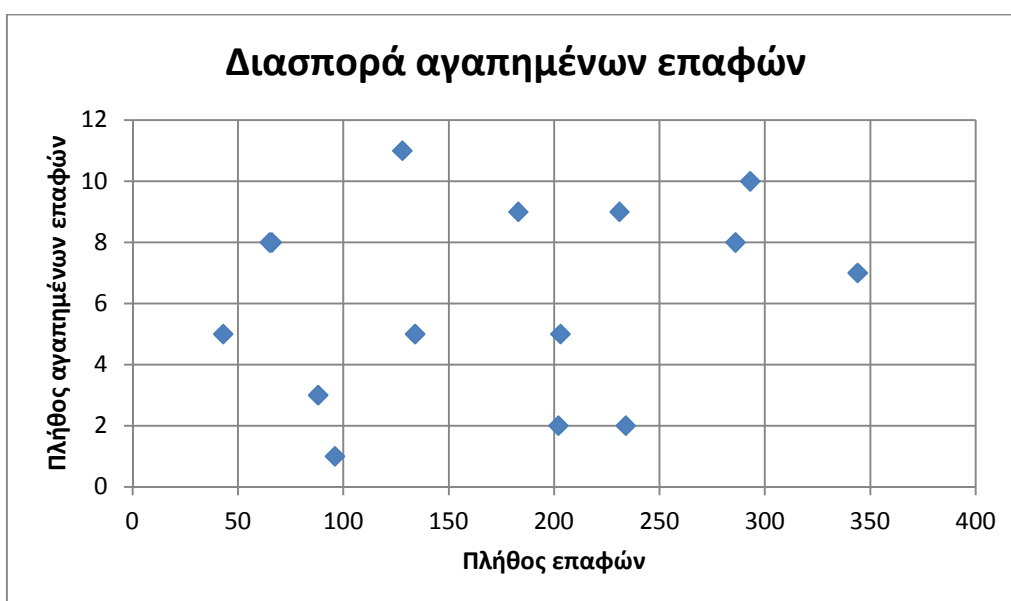
Γράφημα 5. Μέσο μέγεθος των βιβλίων επαφών των χρηστών

Ένα ενδιαφέρον εύρημα ήταν πως 15 από τους συμμετέχοντες είχαν χρησιμοποιήσει τη δυνατότητα των συσκευών Android να ορίζουν επαφές ως αγαπημένες (εικόνα 21), οι οποίες εμφανίζονται στη λίστα αγαπημένων και επιτρέπουν τη γρήγορη επιλογή τους για την έναρξη επικοινωνίας (εικόνα 23). Οι Bentley και Chen (2015) ανέφεραν ότι αυτή ήταν μια δυνατότητα που βρήκαν να χρησιμοποιείται σπάνια στο σύνολο δεδομένων που ανέλυσαν. Αυτές τις επαφές θα μπορούσε να τις χαρακτηρίσει κανείς γενικά σημαντικές, ανεξάρτητα από το πλαίσιο χρήσης. Επιπλέον, θεωρείται ότι εκφράζουν την προσωπική προτίμηση του χρήστη, καθώς είναι ρητά διαχωρισμένες από τις υπόλοιπες. Συνολικά, βρέθηκαν 93 αγαπημένες επαφές, που αποτελούν το 3,5% των επαφών των χρηστών που είχαν χρησιμοποιήσει αυτή τη δυνατότητα. Στο γράφημα 7 παρατηρεί κανείς πως το πλήθος των αγαπημένων επαφών παραμένει σχετικά σταθερό παρά την αύξηση του μεγέθους της λίστας επαφών. Δε βρέθηκε στατιστικά σημαντική συσχέτιση ανάμεσα στο πλήθος των αγαπημένων επαφών και το συνολικό μέγεθος του αποθετηρίου επαφών ($r(15)=0,21$ $p>0,46$). Κατά μέσο όρο καθένας από τους χρήστες που είχαν χρησιμοποιήσει αυτή τη δυνατότητα είχε χαρακτηρίσει 6 επαφές ως αγαπημένες ($\text{mean}=6,2$, $\text{sd}=3,16$, $\text{min}=1$, $\text{max}=11$). Το πλήθος αυτό είναι χαμηλότερο σε σχέση με τα ευρήματα στο πείραμα των Komninos και Liarakis (2009) που παρουσιάζονται αναλυτικά στο (Stefanis,

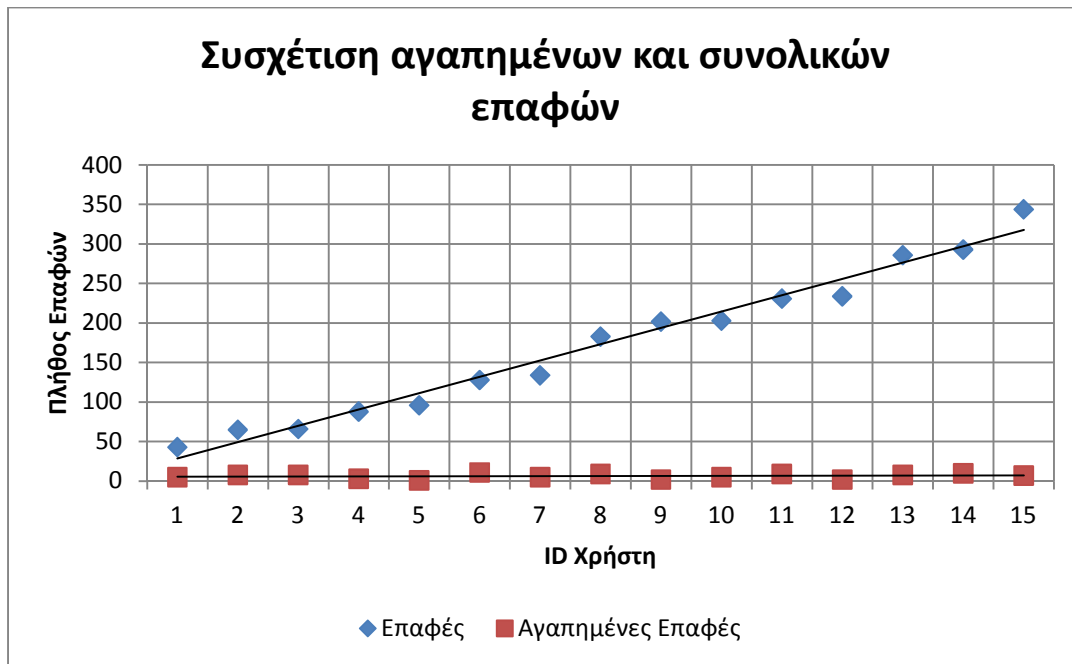
Plessas, Komninos & Garofalakis, 2012b) και τους οδήγησε να προτείνουν το μέγεθος της λίστας των σημαντικών επαφών στις 20 θέσεις. Εντούτοις, σε εκείνη την περίπτωση οι χρήστες είχαν ερωτηθεί για να υποδείξουν ρητά τις σημαντικές επαφές, ενώ στο τρέχον πείραμα τα στοιχεία αντλήθηκαν απευθείας από τις συσκευές. Μια πιθανή εξήγηση θα μπορούσε να είναι πως οι χρήστες τείνουν να υπερεκτιμούν τη σημασία κάποιων επαφών και αυτός είναι και ο λόγος που είναι απρόθυμοι να διαγράψουν εγγραφές από τη λίστα επαφών. Επίσης, οι χρήστες φαίνεται πιθανό να καθορίζουν τις αγαπημένες επαφές μια φορά και στη συνέχεια να μην αλλάζουν αυτή τη ρύθμιση. Από τα τρέχοντα αποτελέσματα φαίνεται ότι το πλήθος επαφών στο οποίο πρέπει μια λίστα να δίνει γρήγορη πρόσβαση είναι περίπου 10 επαφές. Το μέγεθος αυτό της λίστας συμφωνεί με εκείνο που προτείνεται στη μεταγενέστερη έρευνα των Bentley και Chen (2015).



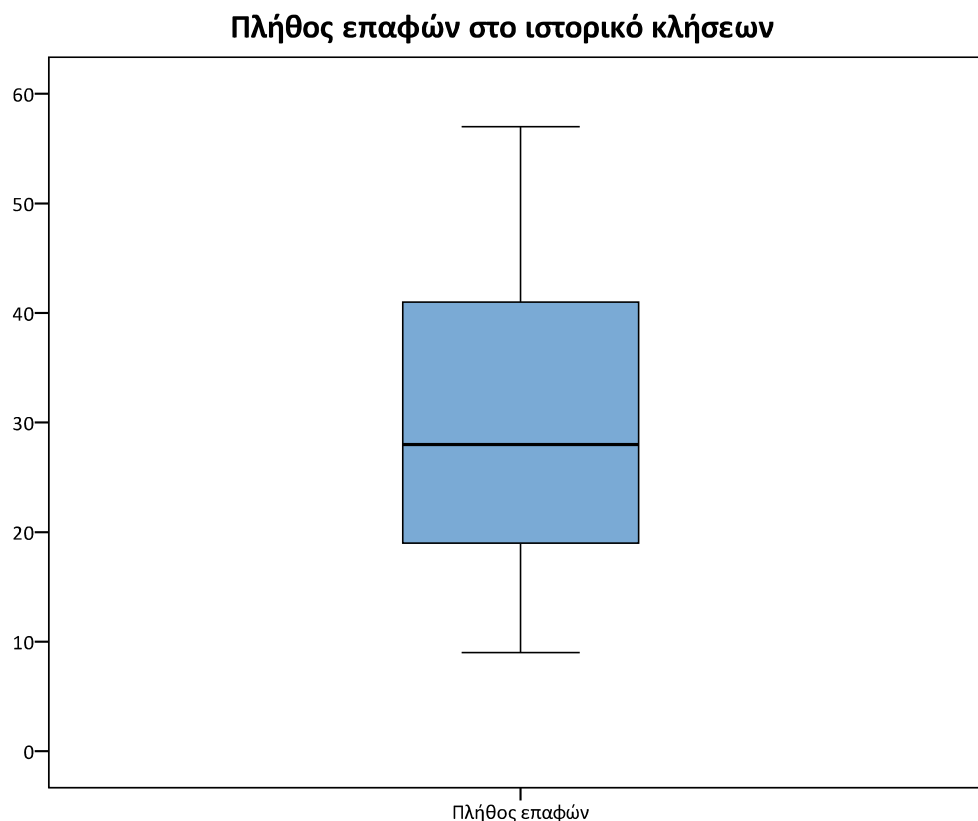
Εικόνα 23. Η λίστα με τις αγαπημένες επαφές στο Android, όπου προωθούνται στην κορυφή πριν από τις συχνά χρησιμοποιούμενες για γρήγορη επικοινωνία



Γράφημα 6. Διασπορά αγαπημένων επαφών σε σχέση με τις συνολικές επαφές



Γράφημα 7. Συσχέτιση αγαπημένων και συνολικών επαφών



Γράφημα 8. Πλήθος διακριτών επαφών στο ιστορικό κλήσεων (τελευταίες 500 κλήσεις)

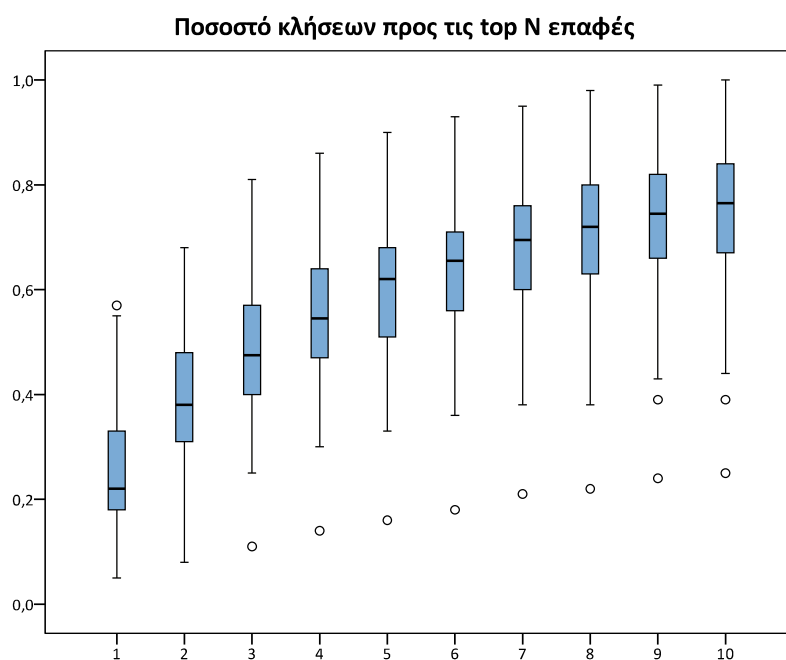
Επειδή οι συσκευές Android περιορίζουν το μέγεθος των αρχείων κλήσεων στις 500 εγγραφές και καθώς η διάρκεια που καλύπτει το ιστορικό αυτό διαφέρει από χρήστη σε χρήστη, ανάλογα με την πυκνότητα της επικοινωνιακής τους συμπεριφορά, δεν

ήταν εφικτό να υπολογιστούν και να συγκριθούν με άλλες έρευνες οι μη χρησιμοποιημένες επαφές ανά μηνιαία χρονικά διαστήματα. Συνολικά, σε αυτές τις 500 κλήσεις αντιπροσωπεύονται κατά μέσο όρο μόνο 30 διακριτές επαφές (mean=30,38, sd=13,45, min=9, max=57) (το 19% των συνολικών επαφών, θυμίζοντας την αρχή Pareto, όπως παρατηρήθηκε και στο (Lugano, 2008)). Η συνολική επικοινωνία κάθε χρήστη πραγματοποιείται με έναν μικρό αριθμό από επαφές, όπως φαίνεται και στο γράφημα 8. Στη μεταγενέστερη έρευνα των Bentley και Chen (2015) οι επαφές που είναι παρούσες στο ιστορικό κλήσεων είναι ακόμα λιγότερες (κατά μέσο όρο 17 και συνολικά το 6% των επαφών), αλλά αυτό πιθανώς οφείλεται στο μεγάλο πλήθος επαφών που έχουν εισαχθεί από κοινωνικά δίκτυα. Στον πίνακα 8 παρουσιάζονται το πλήθος και το ποσοστό των επαφών με τις οποίες υπήρξε τηλεφωνική επικοινωνία στο αρχείο κλήσεων για κάθε χρήστη, ταξινομημένες με βάση την περίοδο που κάλυπτε το ιστορικό κλήσεων. Φαίνεται πως οι περισσότεροι χρήστες έχουν ένα σημαντικό ποσοστό επαφών με το οποίο δεν έχουν επικοινωνήσει κατά τις τελευταίες 500 κλήσεις, γεγονός που επιβεβαιώνει τη δυσκολία που επιβάλλουν οι μη χρησιμοποιούμενες επαφές κατά την αναζήτηση στα ολοένα αυξανόμενα σε μέγεθος αποθετήρια. Μάλιστα, όπως και στο (Bergman, Komninos, Liarokaris & Clarke, 2012), βρέθηκε μια πολύ ισχυρή συσχέτιση ανάμεσα στο μέγεθος των αποθετηρίων και το πλήθος των μη χρησιμοποιούμενων επαφών ($r(34)=0,992$ $p<0,01$).

Πίνακας 8. Επαφές με τις οποίες υπήρξε τηλεφωνική επικοινωνία ανά χρήστη για τη διάρκεια που καλύπτει το αρχείο κλήσεων

ID Χρήστη	Πλήθος επαφών	Πλήθος επαφών με τηλεφωνική επικοινωνία	Ποσοστό επαφών με τηλεφωνική επικοινωνία	Διάρκεια αρχείου κλήσεων (σε μήνες)
8	134	10	7,46%	0,28
25	77	19	24,68%	0,34
28	96	35	36,46%	0,36
10	285	25	8,77%	0,47
15	33	9	27,27%	0,57
35	65	18	27,69%	0,60
43	231	30	12,99%	0,62
23	165	19	11,52%	0,62
4	293	57	19,45%	0,68
36	128	26	20,31%	0,70
33	183	39	21,31%	0,78
44	43	17	39,53%	0,81
37	132	10	7,58%	1,01
3	138	43	31,16%	1,02

39	234	25	10,68%	1,17
14	239	56	23,43%	1,23
38	96	25	26,04%	1,24
34	125	37	29,60%	1,31
42	344	54	15,70%	1,36
1	192	44	22,92%	1,46
2	286	44	15,38%	1,49
22	149	24	16,11%	1,51
9	203	33	16,26%	1,57
6	247	48	19,43%	1,70
5	66	33	50,00%	1,75
19	68	28	41,18%	1,94
24	291	41	14,09%	2,14
13	59	18	30,51%	2,47
17	130	22	16,92%	2,50
40	88	29	32,95%	2,57
26	202	28	13,86%	3,44
41	206	49	23,79%	3,85
18	72	25	34,72%	5,66
21	79	13	16,46%	6,23

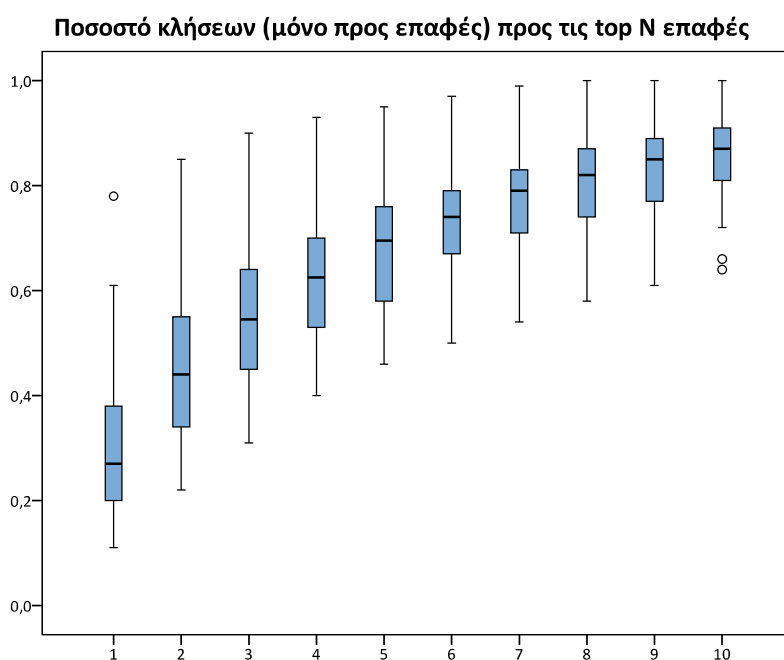


Γράφημα 9. Ποσοστό κλήσεων προς τις top N επαφές σε συχνότητα επικοινωνίας

Όπως προέκυψε από την ανάλυση των κλήσεων (γράφημα 9) με τις top N επαφές, σχεδόν το 59% (mean=58,94%, sd=15,11%) της επικοινωνίας των χρηστών μέσω τηλεφωνικών κλήσεων πραγματοποιείται με τις 5 πιο συχνά χρησιμοποιούμενες επαφές, ενώ σχεδόν το 72% (mean=72,71%, sd=16,25%) της τηλεφωνικής

επικοινωνίας ανταλλάσσεται με τις 10 πιο συχνά χρησιμοποιούμενες επαφές. Οι Bentley και Chen (2015) ανέφεραν ακόμα μεγαλύτερα ποσοστά (84% και 94% αντίστοιχα).

Στο γράφημα 10, εμφανίζεται και πάλι η κατανομή του ποσοστού των κλήσεων από και προς τις N πιο συχνά χρησιμοποιούμενες επαφές, χωρίς να λαμβάνονται υπόψη επικοινωνίες με αριθμούς που δεν ανήκουν σε επαφές και δεν έχουν αποθηκευθεί στο αποθετήριο. Σε αυτή την περίπτωση, το ποσοστό κλήσεων που ανταλλάσσεται με τις 5 πιο συχνά χρησιμοποιούμενες επαφές είναι σχεδόν 70% (mean=69,18%, sd=12,63%), ενώ όταν N=10 το ποσοστό φθάνει το 85% (mean=85,44, sd=9,36%).

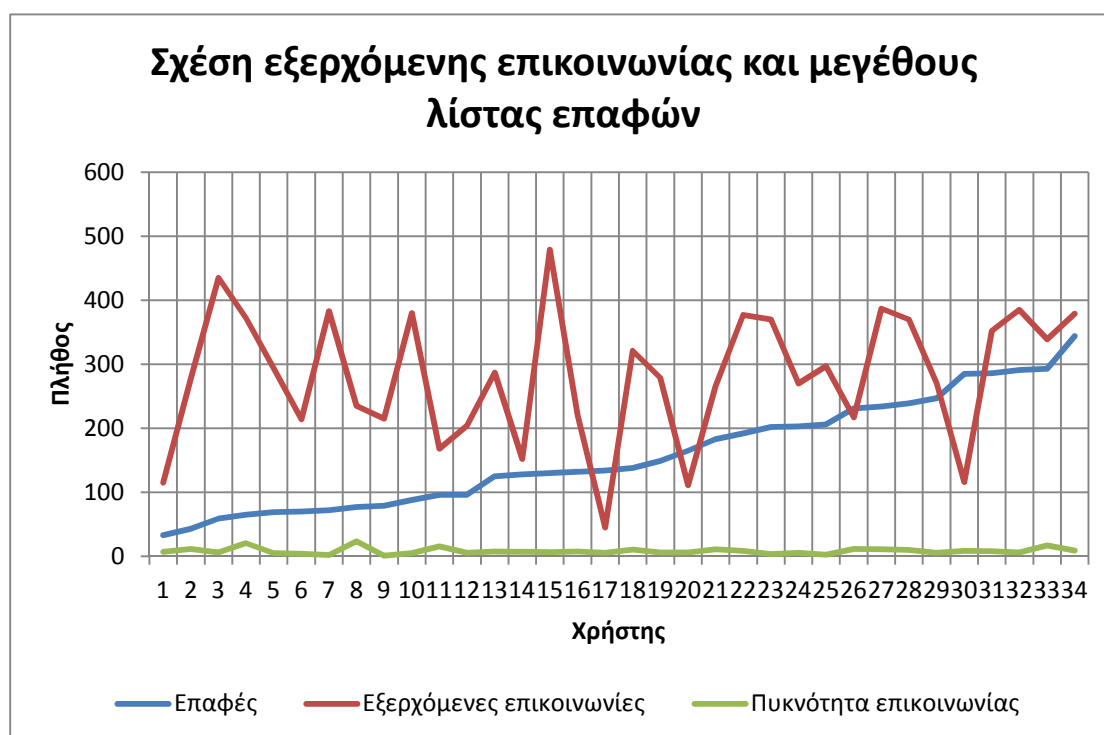


Γράφημα 10. Ποσοστό κλήσεων προς τις top N επαφές σε συχνότητα επικοινωνίας, χωρίς να υπολογίζονται οι κλήσεις προς και από νούμερα που δεν ανήκουν στις επαφές

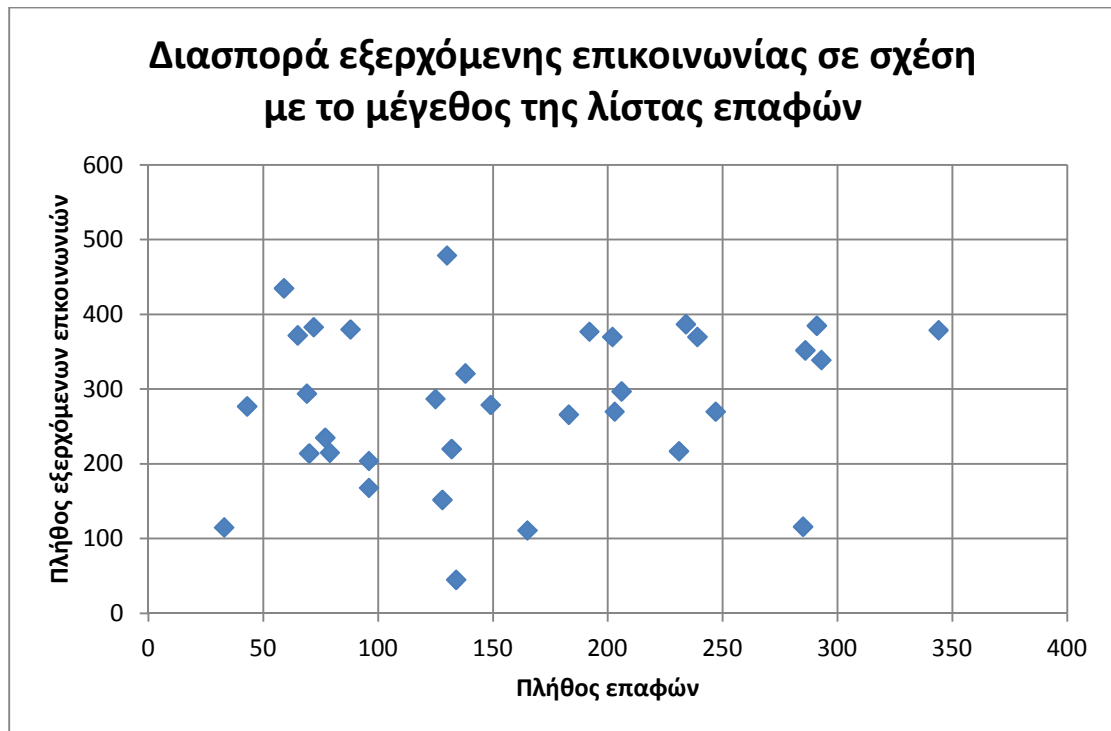
Συνολικά στο σύνολο δεδομένων περιλαμβάνονται 13.669 κλήσεις, από τις οποίες 4.705 (34,42%) ήταν απαντημένες εισερχόμενες, 7.204 (52,70%) εξερχόμενες και 1.760 (12,88%) μη απαντημένες εισερχόμενες (missed calls). Αντίστοιχα ποσοστά βρέθηκαν και στη μελέτη των Bentley και Chen (2015). Μελετώντας τις κλήσεις που πραγματοποιήθηκαν από και προς νούμερα που δεν ανήκουν σε επαφές, βρέθηκε ότι αυτές ήταν 2.119, δηλαδή το 16% των συνολικών κλήσεων, αρκετά μικρότερο ποσοστό από το 31% που ανέφεραν οι Bentley και Chen (2015). Οι εξερχόμενες κλήσεις προς νούμερα που δεν ανήκουν σε επαφές ήταν 1.117 και αποτελούν το 15,5% των συνολικών εξερχόμενων κλήσεων. Όπως και στη μελέτη των Bentley και Chen (2015), από τις κλήσεις προς νούμερα που δεν έχουν αποθηκευθεί σε κάποια επαφή υπερτερούν ελαφρώς οι εξερχόμενες κλήσεις (52,71%) σε σχέση με τις

εισερχόμενες (47,29%). Αντίθετα, οι εισερχόμενες κλήσεις από μη αποθηκευμένους αριθμούς φαίνεται ότι απαντήθηκαν από τους χρήστες στη μεγάλη πλειοψηφία τους (69,86%).

Επειδή, όπως έχει ήδη αναφερθεί, τα αρχεία κλήσεων των συμμετεχόντων κάλυπταν διαφορετικής διάρκειας χρονικά διαστήματα σε ημέρες ($\text{mean}=48,88$, $\text{sd}=41,66$, $\text{min}=8$, $\text{max}=187$), υπολογίστηκε η πυκνότητα της εξερχόμενης επικοινωνίας (πλήθος εξερχόμενων επικοινωνιών / ημέρες που κάλυπτε το αρχείο κλήσεων), συμπεριλαμβανομένων των εξερχόμενων μηνυμάτων, ως δείκτης της καθημερινής εξερχόμενης χρήσης της συσκευής. Στα γραφήματα 11 και 12 παρουσιάζεται η σχέση ανάμεσα στην πυκνότητα της εξερχόμενης επικοινωνίας, το πλήθος των εξερχόμενων επικοινωνιών και το μέγεθος της λίστας επαφών. Είναι εμφανές ότι ο τρόπος με τον οποίο οι χρήστες χρησιμοποιούν τη συσκευή τους για να επικοινωνήσουν δεν εξαρτάται από το μέγεθος του βιβλίου επαφών. Η παρατήρηση αυτή επιβεβαιώνεται και στατιστικά με την εφαρμογή της μετρικής συσχέτισης Spearman's Rank Order, τόσο για τη σχέση ανάμεσα στις εξερχόμενες επικοινωνίες και το μέγεθος της λίστας επαφών ($r_s(34)=0,170$, $p=0,338$), όσο και για τη σχέση ανάμεσα στην πυκνότητα της εξερχόμενης επικοινωνίας και το μέγεθος της λίστας επαφών ($r_s(34)=0,154$, $p=0,383$).

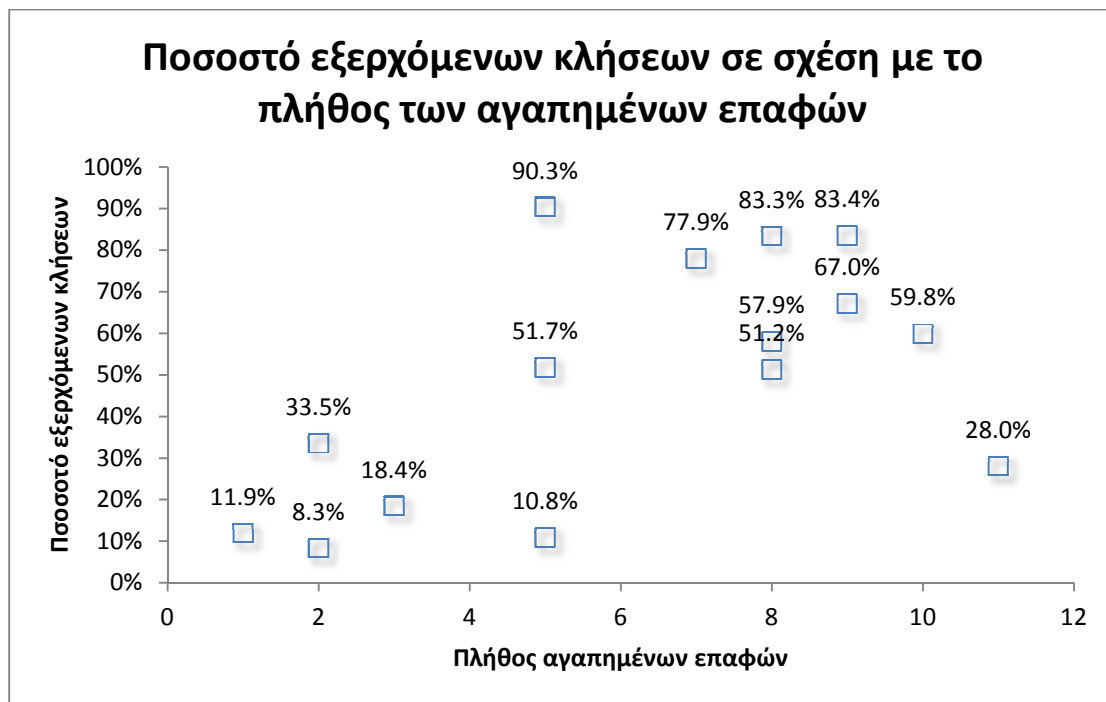


Γράφημα 11. Σχέση εξερχόμενης επικοινωνίας και μεγέθους λίστας επαφών



Γράφημα 12. Διασπορά εξερχόμενης επικοινωνίας σε σχέση με το μέγεθος της λίστας επαφών

Επίσης, για τους χρήστες που είχαν χαρακτηρίσει επαφές ως αγαπημένες, μελετήθηκε το ποσοστό εξερχόμενων κλήσεων προς τις επαφές αυτές (γράφημα 13). Ένα ενδιαφέρον εύρημα ήταν πως για τους περισσότερους χρήστες, το ποσοστό των εξερχόμενων κλήσεων προς τις αγαπημένες επαφές ήταν σημαντικά υψηλό. Ειδικότερα, για το 60% των συμμετεχόντων που χρησιμοποιούσαν τη δυνατότητα του Android να ορίζουν επαφές ως αγαπημένες, περισσότερο από 50% των εξερχόμενων κλήσεών τους ήταν προς αγαπημένες επαφές. Αυτό είναι ένα γεγονός που ενισχύει τη σημασία της προσωπικής προτίμησης ως διάσταση πλαισίου σχετική με τις επαφές. Από την άλλη, υπάρχει ένας αριθμός από συμμετέχοντες που δε φαίνεται να επικοινωνούν ιδιαίτερα με τις αγαπημένες τους επαφές. Ενδεχομένως, η ίδια εξήγηση σχετικά με τη ρύθμιση αυτής της δυνατότητας μόνο μια φορά στην αρχή να ισχύει και σε αυτή την περίπτωση (οι περισσότεροι από αυτούς τους συμμετέχοντες είχαν τις λιγότερες αγαπημένες επαφές). Επιπλέον, το γεγονός ότι κάποιος θεωρεί μια επαφή σημαντική στη ζωή του δε σημαίνει απαραίτητα ότι επικοινωνεί μαζί της συχνά στο τηλέφωνο. Για παράδειγμα, ένας από τους συμμετέχοντες είχε τον πατέρα του ως αγαπημένη επαφή, αλλά δήλωσε ότι δε χρησιμοποιεί συχνά τη συσκευή τηλεφώνου του για να επικοινωνήσει μαζί του, αφού ζουν στο ίδιο μέρος.



Γράφημα 13. Ποσοστό εξερχόμενων κλήσεων σε σχέση με το πλήθος των αγαπημένων επαφών

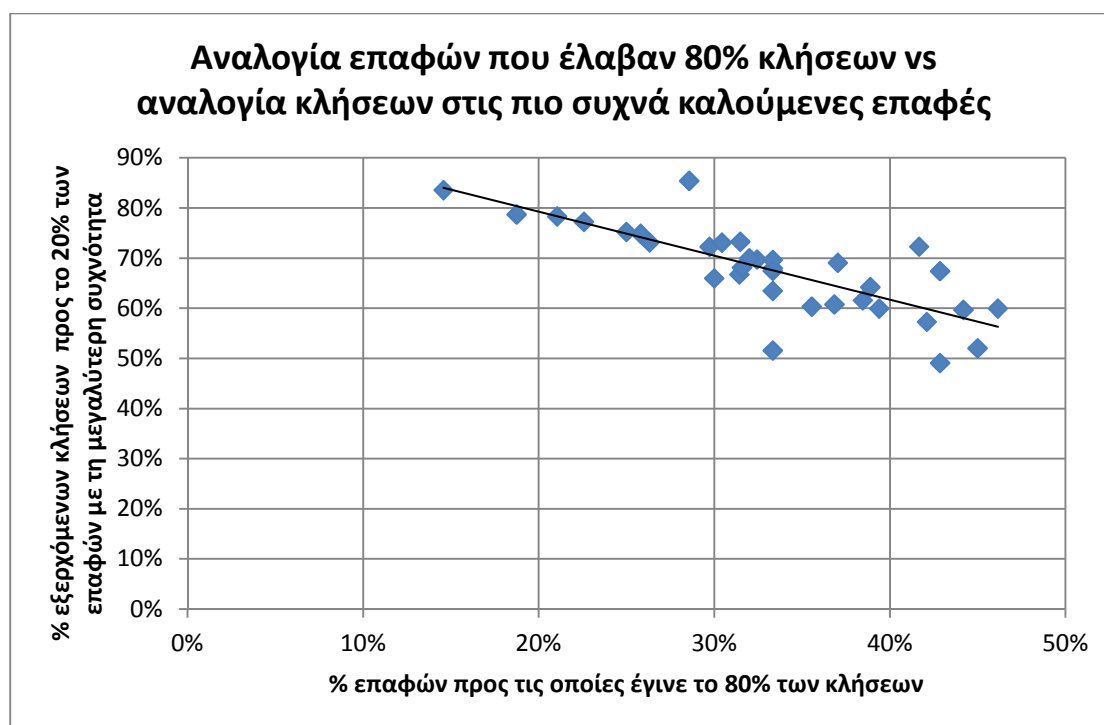
Οι συσκευές Android επιτρέπουν την οργάνωση των επαφών σε ομάδες. Εντούτοις, πρόκειται για μια δυνατότητα που δε χρησιμοποιούσαν οι περισσότεροι από τους συμμετέχοντες, καθώς βρέθηκαν ομάδες μόνο σε 7 από αυτούς. Συνολικά, καταγράφησαν 16 ομάδες ($\text{mean}=2,28$, $\text{sd}=2,21$, $\text{min}=1$, $\text{max}=7$). Οι 6 από τους 7 συμμετέχοντες είχαν μια ομάδα με όνομα «Φίλοι», ενώ 2 είχαν οργανώσει επαφές σε ομάδα με όνομα «Οικογένεια». Οι υπόλοιπες ομάδες αφορούσαν γεωγραφικές περιοχές και συναδέλφους. Συνολικά 142 επαφές ήταν οργανωμένες σε αυτές τις ομάδες ($\text{mean}=8,87$, $\text{sd}=5,86$, $\text{min}=1$, $\text{max}=21$). Αναλυτικά οι ομάδες των επαφών φαίνονται στον πίνακα 9. Μια ενδιαφέρουσα παρατήρηση είναι ότι ο χρήστης με ID 13 έχει κατατάξει τις 54 από τις 59 επαφές του σε κάποια ομάδα, ενώ κάποιες από αυτές τις επαφές τις έχει οργανώσει σε περισσότερες από μια ομάδες. Ο συγκεκριμένος χρήστης έχει πάντως αρκετά μικρότερο πλήθος επαφών σε σχέση με το μέσο όρο και αυτό είναι σύμφωνα με τη βιβλιογραφία ένα χαρακτηριστικό των χρηστών που διατηρούν οργανωμένη τη λίστα επαφών (π.χ. σβήνουν επαφές περιοδικά (Bergman, Komninos, Liarokapis & Clarke, 2012)). Λόγω του μικρού δείγματος χρηστών που χρησιμοποιούσαν αυτή τη δυνατότητα, αλλά και των διαφοροποιήσεων στην οργάνωση των επαφών από κάθε χρήστη, δεν είχε νόημα η περαιτέρω ανάλυση της επικοινωνιακής συμπεριφοράς τους με τις επαφές που είναι οργανωμένες σε ομάδες. Μια εποπτική παρατήρηση, πάντως, δεν έδειξε κάποια εμφανή συσχέτιση με τη συχνότητα επικοινωνίας.

Πίνακας 9. Ομάδες επαφών (contact groups)

ID Χρήστη	Όνομα Ομάδας	Πλήθος Επαφών Ομάδας	Συνολικές Επαφές Χρήστη
1	Σπίτι	7	192
	Φίλοι	15	
4	Friends	7	293
	VIP	1	
	Family	5	
8	ΑΡΣΗΥ	12	134
9	Friends	1	203
13	Πλάτανος	11	59
	Family	12	
	Παιανία	9	
	Πάτρα	21	
	Coworkers	15	
	B.U.D.	4	
	Friends	15	
34	Friends	2	125
40	Friends	5	88

Στη συνέχεια, εξετάστηκε αν κάποια μορφή της αρχής Pareto (Pareto, 1935) ισχύει για τα δεδομένα επικοινωνίας. Γι' αυτό το λόγο, υπολογίστηκε για κάθε χρήστη το ποσοστό των επαφών προς τις οποίες πραγματοποιείται το 80% της εξερχόμενης επικοινωνίας και το ποσοστό των εξερχόμενων κλήσεων που έγιναν προς το 20% των επαφών με την υψηλότερη συχνότητα επικοινωνίας. Σε αυτή την ανάλυση δε λήφθησαν υπόψη επαφές από τα αποθετήρια των χρηστών που δεν είχαν κληθεί καθόλου κατά την περίοδο που κάλυπτε το αρχείο κλήσεων. Μόνο το 15,92% (857) των επαφών στις λίστες επαφών όλων των συμμετεχόντων ήταν παρούσες στα αρχεία κλήσεων, ενώ οι υπόλοιπες 84,08% δεν είχαν χρησιμοποιηθεί καθόλου. Τα αποτελέσματα για κάθε συμμετέχοντα εμφανίζονται στο γράφημα 14. Μια πρώτη παρατήρηση είναι πως η αναλογία κλήσεων προς τις 20% συχνότερα καλούμενες επαφές μειώνεται καθώς η αναλογία των επαφών που αντιστοιχούν στο 80% των κλήσεων αυξάνει. Πρόκειται για μια αναμενόμενη συμπεριφορά και παρατηρεί κανείς ότι υπάρχει μια εμφανής αρνητική συσχέτιση σε αυτή την περίπτωση ($r(34)=-0,78$ $p<0,01$). Παρότι η αρχή Pareto δε φαίνεται να προκύπτει ακριβώς στη μορφή του κανόνα 80/20, για τους περισσότερους συμμετέχοντες φαίνεται να υπάρχει μια σχετικά σταθερή σχέση: το 80% των εξερχόμενων κλήσεων απευθύνονται συνήθως στο 25%-35% των επαφών και το 65%-75% των εξερχόμενων κλήσεων πραγματοποιείται με το 20% των επαφών με την υψηλότερη συχνότητα. Αυτό το εύρημα είναι ενδεικτικό της υψηλής σημασίας της συχνότητας επικοινωνίας ως παράγοντα πλαισίου για την αξιοποίηση των επαφών. Επιπροσθέτως, φαίνεται ότι

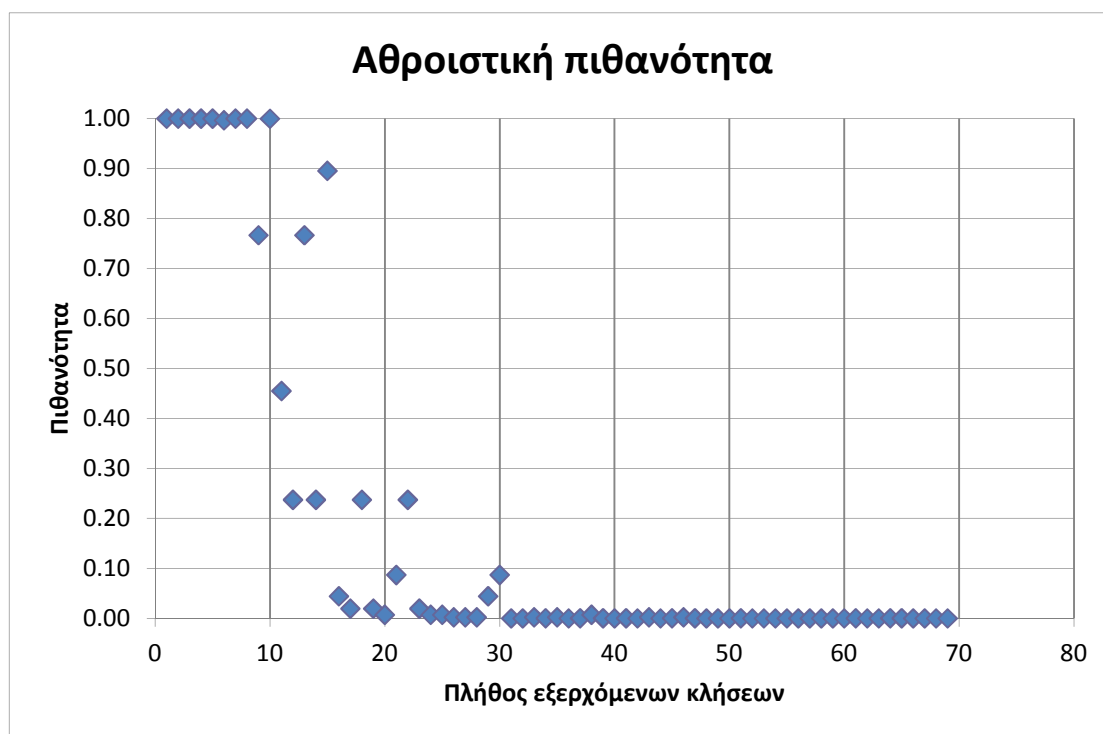
ένα σχετικά μικρό ποσοστό επαφών χρησιμοποιούνται συχνά (υπενθυμίζεται ότι στην ανάλυση δε συμπεριελήφθησαν επαφές που δεν είχαν χρησιμοποιηθεί καθόλου), ενώ οι χρήστες πρέπει να ψάξουν σε ένα μεγάλο αποθετήριο (βιβλίο επαφών) για να ανακτήσουν την απαιτούμενη εγγραφή, ένα γεγονός που οι διεπαφές των λιστών επαφών χρειάζεται να λαμβάνουν υπόψη.



Γράφημα 14. Αναλογία επαφών που έλαβαν 80% κλήσεων προς την αναλογία κλήσεων στις πιο συχνά καλούμενες επαφές

Στη συνέχεια, μελετήθηκε η σχέση ανάμεσα στις συχνότητες κλήσεων και το πλήθος των επαφών που αντιστοιχούν σε κάθε τιμή συχνότητας (δηλαδή το πλήθος των επαφών που έχουν κληθεί μια φορά, δύο φορές κ.ο.κ.). Για να επιτευχθεί αυτό, καθώς τα αρχεία κλήσεων κάλυπταν διαφορετικές χρονικές περιόδους, εφαρμόστηκε κανονικοποίηση στα δεδομένα προεκτείνοντας την επικοινωνιακή συμπεριφορά για μια σταθερή χρονική περίοδο ίση με το μέσο όρο συν μια τυπική απόκλιση για όλους τους χρήστες (δηλαδή 89 ημέρες, $\text{mean}=48$ ημέρες, $\text{sd}=41$). Έπειτα, ταξινομήθηκαν τα ζεύγη των συχνοτήτων κλήσεων και του πλήθους των επαφών με βάση τη συχνότητα κλήσης και υπολογίστηκε η αθροιστική πιθανότητα για κάθε ζεύγος. Όπως φαίνεται στο γράφημα 15, τα δεδομένα μπορούν να χωριστούν σε τρεις διακριτές περιοχές υψηλής, χαμηλής και απρόβλεπτης πιθανοτικής συμπεριφοράς. Ειδικότερα, για την περιοχή χαμηλής πιθανοτικής συμπεριφοράς, αυτή διακρίνεται μετά το πλήθος των 31 κλήσεων. Φαίνεται ότι οι επαφές που έχουν δεχτεί μεγαλύτερο πλήθος κλήσεων ανήκουν σε μια ξεχωριστή ομάδα σπάνιων επαφών. Στην πραγματικότητα, αυτή η

ομάδα αντιπροσωπεύει μόνο το 7,3% του συνόλου των επαφών με τουλάχιστον μια κλήση στην επεκτεταμένη περίοδο (832 επαφές). Η περιοχή υψηλής πιθανότητας αντιπροσωπεύει το 71,2% των επαφών, ενώ η ενδιάμεση περιοχή απροσδιόριστης πιθανότητας το 21,5%. Προφανώς, δεν μπορεί να υποστηριχθεί ότι οι επαφές που ανήκουν στην περιοχή χαμηλής πιθανότητας είναι συνεχώς σημαντικές επειδή έχουν δεχθεί μεγάλο αριθμό από κλήσεις. Πρέπει, όμως, τουλάχιστον για κάποια περίοδο στο ιστορικό της καθημερινής ζωής των χρηστών να υπήρξαν σημαντικές για να δεχθούν τόσο μεγάλο αριθμό κλήσεων.



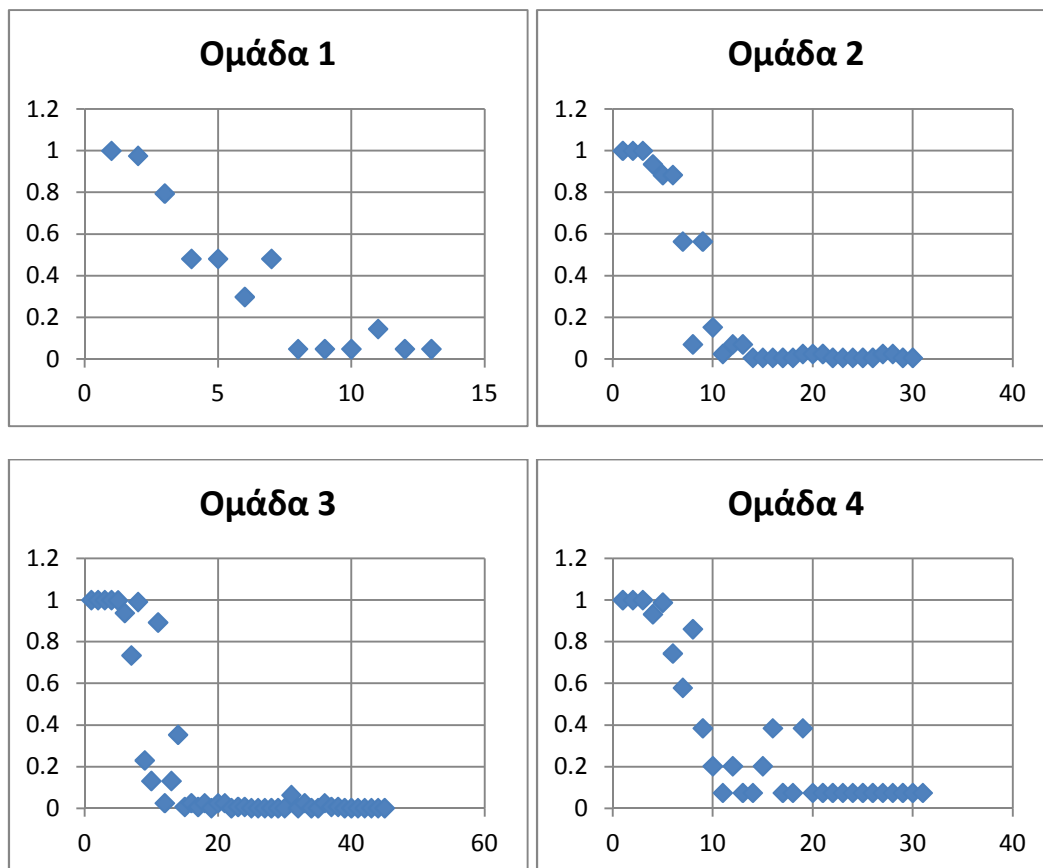
Γράφημα 15. Κατανομή αθροιστικής πιθανότητας για τα ζεύγη του αριθμού των εξερχόμενων κλήσεων και του πλήθους των επαφών

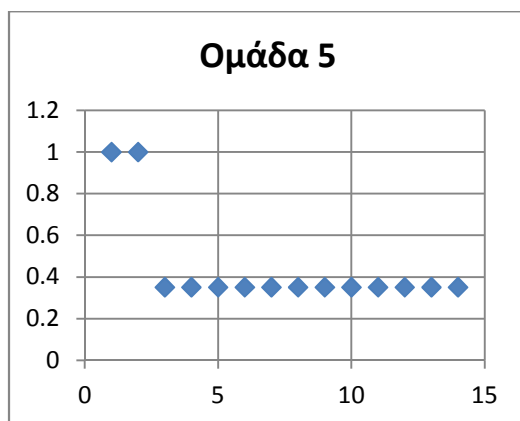
Παρόμοια συμπεριφορά στην κατανομή των αθροιστικών πιθανοτήτων προέκυψε ακόμα και όταν τα δεδομένα δεν κανονικοποιήθηκαν, όπως στην προηγούμενη περίπτωση. Για να διαπιστωθεί εάν αυτό το πρότυπο συμπεριφοράς επαναλαμβανόταν για τις διαφορετικές διάρκειες κάλυψης των αρχείων κλήσεων, το σύνολο δεδομένων διαχωρίστηκε σε 5 διακριτές ομάδες ανάλογα με το διάστημα που κάλυπτε το ιστορικό κλήσης (πίνακας 10) και εφαρμόστηκε η ίδια ανάλυση για κάθε ομάδα. Από την ανάλυση παρατηρήθηκε το ίδιο πρότυπο συμπεριφοράς στην κατανομή πιθανοτήτων για κάθε ομάδα, ανεξάρτητα από το μέγεθος της αντίστοιχης χρονικής περιόδου (γράφημα 16). Μια ιδιαίτερη παρατήρηση είναι το γεγονός πως το μέγεθος της περιοχής χαμηλής αθροιστικής πιθανότητας (και επομένως των επαφών με υψηλή πιθανότητα κλήσης) παραμένει το ίδιο (γράφημα 17), περίπου γύρω από το

σημείο του 11% (mean=10,93%, sd=1,7%). Τα ευρήματα για την ομάδα 5 δείχνουν μια ακανόνιστη συμπεριφορά, αλλά αυτό οφείλεται στο γεγονός ότι η ομάδα περιλαμβάνει μόνο δύο χρήστες. Αυτά τα αποτελέσματα, υποδεικνύουν ένα κρίσιμο σημείο για τις επαφές, όπου πραγματοποιείται μια μετάβαση από την περιοχή των κοινών στην περιοχή των σπάνιων (και κατ' επέκταση σημαντικών), σε σχέση με τη συχνότητα χρήσης τους.

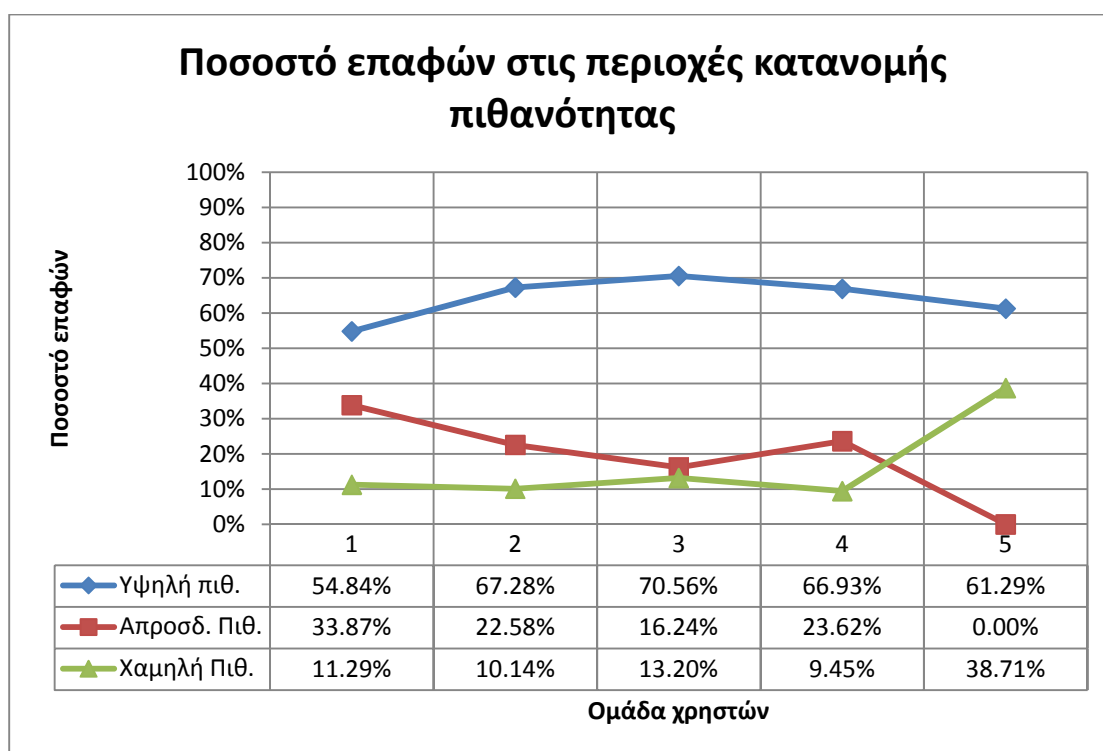
Πίνακας 10. Ομάδες χρηστών ανάλογα με τη χρονική διάρκεια του ιστορικού κλήσεων

Ομάδα	Χρήστες	Ελαχ. Ημέρες	Μεγ. Ημέρες	Μ.Ο. Ημερών	Ημέρες Τυπ. Απόκλισης
1	4	8	14	11	7
2	10	17	31	24	15
3	13	35	64	49.5	30
4	5	74	115	94.5	42
5	2	170	187	178.5	18





Γράφημα 16. Κατανομή αθροιστικής πιθανότητας για τα ζεύγη του αριθμού των εξερχόμενων κλήσεων και του πλήθους των επαφών, ανά ομάδα χρηστών

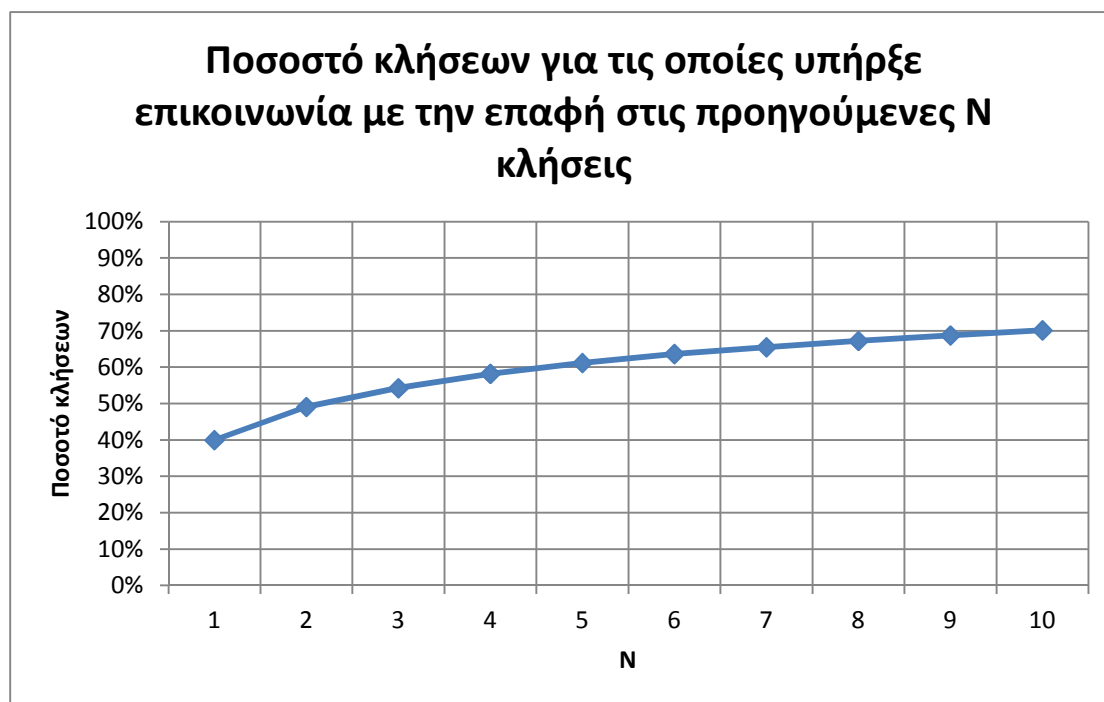


Γράφημα 17. Ποσοστό επαφών στην υψηλή, απροσδιόριστη και χαμηλή περιοχή κατανομής πιθανότητας ανά ομάδα χρηστών

Πρόκειται για ένα σημαντικό εύρημα, επειδή δείχνει ότι μπορεί να διαχωριστεί το πλήθος των επαφών που είναι εν δυνάμει σημαντικές ως προς το κριτήριο της συχνότητας για οποιοδήποτε χρήστη, άσχετα από το σύνολο δεδομένων που καλύπτει οποιαδήποτε υπό εξέταση χρονική περίοδο. Συνεπώς, μπορούν να εντοπιστούν οι επαφές που είναι «καθολικά» σημαντικές για τον χρήστη με βάση το κριτήριο της συχνότητας επικοινωνίας, για ολόκληρο το ιστορικό κλήσεων του, καθώς και εκείνες οι επαφές που είναι «τοπικά» σημαντικές όσον αφορά προκαθορισμένα χρονικά πλαίσια (π.χ. αυτή την εβδομάδα, τον προηγούμενο μήνα).

Επιπροσθέτως, με βάση το γράφημα 17 και θεωρώντας ότι η συχνότητα επικοινωνίας είναι το πρωτεύον κριτήριο, μπορεί να περιοριστεί περαιτέρω η εφαρμογή άλλων κριτηρίων μόνο στο σώμα των επαφών που είναι σημαντικές για το τοπικό πλαίσιο.

Στη συνέχεια, υπολογίστηκε το ποσοστό των κλήσεων για τις οποίες υπήρξε επικοινωνία με την κληθείσα επαφή κατά τις προηγούμενες N κλήσεις, για $N=1\ldots 10$ (γράφημα 18). Βρέθηκε ότι για το 40% περίπου των κλήσεων που περιλαμβάνονται στο σύνολο δεδομένων, η προηγούμενη τηλεφωνική επικοινωνία ήταν με την ίδια επαφή. Καθώς το N αυξάνει, το ποσοστό αυτό αυξάνεται με μειούμενο ρυθμό, έως περίπου το 70% των κλήσεων για $N=10$. Φαίνεται, συνεπώς, ότι η λίστα πρόσφατων κλήσεων που διαθέτουν οι περισσότερες συσκευές μπορεί να διευκολύνει την ανάκτηση των επαφών το πολύ στο 70% των περιπτώσεων. Οι Bentley και Chen (2015) αργότερα βρήκαν ότι σε αυτή τη λίστα ($N=10$) περιλαμβάνονται μόνο 6 από τις περισσότερες χρησιμοποιούμενες επαφές των χρηστών.



Γράφημα 18. Ποσοστό κλήσεων για τις οποίες υπήρξε επικοινωνία με την επαφή στις προηγούμενες N κλήσεις

3.2.3 Συμπεράσματα

Από το πείραμα που πραγματοποιήθηκε εξήχθησαν χρήσιμα συμπεράσματα για τη σύνθεση του αποθετηρίου επαφών και τα πρότυπα επικοινωνίας των χρηστών σε συσκευές κινητού υπολογισμού.

Αρχικά, επιβεβαιώθηκε το γεγονός ότι οι χρήστες για την ανάκτηση μιας επαφής χρειάζεται να πραγματοποιήσουν αναζήτηση σε ένα αρκετά μεγάλο σώμα από

εγγραφές (158 κατά μέσο όρο). Το πλήθος αυτό των επαφών είναι σημαντικά μεγαλύτερο σε σχέση με προηγούμενες μελέτες, ενώ μεταγενέστερες αναλύσεις έδειξαν ότι αυξάνεται με γρήγορους ρυθμούς (φτάνοντας τις 308 επαφές κατά μέσο όρο στην πιο πρόσφατη ανάλυση των Bentley και Chen (2015)). Επιπλέον, ένα σημαντικό μέρος από αυτές τις επαφές δε χρησιμοποιείται τόσο συχνά ή και καθόλου, καθώς μόνο το 20% των συνολικών επαφών ήταν παρούσες στα αρχεία κλήσεων των χρηστών. Ουσιαστικά, το εύρημα αυτό επιβεβαιώνει τους Bergman, Komninos, Liarokapis και Clarke (2012), οι οποίοι εφάρμοσαν την αρχή του υποβιβασμού, για να υποβιβάσουν οπτικά τις μη χρησιμοποιούμενες επαφές και να προβιβάσουν αυτές που χρησιμοποιούνται. Αντίστοιχα συμπεράσματα για τις μη χρησιμοποιούμενες επαφές προέκυψαν και από τα αποτελέσματα που παρουσίασαν οι Bentley και Chen (2015). Συνεπώς, επιβεβαιώθηκε το γεγονός πως ένα μεγάλο μέρος των στοιχείων που περιλαμβάνονται στο αποθετήριο των επαφών δε χρησιμοποιούνται, γεγονός το οποίο δυσκολεύει σημαντικά την ανάκτηση λόγω των περιορισμών που χαρακτηρίζουν τις οθόνες και τις μεθόδους εισόδου των συσκευών κινητού υπολογισμού.

Στη συνέχεια, αναλύθηκαν διεξοδικά τα πρότυπα επικοινωνίας και εντοπίστηκαν διαστάσεις πλαισίου που σχετίζονται με την επικοινωνιακή συμπεριφορά των χρηστών. Πρόκειται για την πρώτη ανάλυση αυτού του είδους, καθώς στη βιβλιογραφία προτείνονται και χρησιμοποιούνται διαστάσεις πλαισίου σε προσαρμοστικές διεπαφές που διευκολύνουν την ανάκτηση των επαφών, χωρίς να παρουσιάζεται κάποια εις βάθος έρευνα και αιτιολόγηση. Ο εντοπισμός των διαστάσεων πλαισίου που μπορούν να χρησιμοποιηθούν για την αξιοποίηση των επαφών είναι μια σημαντική συνεισφορά της διδακτορικής έρευνας.

Η συχνότητα επικοινωνίας είναι ένας σημαντικός παράγοντας χρησιμοποίησης των επαφών, καθώς βρέθηκε ότι το μεγαλύτερο μέρος της εξερχόμενης επικοινωνίας των χρηστών περιλαμβάνει ένα μικρό μόνο αριθμό επαφών από το συνήθως μεγάλο πλήθος που περιέχεται στο αποθετήριο. Είναι ενδεικτικό ότι σχεδόν το 59% της τηλεφωνικής επικοινωνίας των χρηστών πραγματοποιείται με τις 5 πιο συχνά χρησιμοποιούμενες επαφές, ενώ το 72% με τις 10 πιο συχνά χρησιμοποιούμενες επαφές. Τα ποσοστά αυτά (που αυξάνονται σε 70% και 85% αντίστοιχα όταν δε λαμβάνονται υπόψη κλήσεις με αριθμούς που δεν αντιστοιχούν σε επαφές) είναι χαμηλότερα από αυτά που αναφέρουν οι Bentley και Chen (2015) και υποδηλώνουν ότι ένα σημαντικό ποσοστό της επικοινωνίας των χρηστών από τις συσκευές κινητού υπολογισμού ανταλλάσσεται με επαφές που δεν περιλαμβάνονται στη λίστα πιο συχνά χρησιμοποιούμενων επαφών. Μια μορφή της αρχής Pareto φαίνεται να ισχύει,

καθώς το 80% των εξερχόμενων κλήσεων απευθύνονται συνήθως στο 25%-35% των επαφών και το 65%-75% των εξερχόμενων κλήσεων πραγματοποιείται με το 20% των επαφών με την υψηλότερη συχνότητα.

Η προσφατότητα της επικοινωνίας είναι ακόμα ένας παράγοντας που παίζει σημαντικό ρόλο στην ανάκτηση και χρησιμοποίηση των επαφών. Από την ανάλυση που πραγματοποιήθηκε βρέθηκε ότι για το 40% περίπου των κλήσεων η αμέσως προηγούμενη κλήση ήταν με την ίδια επαφή. Ελέγχοντας ακόμα περισσότερες από τις προηγούμενες επικοινωνίες, το ποσοστό αυτό αυξάνει σχεδόν στο 70% (για τις 10 προηγούμενες κλήσεις). Εντούτοις, υπάρχει ένα σημαντικό ποσοστό κλήσεων για τις οποίες η προσφατότητα δεν αποτελεί επαρκή παράγοντα πλαισίου. Επομένως, σε αυτές τις περιπτώσεις, η χρησιμοποίηση προσαρμοστικών διεπαφών που βασίζονται μόνο στην προσφατότητα (όπως είναι η λίστα πρόσφατων κλήσεων) αποτυγχάνει στην ανάκτηση της επιθυμητής επαφής. Μάλιστα, όπως παρατήρησαν και οι Bentley και Chen (2015), η λίστα πρόσφατων κλήσεων μπορεί να είναι μια καλή επιλογή για την επικοινωνία με «τοπικά» σημαντικές επαφές, αλλά δεν προσφέρει εύκολη πρόσβαση στις «καθολικά» σημαντικές επαφές, δηλαδή σε αυτές που χρησιμοποιούνται πιο συχνά.

Επιπλέον, η προσωπική προτίμηση όπως εκφράζεται μέσω του χαρακτηρισμού επαφών ως αγαπημένων φαίνεται να είναι ένας σημαντικός παράγοντας πλαισίου για τη χρησιμοποίηση επαφών, για κάποιους από τους χρήστες που συμμετείχαν στο πείραμα που διοργανώθηκε. Οι επαφές αυτές μπορεί να θεωρηθεί ότι αποτελούν τις «καθολικά» σημαντικές επαφές, ανεξάρτητα από τη συχνότητα επικοινωνίας. Το μέσο πλήθος τους ανά χρήστη (περίπου 6) δείχνει ότι μια προσαρμοστική διεπαφή δε χρειάζεται να έχει περισσότερες από 10 θέσεις, ανεξάρτητα από το μέγεθος του αποθετηρίου, όπως πρότειναν στη συνέχεια και οι Bentley και Chen (2015) και σε αντίθεση με τους Komninos και Liaroikaris (2009) που είχαν θέσει αυτό το όριο στις 20 επαφές. Το συμπέρασμα αυτό ενισχύεται και από το υψηλό ποσοστό κλήσεων με τις 10 περισσότερο χρησιμοποιούμενες επαφές. Οι περισσότεροι από τους χρήστες που είχαν χαρακτηρισμένες επαφές ως αγαπημένες, πραγματοποιούσαν ένα μεγάλο μέρος της επικοινωνίας τους με αυτές τις επαφές και ειδικά όσοι είχαν περισσότερες από 5 αγαπημένες επαφές. Εντούτοις, βρέθηκαν και χρήστες (κυρίως όσοι είχαν λιγότερες από 5 αγαπημένες επαφές) οι οποίοι δεν επικοινωνούσαν σε μεγάλο ποσοστό με αυτές τις επαφές.

Μια ακόμα σημαντική συνεισφορά αυτής της έρευνας ήταν η απόδειξη πως η επικοινωνιακή συμπεριφορά των χρηστών παρουσιάζει ένα επαναλαμβανόμενο

πρότυπο χρήσης, το οποίο μπορεί να αξιοποιηθεί για το σχεδιασμό διεπαφών ανάκτησης που χρησιμοποιούν το πλαίσιο, ανεξάρτητα από τη χρονική κλίμακα των παρατηρούμενων αλληλεπιδράσεων. Όπως έχει ήδη αναφερθεί, η συχνότητα επικοινωνίας, η προσφατότητα και η προσωπική προτίμηση μπορεί να μην είναι τα μόνα κριτήρια για τη σημαντικότητα των επαφών. Η έρευνα που παρουσιάστηκε ανοίγει ερωτήματα σχετικά με άλλες πληροφορίες πλαισίου, όπως η δραστηριότητα, οι προγραμματισμένες εργασίες ή η τοποθεσία, που όμως δεν ήταν διαθέσιμες στο υπό εξέταση σύνολο δεδομένων. Σε επόμενο κεφάλαιο μελετάται ξεχωριστά η επίπτωση της τοποθεσίας και του χρονικού πλαισίου, καθώς οι συμμετέχοντες υπέδειξαν ότι οι συνήθειές τους, όσον αφορά τις κλήσεις, εξαρτώνται από την τοποθεσία στην οποία βρίσκονται. Για παράδειγμα, όταν βρίσκονται στη δουλειά τείνουν να επικοινωνούν με συναδέλφους. Τα απογεύματα, όταν βρίσκονται σπίτι ή για διασκέδαση επικοινωνούν σπάνια για θέματα εργασίας, περνώντας τον περισσότερο χρόνο επικοινωνίας μιλώντας με φίλους και συγγενείς. Οι προγραμματισμένες εργασίες που μπορούν να αντληθούν από το ημερολόγιο των χρηστών και σχετίζονται με κάποια επαφή θα μπορούσαν, επίσης, να χρησιμοποιηθούν για να αυξηθεί προσωρινά η σημαντικότητα της επαφής, ακόμα κι αν πρόκειται για επαφή που δε χρησιμοποιείται τακτικά. Πάντως, το μικρό πλήθος διαστάσεων πλαισίου που προέκυψαν ως σχετικές με το πρόβλημα της αξιοποίησης των επαφών από την ανάλυση των δεδομένων του πειράματος, εξάλειψε, τουλάχιστον σε αυτή τη φάση, την ανάγκη για την εφαρμογή τεχνικών μείωσης διαστάσεων στο διάνυσμα αναπαράστασης του πλαισίου.

3.3 Σύνοψη

Για την αναπαράσταση του πλαισίου προτάθηκε ένα πολυδιάστατο διανυσματικό μοντέλο, ώστε να αξιοποιηθεί ο συνδυασμός των διαστάσεων πλαισίου που σχετίζονται με τη χρησιμοποίηση των επαφών. Οι υποψήφιες διαστάσεις περιλαμβάνουν τη συχνότητα και την προσφατότητα της επικοινωνίας, την τοποθεσία, την προσωπική προτίμηση, το χρονικό πλαίσιο, τις προγραμματισμένες εργασίες κ.α. Ένα τέτοιο μοντέλο διευκολύνει και την εφαρμογή τεχνικών μείωσης διαστάσεων (όπως ο αλγόριθμος Fastmap, ο οποίος φαίνεται να είναι κατάλληλος για συσκευές κινητού υπολογισμού) που μπορεί να οδηγήσουν σε χρήσιμες οπτικοποιήσεις δύο και τριών διαστάσεων.

Ένα πείραμα κατά το οποίο αναλύθηκαν τα αποθετήρια επαφών και τα πρότυπα επικοινωνίας των χρηστών που συμμετείχαν, επιβεβαίωσε το μεγάλο πλήθος επαφών στο οποίο χρειάζεται να πραγματοποιηθεί η αναζήτηση για την έναρξη μιας επικοινωνίας, καθώς και το μεγάλο πλήθος μη χρησιμοποιούμενων επαφών.

Επιπλέον, τα αποτελέσματα του πειράματος οδήγησαν στο συμπέρασμα πως οι υπάρχουσες διεπαφές αξιοποίησης των επαφών σε συσκευές κινητού υπολογισμού (λίστα πρόσφατων κλήσεων, αγαπημένων επαφών και συχνά χρησιμοποιούμενων επαφών) αδυνατούν να εξυπηρετήσουν τις ανάγκες των χρηστών.

Από την ανάλυση των δεδομένων προέκυψε πως η συχνότητα και η προσφατότητα επικοινωνίας αποτελούν σημαντικές διαστάσεις πλαισίου για το υπό εξέταση πρόβλημα. Επίσης, η προσωπική προτίμηση είναι μια διάσταση που φαίνεται να παίζει σημαντικό ρόλο για τους περισσότερους από τους χρήστες που χρησιμοποιούσαν την αντίστοιχη δυνατότητα των συσκευών να ορίζουν επαφές ως αγαπημένες.

4. Αξιοποίηση Συχνότητας και Προσφατότητας ως Παραγόντων Κοινωνικού Πλαισίου Συσκευών Κινητού Υπολογισμού

Η διαδικασία αξιοποίησης των επαφών για την έναρξη μιας εξερχόμενης επικοινωνίας στις συσκευές κινητού υπολογισμού περιλαμβάνει την ανάκτηση της επιθυμητής επαφής από μια από τις διαθέσιμες διεπαφές, όπως είναι η λίστα επαφών, η λίστα πρόσφατων κλήσεων ή η λίστα συχνά χρησιμοποιούμενων επαφών. Για τη βελτίωση της ανάκτησης, δηλαδή την πιο εύκολη και γρήγορη εύρεση της επιθυμητής επαφής, μπορεί να εφαρμοστεί ένας αλγόριθμος πρόβλεψης που υπολογίζει για κάθε επαφή την πιθανότητα να είναι αυτή που θα χρειαστεί ο χρήστης και παρουσιάζει τις πιο πιθανές επαφές. Στην ουσία, σε μια τέτοια προσέγγιση εφαρμόζεται κάποια τεχνική υπολογισμού της πιθανότητας χρησιμοποίησης της επαφής, συνήθως αξιοποιώντας το πλαίσιο και στη συνέχεια το πρόβλημα μετατρέπεται σε πρόβλημα ταξινόμησης των επαφών και επιλογής. Στα πλαίσια της διδακτορικής έρευνας έγινε προσπάθεια να απαντηθεί το πρόβλημα της ανάκτησης επαφών όταν πραγματοποιείται μια εξερχόμενη κλήση. Παρότι η επικοινωνία στις κινητές συσκευές λαμβάνει χώρα συχνά μέσω διάφορων καναλιών (π.χ. Skype, Facebook κ.λπ.) και όχι μόνο μέσω κλήσεων και μηνυμάτων, υπάρχει αδυναμία συλλογής όλων αυτών των δεδομένων επικοινωνίας από τις διαφορετικές εφαρμογές για τις συσκευές κινητού υπολογισμού, λόγω περιορισμών στα δικαιώματα πρόσβασης που αυτές θέτουν.

Σε αυτό το κεφάλαιο παρουσιάζεται αναλυτικά η προτεινόμενη μέθοδος πρόβλεψης και παραγωγής προτάσεων, τα πειραματικά αποτελέσματα από την εφαρμογή της σε δύο διαφορετικά σύνολα δεδομένων μέσω προσομοίωσης και τα αποτελέσματα από την εφαρμογή της σε ένα πείραμα πεδίου.

4.1 Αλγόριθμος αξιοποίησης πλαισίου για την πρόβλεψη εξερχόμενων κλήσεων

Στο προηγούμενο κεφάλαιο, η ανάλυση των αποθετηρίων επαφών των χρηστών και των προτύπων επικοινωνίας τους έδειξε ότι η συχνότητα και η προσφατότητα της επικοινωνίας είναι δύο παράγοντες πλαισίου που παίζουν σημαντικό ρόλο στην επικοινωνιακή συμπεριφορά τους. Επομένως, οι δύο αυτοί παράγοντες ταυτοποιήθηκαν ως διαστάσεις του πλαισίου που μπορούν να χρησιμοποιηθούν σε έναν αλγόριθμο πρόβλεψης των επαφών που είναι πιο πιθανό να κληθούν ανά πάσα στιγμή. Επίσης, για κάποιους χρήστες βρέθηκε ότι η προσωπική προτίμηση, όπως

εκφράζεται με τον χαρακτηρισμό των επαφών ως αγαπημένων, σχετίζεται με υψηλή συχνότητα επικοινωνίας, αν και για άλλους δε φαίνεται να ισχύει κάτι τέτοιο, γεγονός που καθιστά αυτή τη διάσταση διφορούμενη. Καθώς, ο αριθμός αυτών των χρηστών ήταν μικρός σε σχέση με το συνολικό δείγμα συμμετεχόντων, αποφασίστηκε προς το παρόν να μην ενσωματωθεί αυτή η διάσταση στον αλγόριθμο πρόβλεψης. Άλλοι παράγοντες, όπως η τοποθεσία δεν ήταν δυνατό να αξιολογηθούν, καθώς δεν υπήρχαν διαθέσιμα δεδομένα. Τελικά, ο αλγόριθμος που περιγράφεται στη συνέχεια αξιοποιεί τις διαστάσεις της συχνότητας και προσφατότητας της επικοινωνίας από το κινητό κοινωνικό πλαίσιο για να προτείνει τις επαφές που είναι πιο πιθανό να κληθούν σε μια χρονική στιγμή t . Η ιδέα της αξιοποίησης αυτών των δυο διαστάσεων για το σχεδιασμό μιας προσαρμοζόμενης λίστας επαφών είχε παρουσιαστεί και στο (Jung, Anttila & Blom, 2008).

Σύμφωνα με το μοντέλο αναπαράστασης του πλαισίου που προτάθηκε στο προηγούμενο κεφάλαιο, τα αντικείμενα προσωπικής πληροφορίας αναπαρίστανται ως διανύσματα επαυξημένα με πληροφορίες πλαισίου. Για παράδειγμα, για n διαστάσεις πλαισίου ένα αντικείμενο μπορεί να αναπαρασταθεί ως $(x_1, x_2, \dots, x_n)$, όπου x_i η τιμή της διάστασης i . Όταν οι διαστάσεις είναι η συχνότητα και η προσφατότητα της επικοινωνίας, το πλαίσιο κάθε επαφής c_k από το σύνολο των επαφών $C = (c_1, c_2, \dots, c_n)$ που βρίσκονται στο αποθετήριο της συσκευής του χρήστη τη χρονική στιγμή t μπορεί να αναπαρασταθεί ως

$$c_k = (f_k, r_k)$$

όπου f_k και r_k είναι οι τιμές των διαστάσεων που αντιπροσωπεύουν τη συχνότητα και την προσφατότητα της επικοινωνίας για την επαφή αυτή. Για κάθε διάσταση υπάρχει μια συνάρτηση σκορ (score function) που υπολογίζει την τιμή της από το ιστορικό κλήσεων του χρήστη τη χρονική στιγμή t . Για τη συχνότητα επικοινωνίας η συνάρτηση είναι η $F_t(c_k)$ και υπολογίζει το ποσοστό των κλήσεων με την επαφή c_k , ενώ για την προσφατότητα είναι η $R_t(c_k)$ και υπολογίζει τη χρονική απόσταση από την τελευταία κλήση προς την επαφή. Επομένως, το πλαίσιο για την επαφή c_k αναπαρίσταται ως

$$c_k = (F_t(c_k), R_t(c_k))$$

Σε δύο ερευνητικές εργασίες (Phithakkitnukoon & Dantu, 2008; Phithakkitnukoon, Dantu, Claxton & Eagle, 2011) έχει αποδειχθεί ότι μόνο ένα μέρος των πρόσφατων δεδομένων από το ιστορικό των χρηστών είναι επαρκές για την πρόβλεψη της μελλοντικής συμπεριφοράς τους. Επομένως, η συνάρτηση υπολογισμού της συχνότητας δε χρειάζεται να αναλύσει ολόκληρο το ιστορικό κλήσεων των χρηστών,

αλλά μόνο ένα πρόσφατο μέρος. Για το σκοπό αυτό ορίζεται ένα παράθυρο εκπαίδευσης t_w σε ημέρες, από τη χρονική τιμή t και προς το παρελθόν, τις κλήσεις του οποίου χρησιμοποιεί η συνάρτηση F_t για να υπολογίσει το ποσοστό επικοινωνίας με κάθε επαφή. Επιπλέον, η συνάρτηση λαμβάνει υπόψη της τόσο τις εισερχόμενες όσο και τις εξερχόμενες κλήσεις που έχουν πραγματοποιηθεί με κάθε επαφή, ώστε να συλλάβει την επίπτωση της αμοιβαιότητας (Phithakkitnukoon, Dantu, Claxton & Eagle, 2011). Τελικά, δεδομένων του συνόλου επαφών C ενός χρήστη και του παραθύρου εκπαίδευσης t_w , η συνάρτηση F_t ορίζεται ως

$$F_t: C \rightarrow [0,1]$$

$$F_t(c_k) = \frac{\sum \text{κλήσεων με τη } c_k \text{ στο } t_w}{\sum \text{κλήσεων στο } t_w}$$

Εντός του παραθύρου εκπαίδευσης t_w ορίζεται ένα παράθυρο προσφατότητας r_w σε ώρες, το οποίο περιέχει τις κλήσεις από την αρχή του t_w (δηλαδή από τη χρονική στιγμή t) και προς τα πίσω στο χρόνο. Στο προηγούμενο κεφάλαιο, από την ανάλυση των προγενέστερων N κλήσεων πριν από κάθε κλήση προέκυψε πως το όφελος για την εύρεση της επαφής που καλείται μειώνεται διαρκώς, καθώς αυξάνεται το πλήθος των προγενέστερων κλήσεων που αναλύονται. Επομένως, για τον υπολογισμό της τιμής της διάστασης της προσφατότητας επικοινωνίας δε χρειάζεται να αναλυθούν όλες οι κλήσεις μέσα στο διάστημα t_w , αλλά μόνο ένας μικρός αριθμός από προηγούμενες κλήσεις και συγκεκριμένα όσες βρίσκονται στο διάστημα r_w . Η χρήση ενός χρονικού παραθύρου, αντί για ένα παράθυρο που βασίζεται στον αριθμό των προγενέστερων κλήσεων, προτιμήθηκε για να αντισταθμιστεί το φαινόμενο μεγάλων χρονικών διαστημάτων χωρίς να έχουν πραγματοποιηθεί κλήσεις, όπως για παράδειγμα κατά τη διάρκεια της νύχτας. Οι κλήσεις στο διάστημα r_w χρησιμοποιούνται για να δώσουν προσωρινή σημασία στις επαφές, ενώ οι κλήσεις στο διάστημα t_w για να τους δώσουν καθολική σημασία. Η συνάρτηση $R_t(c_k)$ υπολογίζει την τιμή της προσφατότητας ως το ποσοστό του χρόνου που έχει μεσολαβήσει από την αρχή του r_w (δηλαδή από τη χρονική στιγμή t) έως τη χρονική στιγμή που έχει πραγματοποιηθεί η πιο πρόσφατη επικοινωνία με την επαφή c_k μέσα στο r_w . Δεδομένου του r_w , η συνάρτηση R_t ορίζεται ως

$$R_t: C \rightarrow [0,1]$$

$$R_t(c_k) = 1 - \frac{t - t_{\text{πιο πρόσφατη επικοινωνία με την επαφή στο } r_w}}{r_w}$$

Στη συνέχεια, πραγματοποιείται μίξη των επιμέρους σκορ των διαστάσεων του πλαισίου (score-based fusion (Hsu, Chung & Kristal, 2006))), οπότε σε κάθε επαφή ανατίθεται ένα συνολικό σκορ ως εξής:

$$\Pi_t(c_k) = w_f \times F_t(c_k) + w_r \times R_t(c_k)$$

Τα w_f και w_r είναι τα βάρη στάθμισης των διαστάσεων. Επειδή το πεδίο τιμών και των δύο συναρτήσεων σκορ F_t και R_t είναι το $[0,1]$ δεν απαιτείται κανονικοποίηση των επιμέρους σκορ πριν από τη μίξη.

Τέλος, οι επαφές ταξινομούνται με βάση το συνολικό σκορ σε φθίνουσα κατάταξη και ο αλγόριθμος προτείνει τις top-k επαφές ως τις πιο πιθανές να κληθούν τη χρονική στιγμή t .

Τα βήματα του αλγορίθμου παρουσιάζονται συνοπτικά στον πίνακα 11.

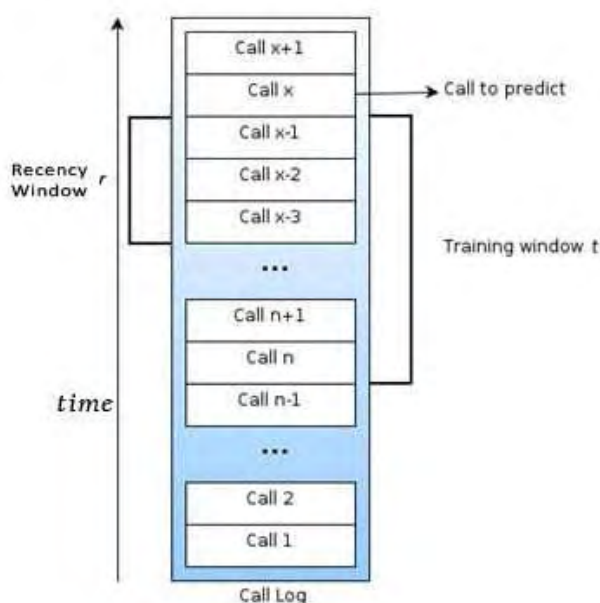
Πίνακας 11. Αλγόριθμος αξιοποίησης πλαισίου για την πρόβλεψη εξερχόμενων κλήσεων

<i>Αλγόριθμος πρόβλεψης πιθανότερων επαφών για έναρξη εξερχόμενης κλήσης</i>	
Παράμετροι: παράθυρο εκπαίδευσης t_w , παράθυρο προσφατότητας r_w , βάρος διάστασης συχνότητας επικοινωνίας w_f , βάρος διάστασης προσφατότητας επικοινωνίας w_r , πλήθος επαφών k .	
Τη χρονική στιγμή t εκτέλεσης του αλγορίθμου:	
1.	Υπολογισμός του σκορ της διάστασης συχνότητας επικοινωνίας $F_t(c_k)$ για κάθε επαφή c_k που έχει κληθεί εντός του παραθύρου εκπαίδευσης t_w
2.	Υπολογισμός του σκορ της διάστασης προσφατότητας επικοινωνίας $R_t(c_k)$ για κάθε επαφή c_k που έχει κληθεί εντός του παραθύρου προσφατότητας r_w
3.	Υπολογισμός του συνολικού σκορ $\Pi_t(c_k) = w_f \times F_t(c_k) + w_r \times R_t(c_k)$ για κάθε επαφή c_k που έχει κληθεί εντός του παραθύρου εκπαίδευσης t_w
4.	Ταξινόμηση των επαφών σε φθίνουσα σειρά με βάση το συνολικό σκορ τους
5.	Επιλογή των k κορυφαίων επαφών από την ταξινομημένη λίστα

4.2 Προκαταρκτικό πείραμα αξιολόγησης αλγορίθμου

Για την αξιολόγηση του αλγορίθμου εφαρμόστηκε ένα πείραμα προσομοίωσης (Plessas, Stefanis, Komninos & Garofalakis, 2013) με τα αρχεία κλήσεων των χρηστών του συνόλου δεδομένων. Η διαδικασία της προσομοίωσης για κάθε χρήστη παρουσιάζεται στην εικόνα 24 και έχει ως εξής: έστω ότι πρέπει να προβλεφθεί ποια επαφή κλήθηκε στη θέση x του αρχείου κλήσεων. Αυτή η αλληλεπίδραση πρέπει να αναπαριστά μια εξερχόμενη κλήση, καθώς ο αλγόριθμος υποστηρίζει την ανάκτηση των επαφών και επομένως δεν πραγματοποιεί προβλέψεις εισερχόμενων κλήσεων. Τότε, συγκεντρώνονται όλες οι κλήσεις που έχουν λάβει χώρα τις προηγούμενες t_w ημέρες από τη χρονική στιγμή που αντιστοιχεί στη θέση x . Αυτό το παράθυρο t_w περιλαμβάνει ένα μεταβλητό πλήθος από κλήσεις, οι οποίες χρησιμοποιούνται ως

δεδομένα εκπαίδευσης με βάση τα οποία πραγματοποιείται η πρόβλεψη. Οι κλήσεις αυτές αναλύονται για να υπολογιστεί το σκορ της συχνότητας επικοινωνίας για κάθε επαφή. Επίσης, συγκεντρώνονται όλες οι κλήσεις που έχουν γίνει t_w ώρες πριν από τη χρονική στιγμή που αντιστοιχεί στη θέση x και υπολογίζεται το σκορ της προσφατότητας της επικοινωνίας για κάθε επαφή. Τα επιμέρους σκορ συνδυάζονται και αφού παραχθεί μια πρόβλεψη, συγκρίνεται με την πραγματική κλήση που έγινε στη θέση x και καταγράφεται το αποτέλεσμα. Με αυτόν τον τρόπο, διατρέχεται όλο το αρχείο κλήσεων του χρήστη, κλήση προς κλήση και γίνεται προσπάθεια επιτυχούς πρόβλεψης της κάθε μιας (προφανώς οι κλήσεις των πρώτων t_w ημερών από το ιστορικό χρησιμοποιούνται μόνο ως δεδομένα εκπαίδευσης και όχι για την πραγματοποίηση προβλέψεων).



Εικόνα 24. Σχηματική αναπαράσταση της διαδικασίας εκτέλεσης της προσομοίωσης εφαρμογής του αλγορίθμου πρόβλεψης

Πριν την πραγματοποίηση των πειραμάτων, χρειαζόταν να καθοριστεί ένα κατάλληλο μήκος για το παράθυρο εκπαίδευσης, όπως και για το παράθυρο προσφατότητας. Η εμπειρική παρατήρηση των αρχείων κλήσεων των χρηστών έδειξε ότι η προσφατότητα της επικοινωνίας έχει επιρροή, με φθίνουσα επίπτωση, για μια περίοδο 6 ωρών. Ο επιπλέον πειραματισμός με τα διαθέσιμα δεδομένα οδήγησε σε ελαφρώς καλύτερα αποτελέσματα για ένα χρονικό διάστημα 12 ωρών, επομένως το μήκος του παραθύρου της προσφατότητας τέθηκε στις 12 ώρες. Οι Phithakkitnukoon, Dantu, Claxton και Eagle (2011) έδειξαν ότι η ακρίβεια του αλγορίθμου τους για την πρόβλεψη κλήσεων δε βελτιώθηκε σύμφωνα με το μέγεθος των δεδομένων εκπαίδευσης. Αυτό είναι ένα λογικό αποτέλεσμα, αφού καθώς οι

άνθρωποι αλλάζουν τα πρότυπα συμπεριφοράς τους και ενδεχομένως αλληλεπιδρούν πιο κοντά με διαφορετικές κοινωνικές ομάδες με την πάροδο του χρόνου, τα παλιότερα δεδομένα όχι μόνο γίνονται πλεονάζοντα, αλλά μπορεί να έχουν και αρνητικό αντίκτυπο στην επιτυχία των προβλέψεων. Στην περίπτωση που παρουσιάζεται, επειδή, όπως αναλύεται και στην επόμενη ενότητα, το μήκος των αρχείων κλήσεων ήταν 52 ημέρες κατά μέσο όρο, δεν υπήρχε η δυνατότητα για πειραματισμό με παράθυρο εκπαίδευσης πολύ μεγάλης διάρκειας. Το παράθυρο εκπαίδευσης έπρεπε να είναι αρκετά μεγάλο ώστε να παρέχει επαρκή δεδομένα, εντούτοις ένα μήκος μεγαλύτερο από 2 εβδομάδες θεωρήθηκε ότι θα αποτύγχανε να συλλάβει τις δυναμικές αλλαγές στη συμπεριφορά πραγματοποίησης κλήσεων των χρηστών (π.χ. στην περίπτωση που κάποιος έπαιρνε άδεια μιας εβδομάδας για διακοπές). Δοκιμάζοντας, λοιπόν, τον αλγόριθμο πρόβλεψης με ίσα βάρη για τις διαστάσεις της συχνότητας και προσφατότητας, για μήκος παραθύρου 10 και 15 ημέρες, βρέθηκαν κατά μέσο όρο ελαφρώς υψηλότερα ποσοστά επιτυχίας για την τιμή των 10 ημερών. Συνεπώς, το μήκος του παραθύρου εκπαίδευσης t_w τέθηκε στις 10 ημέρες.

Τέλος, χρειαζόταν να αποφασιστούν κατάλληλα μήκη για τη λίστα προτάσεων για την εκτέλεση των πειραματικών προσομοιώσεων. Οι Phithakkitnukoon, Dantu, Claxton και Eagle (2011) πειραματίστηκαν με διάφορα μεγέθη της λίστας πρόβλεψης (έως 20 θέσεις), εντούτοις θεωρήθηκε ότι μια διεπαφή για συσκευές κινητού υπολογισμού θα έπρεπε να παρέχει γρήγορη πρόσβαση στις επαφές, χωρίς να χρειάζεται κάθετη κύλιση στην οθόνη. Επιπλέον, τα αποτελέσματα ανάλυσης των προτύπων επικοινωνίας που παρουσιάστηκαν στο προηγούμενο κεφάλαιο (Stefanis, Plessas, Komninos & Garofalakis, 2012,), όπως και τα ευρήματα των Bentley και Chen (2015) οδηγούν στο σχεδιασμό διεπαφών με όχι περισσότερες από 10 θέσεις. Τελικά, για την αποφυγή της κάθετης κύλισης αποφασίστηκε ο πειραματισμός με έως 8 προβλέψεις, καθώς αυτό το πλήθος επαφών μπορεί να εμφανιστεί σε μια μόνο οθόνη συσκευών κινητού υπολογισμού. Τελικά, πραγματοποιήθηκαν πειράματα με 1 (κατευθείαν επιτυχία ή αποτυχία), 3 (λογικός αριθμός για widgets οθονών κινητών συσκευών), 5 (για απευθείας σύγκριση με άλλες εργασίες (π.χ. Lee, Seo & Lee, 2010; Barzaiq & Loke, 2011; Phithakkitnukoon, Dantu, Claxton & Eagle, 2011)) και 8 (μέγιστος αριθμός επαφών που χωράνε σε μια οθόνη κινητής συσκευής) θέσεις.

4.2.1 Σύνολο δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε για την πειραματική αξιολόγηση του αλγορίθμου μέσω προσομοίωσης ήταν αυτό που παρουσιάστηκε στο προηγούμενο κεφάλαιο για την ανάλυση του αποθετηρίου επαφών και των προτύπων επικοινωνίας

των χρηστών. Αντί για 34 χρήστες, όμως, χρησιμοποιήθηκαν τα δεδομένα μόνο από 25 (22 άνδρες και 3 γυναίκες, με ηλικίες από 19 έως 39 ετών), καθώς εφαρμόστηκαν επιπλέον κριτήρια για το φιλτράρισμα χρηστών με χαρακτηριστικά που τους καθιστούσαν μη έγκυρους για συμμετοχή στη διαδικασία αξιολόγησης. Πιο συγκεκριμένα, τα ακριβή κριτήρια που εφαρμόστηκαν για να συμπεριληφθούν τα δεδομένα των χρηστών στην ανάλυση ήταν:

- το πλήθος των επαφών (μεγαλύτερο από 20)
- το πλήθος των κλήσεων στο αρχείο κλήσεων (περισσότερες από 150)
- η περίοδος που κάλυπτε το αρχείο κλήσεων (περισσότερες από 15 ημέρες)

Συνολικά, οι λίστες επαφών των χρηστών περιελάμβαναν 4.185 εγγραφές. Κατά μέσο όρο, κάθε λίστα επαφών περιείχε 167 επαφές (mean=167,4, sd=87,6, min=33, max=344). Δεκατρείς από τους χρήστες είχαν χρησιμοποιήσει τη δυνατότητα των συσκευών να ορίζουν επαφές ως αγαπημένες. Τα αρχεία κλήσεων κάλυπταν διαφορετικές περιόδους σε ημέρες (mean=52,80, sd=35,23, min=18, max=170), καθώς οι συσκευές Android περιορίζουν τις κλήσεις που διατηρούνται στο ιστορικό σε 500 (mean=449,88, sd=98,12, min=182, max=500). Τα χαρακτηριστικά των χρηστών φαίνονται αναλυτικά στον πίνακα 12.

Πίνακας 12. Χαρακτηριστικά των χρηστών του συνόλου δεδομένων

ID Χρήστη	Επαφές	Αγαπημένες	Κλήσεις	Ημέρες
2	286	8	499	45
4	293	10	500	20
5	70	8	366	52
6	247		500	51
9	203	5	500	47
13	59		500	74
14	239		500	37
17	130		500	75
18	72		383	170
19	69		500	58
22	149		500	45
24	291		499	64
26	202	2	490	103
33	183	9	500	24
34	125		500	39
35	65	8	498	18
36	128	11	182	21
37	132		203	30
38	96		270	37
39	234	2	499	35

40	88	3	500	77
41	206		499	115
42	344	7	500	41
43	231	9	359	19
44	43	5	500	23
Mean	167,4		449,88	52,8
SD	87,60		98,12	35,23

Οι Lee, Seo και Lee (2010), σε μια αντίστοιχη μελέτη αξιολόγησης ενός αλγορίθμου πρόβλεψης κλήσεων για μια προσαρμοζόμενη διεπαφή, κατέταξαν τους χρήστες τους σε 3 ομάδες ανάλογα με την καταλληλότητα των προτάσεων της διεπαφής και παρατήρησαν διαφορετικές επικοινωνιακές συμπεριφορές για κάθε ομάδα. Είναι λογικό, ανάλογα με τον τρόπο επικοινωνίας των χρηστών, για κάποιους χρήστες να είναι πιο εύκολη η πρόβλεψη των επαφών που είναι πιο πιθανό να καλέσουν (π.χ. αν κατευθύνουν το μεγαλύτερο μέρος της επικοινωνίας τους σε ένα μικρό αριθμό επαφών), ενώ για άλλους (π.χ. αυτούς που διαμοιράζουν την επικοινωνία τους σε πολλές διαφορετικές επαφές) πιο δύσκολη. Στη συνέχεια του κειμένου, χρησιμοποιείται ο όρος «κοινωνικότητα», όχι με την κυριολεκτική έννοια, αλλά για να εκφράσει αυτό το πρότυπο της εισερχόμενης και εξερχόμενης επικοινωνίας που λαμβάνει χώρα στη συσκευή ενός χρήστη.

Ένας σημαντικός στόχος πριν την εκτέλεση των πειραμάτων ήταν η οργάνωση των χρηστών σε ομάδες, ανάλογα με την επικοινωνιακή τους συμπεριφορά και επομένως την ευκολία ή δυσκολία πρόβλεψής της. Η ανάγκη γι' αυτή την κατηγοριοποίηση ήταν ακόμα μεγαλύτερη εξαιτίας της φύσης των αρχείων κλήσεων, τα οποία ποικίλαν σε διάρκεια και πυκνότητα. Καθώς δεν ήταν επιθυμητή η τροποποίηση των αρχείων με την εφαρμογή κάποιας μορφής κανονικοποίησης, η κατάταξη των χρηστών σε ομάδες με βάση την «κοινωνικότητα» ήταν ένα σημαντικό βήμα για την εξασφάλιση της ακεραιότητας των συμπερασμάτων. Μια προφανής ιδέα ήταν η χρήση της πυκνότητας της επικοινωνίας, δηλαδή το πλήθος των κλήσεων ανά ημέρα. Όμως, μια τέτοια επιλογή δε θα αντικατόπτριζε αποτελεσματικά τη συμπεριφορά των χρηστών ως προς την «κοινωνικότητά» τους. Ένας χρήστης θα μπορούσε να πραγματοποιεί πολλές κλήσεις κάθε μέρα, αλλά αυτές να είναι προς πολύ λίγες διακριτές επαφές. Επομένως, προτάθηκε μια μετρική για την «κοινωνικότητα» των χρηστών (Κ-Μετρική), που βασίζεται στα ποσοστά κλήσεων προς κάθε επαφή και η οποία υπολογίζεται ως εξής: για κάθε χρήστη δημιουργείται μια λίστα με όλες τις επαφές προς τις οποίες έχει πραγματοποιήσει κλήσεις για το διάστημα που καλύπτει το ιστορικό κλήσεων. Για κάθε μια από αυτές τις επαφές υπολογίζεται το ποσοστό κλήσεων που της αναλογούν σε σχέση με τις συνολικές εξερχόμενες κλήσεις. Αφού

ταξινομηθούν οι επαφές σε φθίνουσα σειρά με βάση το ποσοστό κλήσεων, υπολογίζεται για κάθε επαφή η διαφορά του ποσοστού από αυτό της προηγούμενης στην κατάταξη (π.χ. αν ένας χρήστης έχει καλέσει 3 επαφές με ποσοστά 80%, 15% και 5%, οι διαφορές θα είναι 80%, 65% και 10%). Στη συνέχεια, υπολογίζεται ο μέσος όρος των διαφορών και η τιμή αυτή είναι η Κ-Μετρική, η οποία χρησιμοποιείται για να αναπαραστήσει την «κοινωνικότητα» της επικοινωνίας. Όσο χαμηλότερη είναι η τιμή της μετρικής, τόσο πιο «κοινωνικός» είναι ένας χρήστης, καθώς αυτό σημαίνει ότι ο χρήστης καλεί περισσότερες επαφές με μικρότερη συχνότητα. Ως παράδειγμα, μπορεί να θεωρήσει κανείς την παρακάτω περίπτωση: ένας χρήστης Α για μια συγκεκριμένη χρονική περίοδο επικοινωνεί μόνο με 3 επαφές (40%, 40% και 20% αντίστοιχα), ενώ για την ίδια περίοδο ένας χρήστης Β επικοινωνεί με 6 επαφές (με σχετικές συχνότητες 40%, 20%, 15%, 10%, 10% και 5%) και ένας χρήστης Γ με 10 επαφές (20%, 20%, 15%, 15%, 10%, 8%, 5%, 3%, 2%, 2%). Τότε, η μετρική της «κοινωνικότητας» Κ δείχνει ότι ο χρήστης Γ ($K = 3,8$) είναι ο πιο «κοινωνικός», ο χρήστης Β ($K = 12,5$) ακολουθεί και ο χρήστης Α ($K = 20$) είναι ο λιγότερο «κοινωνικός» χρήστης.

Πίνακας 13. Κατανομή χρηστών σε ομάδες με βάση την τιμή της μετρικής της «κοινωνικότητας»

Ομάδα 1		Ομάδα 2		Ομάδα 3	
ID	K-Μετρική	ID	K-Μετρική	ID	K-Μετρική
13	0,072	17	0,027	2	0,011
35	0,078	18	0,042	4	0,007
37	0,150	22	0,029	5	0,009
		26	0,036	6	0,012
		33	0,033	9	0,021
		34	0,033	14	0,007
		39	0,037	19	0,014
		42	0,024	24	0,011
		44	0,024	36	0,018
				38	0,021
				40	0,009
				41	0,007
				43	0,016

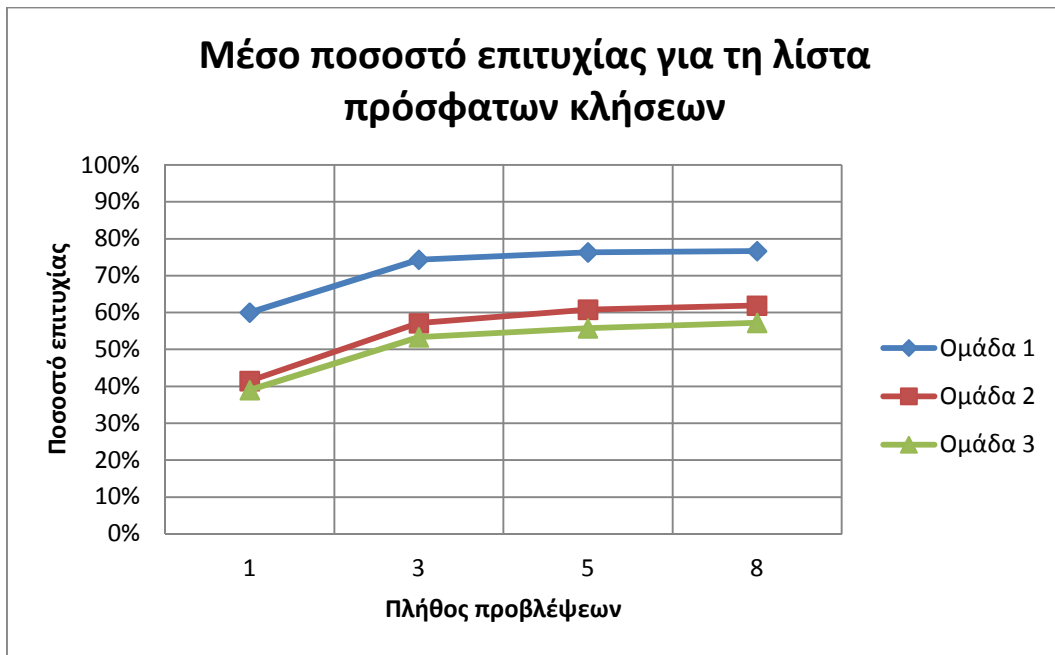
Ο υπολογισμός της μετρικής για τους χρήστες του συνόλου δεδομένων, επέτρεψε την εκτέλεση του αλγορίθμου συσταδοποίησης k-means, για τιμές του $k=1,2,3$ και 4 (καθώς ο εμπειρικός κανόνας επιλογής του k έδειχνε ότι η τιμή του θα έπρεπε να είναι μεταξύ 3 και 4). Τελικά, τα βέλτιστα αποτελέσματα βρέθηκαν για $k=3$, καθώς γι' αυτή

την τιμή μειώθηκε το άθροισμα τετραγώνων εντός των συστάδων (within clusters sum of squares). Συνεπώς, όπως και στην έρευνα των Lee, Seo και Lee (2010), οι χρήστες μπορούν να διαχωριστούν σε 3 ομάδες, όπως φαίνεται στον πίνακα 13. Η ομάδα 1 είναι η λιγότερο «κοινωνική», δηλαδή άνθρωποι που τείνουν να κάνουν τις περισσότερες κλήσεις προς ένα πολύ μικρό αριθμό επαφών και θα μπορούσε να υποθέσει κανείς ότι είναι πιο πιθανό να παρουσιάσουν τακτική και επομένως προβλέψιμη συμπεριφορά. Η ομάδα 2 περιλαμβάνει τους μέτρια «κοινωνικούς» χρήστες, ενώ η ομάδα 3 περιέχει τους περισσότερο «κοινωνικούς» χρήστες, που βρίσκονται σε διαρκή επικοινωνία με ένα μεγαλύτερο πλήθος επαφών και συνεπώς η συμπεριφορά τους είναι πιθανότατα πιο δύσκολο να προβλεφθεί. Η ομάδα 1 περιλαμβάνει μόνο 3 χρήστες, οπότε τα δεδομένα που την αφορούν παρουσιάζονται με επιφύλαξη όσον αφορά τη σημασία τους.

4.2.2 Πειραματικά αποτελέσματα

4.2.2.1 Πειράματα βάσης σύγκρισης

Εκτός από την πραγματοποίηση αναζήτησης στις αλφαβητικές λίστες επαφών, στις συσκευές κινητού υπολογισμού υπάρχουν διαθέσιμες συνήθως δύο ακόμα μέθοδοι για την ανάκτηση των επαφών όταν ο χρήστης θέλει να ξεκινήσει μια νέα κλήση: η λίστα με τις πιο συχνά καλούμενες επαφές και η λίστα με τις πιο πρόσφατες κλήσεις. Αυτές οι δύο μέθοδοι αντιστοιχούν σε εκτελέσεις του αλγορίθμου πρόβλεψης με ζεύγη βαρών ($w_f=1, w_r=0$) και ($w_f=0, w_r=1$) αντίστοιχα. Αυτές οι εκτελέσεις θεωρούνται πειράματα βάσης σύγκρισης, τα αποτελέσματα των οποίων, συγκρινόμενα με τις εκτελέσεις όπου και οι δύο διαστάσεις συνδυάζονται, αποκαλύπτουν τη βελτίωση της προτεινόμενης προσέγγισης σε σχέση με τις παραδοσιακές μεθόδους ανάκτησης. Στα γραφήματα 19 και 20 παρουσιάζεται η μέση απόδοση των δύο μεθόδων για κάθε ομάδα χρηστών και για διαφορετικά μήκη της λίστας πρόβλεψης. Ένα ενδιαφέρον συμπέρασμα που προκύπτει είναι πως οι πιο «κοινωνικοί» χρήστες (ομάδες 2 και 3) των κινητών συσκευών έχουν τη δυνατότητα να βρουν την επαφή που θέλουν να καλέσουν στην κορυφή της λίστας κλήσεων μόνο στο 40% των περιπτώσεων και στην κορυφή της λίστας συχνών επαφών μόνο στο 20-35% των περιπτώσεων. Επιπρόσθετα, γι' αυτές τις ομάδες, η απόδοση αυξάνεται γύρω στο 60% καθώς κοιτούν περισσότερο το ιστορικό κλήσεων, χωρίς ιδιαίτερη βελτίωση μετά τις 3 πρώτες εγγραφές, ενώ φαίνεται να υπάρχει μια σχεδόν γραμμική αύξηση στην πιθανότητα να βρουν την επαφή που θέλουν να καλέσουν καθώς το μέγεθος της λίστας πιο συχνών επαφών αυξάνεται. Η ομάδα 1 (λιγότερο «κοινωνικοί» χρήστες) παρουσιάζει παρόμοια συμπεριφορά, αλλά με πολύ καλύτερη απόδοση ως προς την επιτυχία ανάκτησης, όπως θα περίμενε κανείς.



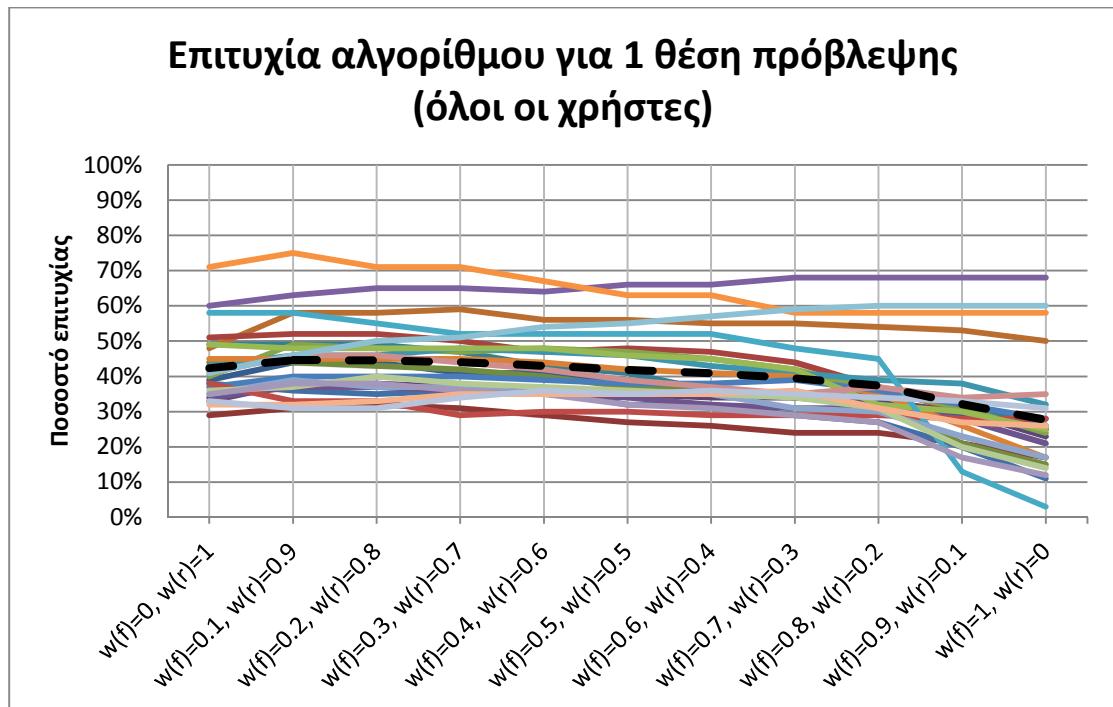
Γράφημα 19. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της προσφατότητας ($w_f=0$, $w_r=1$, ισοδυναμεί με τη λίστα πρόσφατων κλήσεων)



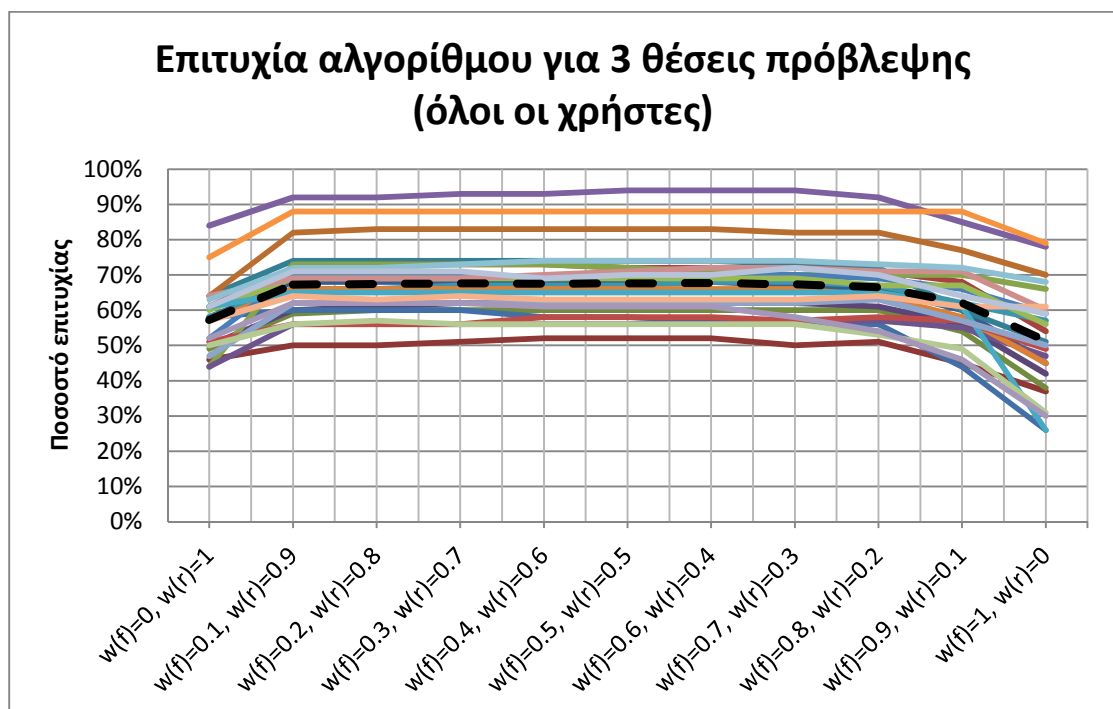
Γράφημα 20. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της συχνότητας ($w_f=1$, $w_r=0$, ισοδυναμεί με τη λίστα συχνών επαφών)

4.2.2.2 Κυρίως πειράματα

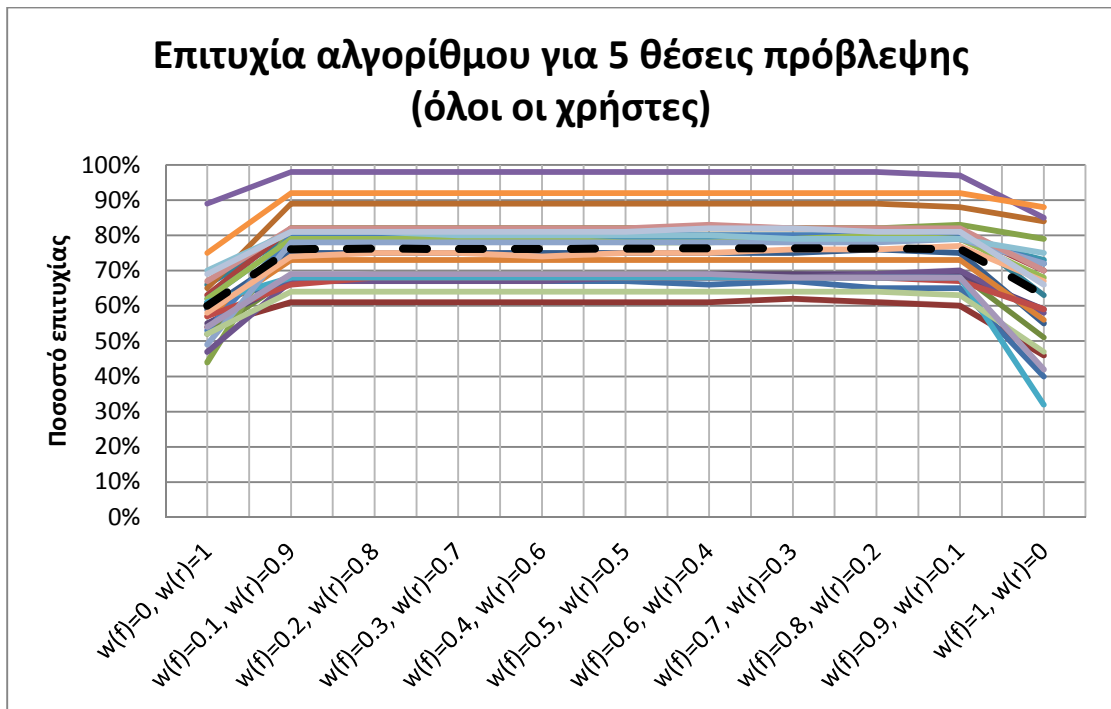
Αρχικά πραγματοποιήθηκε ένα σύνολο από πειραματικές προσομοιώσεις όπου μελετήθηκε μόνο η περίπτωση να βρίσκεται η καλούμενη επαφή στη λίστα των προβλέψεων ή όχι, χωρίς να λαμβάνεται υπόψη η θέση της σε αυτή.



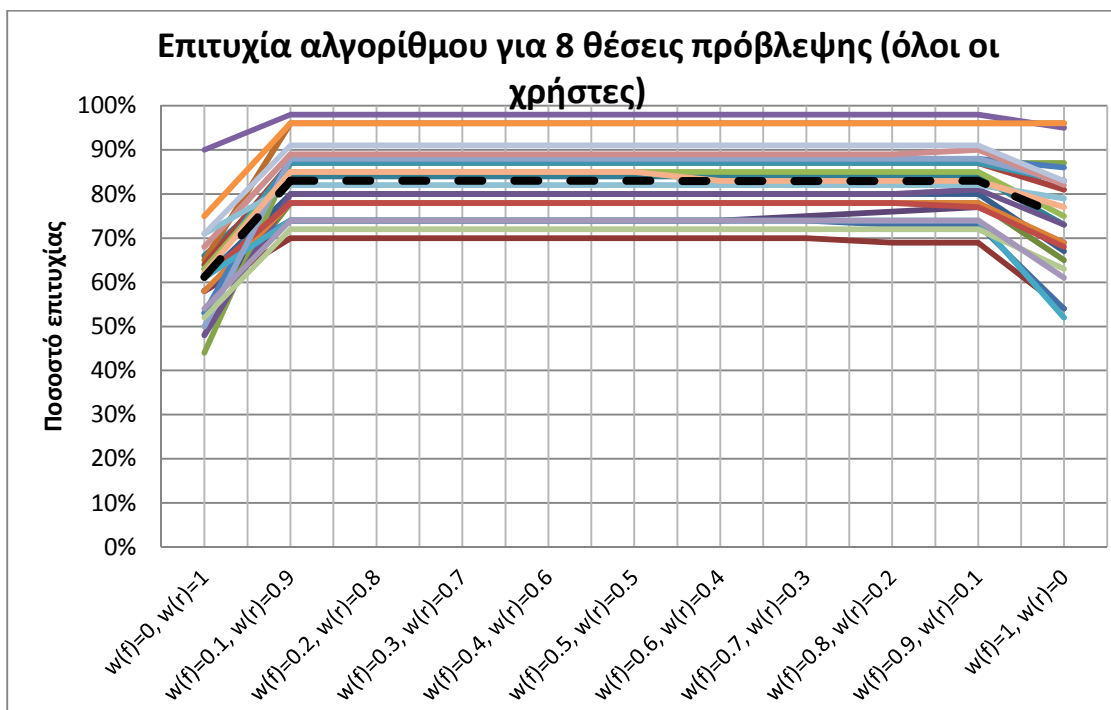
Γράφημα 21. Επιτυχία αλγορίθμου για λίστα πρόβλεψης μιας θέσης. Ο μέσος όρος παρουσιάζεται με διακεκομμένη μαύρη γραμμή.



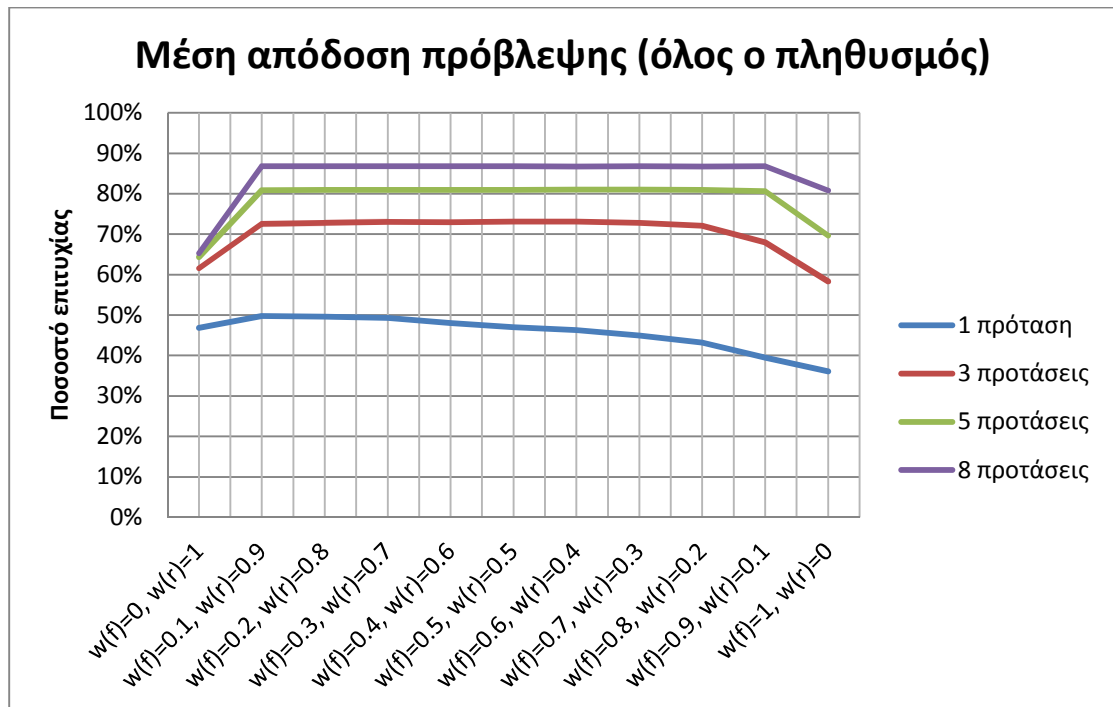
Γράφημα 22. Επιτυχία αλγορίθμου για λίστα πρόβλεψης 3 θέσεων



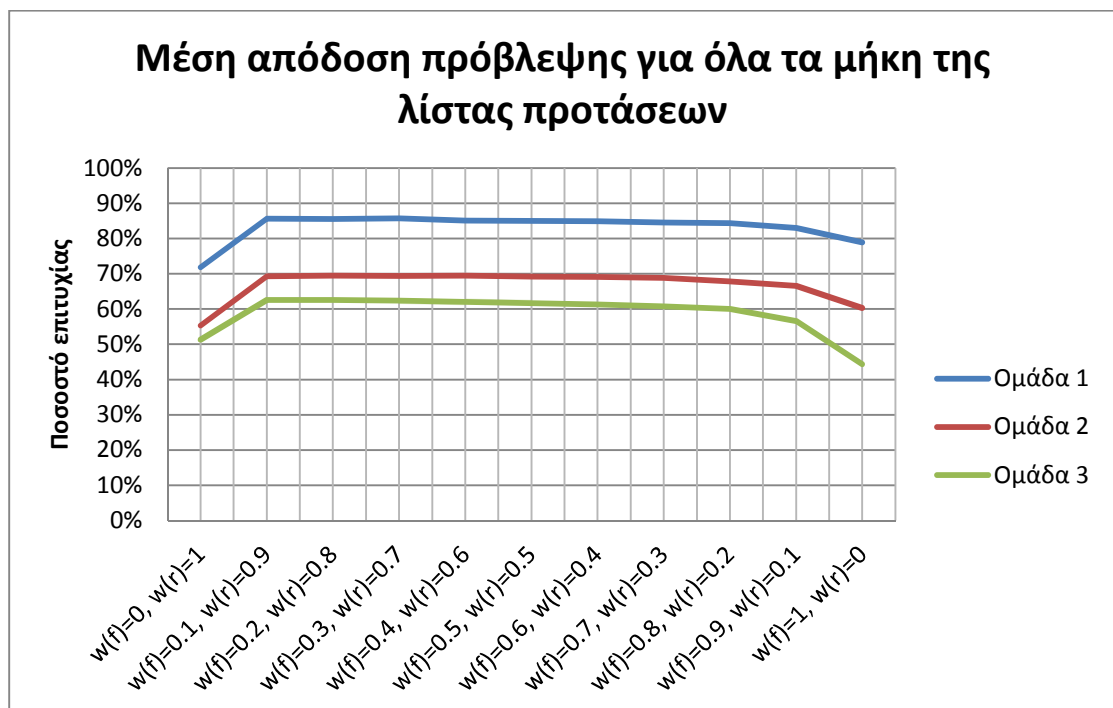
Γράφημα 23. Επιτυχία αλγορίθμου για λίστα πρόβλεψης 5 θέσεων



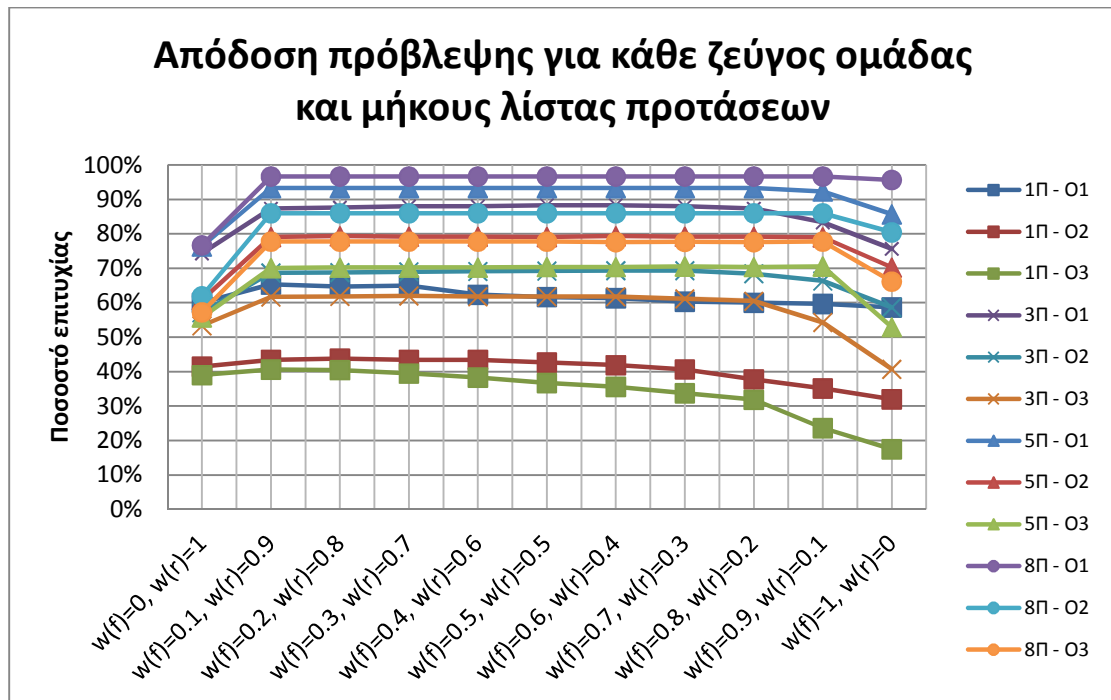
Γράφημα 24. Επιτυχία αλγορίθμου για λίστα πρόβλεψης 8 θέσεων



Γράφημα 25. Μέση απόδοση πρόβλεψης για όλο τον πληθυσμό ανά μήκος λίστας προτάσεων



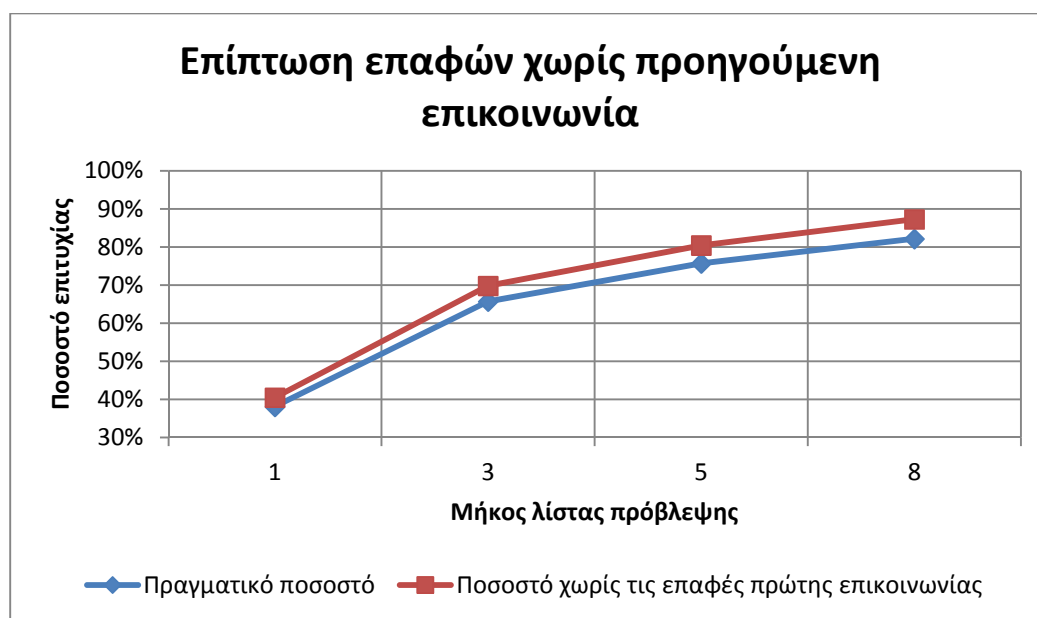
Γράφημα 26. Μέση απόδοση πρόβλεψης για όλα τα μήκη της λίστας προτάσεων ανά ομάδα χρηστών



Γράφημα 27. Απόδοση πρόβλεψης για κάθε ζεύγος ομάδας και μήκους λίστας προτάσεων

Τα παραπάνω γραφήματα παρουσιάζουν τα ποσοστά επιτυχίας για όλους τους χρήστες. Στην αρχή (γραφήματα 21 έως 24) παρουσιάζονται αναλυτικά τα αποτελέσματα για την περίπτωση των 1, 3, 5 και 8 θέσεων πρόβλεψης για όλους τους χρήστες και στη συνέχεια παρουσιάζεται το μέσο ποσοστό επιτυχίας για όλους τους χρήστες ανά μήκος της λίστας πρόβλεψης (γράφημα 25). Τέλος, το γράφημα 26 δείχνει την απόδοση του αλγορίθμου για όλα τα μήκη της λίστας πρόβλεψης ανά ομάδα «κοινωνικότητας», ενώ το γράφημα 27 δείχνει την απόδοση του αλγορίθμου για όλα τα ζεύγη ομάδων χρηστών και μεγέθους λίστας προτάσεων. Μια πρώτη παρατήρηση που είναι αμέσως εμφανής είναι πως ο συνδυασμός των διαστάσεων της συχνότητας και προσφατότητας της επικοινωνίας προσφέρει καλύτερα αποτελέσματα σε σχέση με τη χρήση κάθε διάστασης ξεχωριστά. Αυτό το εύρημα υποδεικνύει ότι οι διεπαφές που υπάρχουν διαθέσιμες στις συσκευές κινητού υπολογισμού για την αξιοποίηση των επαφών και βασίζονται μόνο στη συχνότητα και την προσφατότητα της επικοινωνίας δεν είναι σε καμία περίπτωση βέλτιστες. Μια διεπαφή που θα παρείχε προτάσεις επαφών για κλήση, βασισμένη σε συνδυασμό των δύο διαστάσεων, θα ήταν πολύ πιο αποτελεσματική. Είναι ξεκάθαρο από τα γραφήματα ότι η ομάδα 1 έχει σταθερά τα μεγαλύτερα ποσοστά επιτυχίας, ενώ οι ομάδες 2 και 3 ακολουθούν. Αυτό επιβεβαιώνει την υπόθεση ότι η ομάδα 1 παρουσιάζει την πιο προβλέψιμη συμπεριφορά (συχνές κλήσεις προς ένα μικρό αριθμό επαφών). Επιπρόσθετα, παρατηρεί κανείς ότι καθώς το μήκος της λίστας των προτεινόμενων επαφών αυξάνεται, ο ρόλος των βαρών γίνεται λιγότερο σημαντικός.

Για μικρές λίστες προβλέψεων (1-3 θέσεις) το βάρος της διάστασης της προσφατότητας της επικοινωνίας φαίνεται να παίζει πιο σημαντικό ρόλο στην επιτυχή πρόβλεψη, επομένως η προσφατότητα της επικοινωνίας φαίνεται να είναι πιο σημαντικός παράγοντας από τη συχνότητα για τον καθορισμό της σημασίας μιας επαφής.

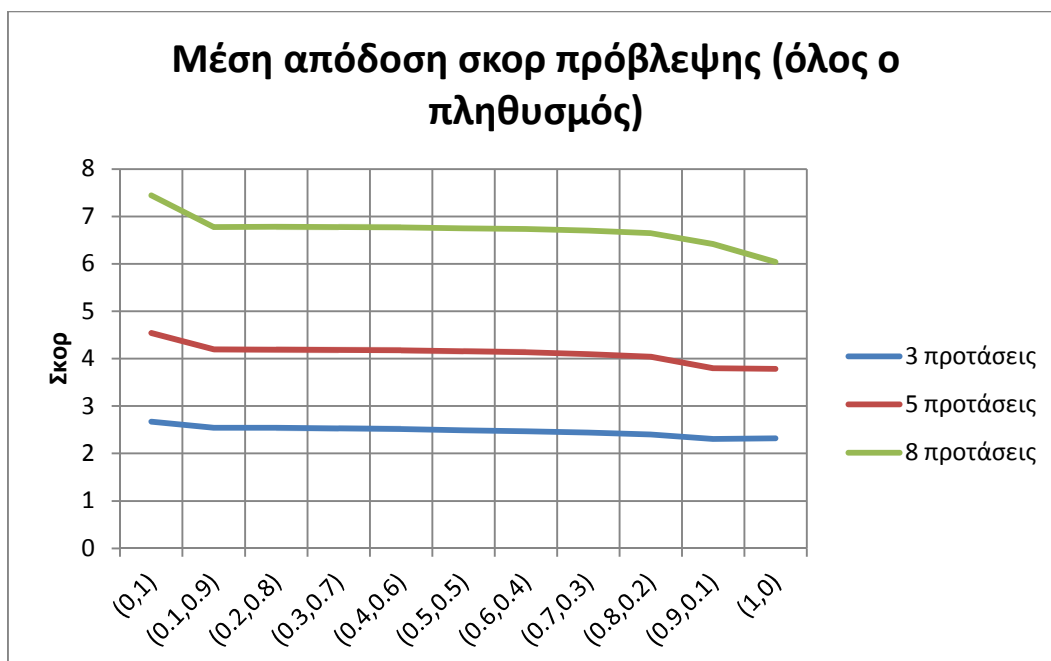


Γράφημα 28. Επίπτωση των επαφών με τις οποίες δεν υπάρχει προηγούμενη επικοινωνία στην επιτυχία των προβλέψεων

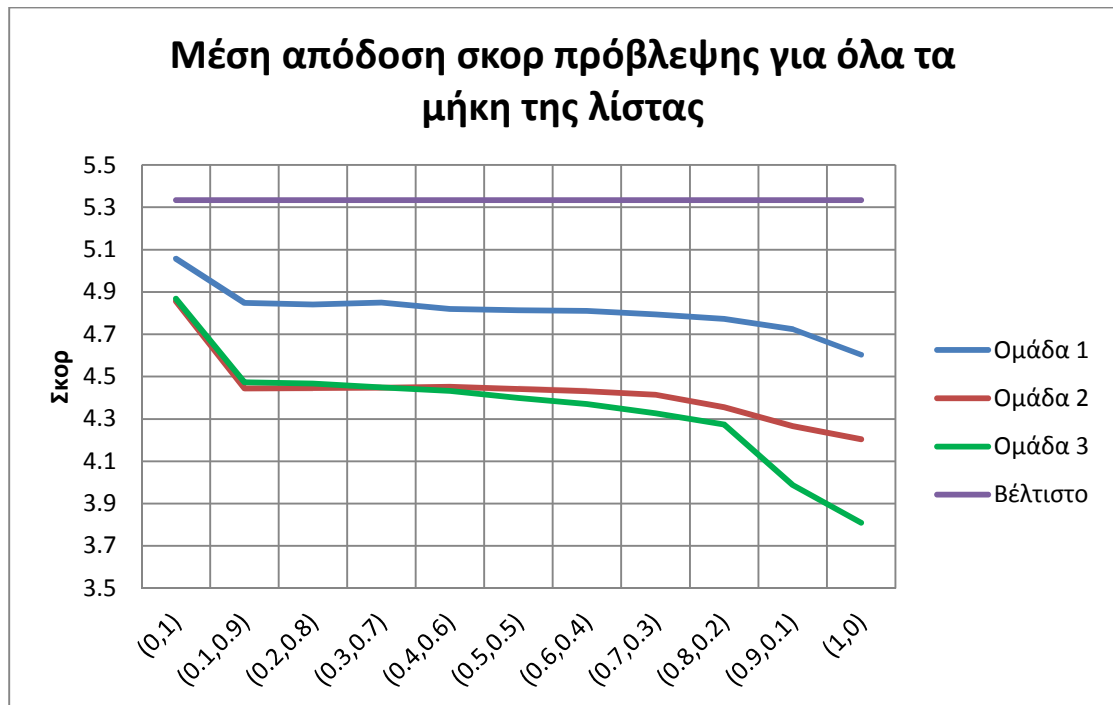
Στο γράφημα 28, παρουσιάζεται η επίπτωση που έχουν στην επιτυχία πρόβλεψης, για το σύνολο των χρηστών, επαφές οι οποίες καλούνται για πρώτη φορά. Αυτές οι κλήσεις δεν μπορούν να προβλεφθούν μόνο από την ανάλυση του ιστορικού κλήσεων και ενδεχομένως για το σκοπό αυτό να χρειάζονται και άλλες πληροφορίες πλαισίου, οι οποίες στο υπό εξέταση σύνολο δεδομένων δεν ήταν διαθέσιμες. Παρατηρεί κανείς ότι όσο μεγαλώνει η λίστα προτάσεων τόσο μεγαλύτερη είναι και η διαφορά ανάμεσα στο πλασματικό ποσοστό επιτυχίας του αλγορίθμου (όταν δηλαδή δεν πραγματοποιούνται προβλέψεις για επαφές για τις οποίες δεν υπάρχουν προηγούμενες επικοινωνίες) και στο πραγματικό ποσοστό επιτυχίας. Για μεγαλύτερα μήκη της λίστας πρόβλεψης δεν αναμένεται σημαντική περαιτέρω αύξηση. Αντίστοιχο συμπέρασμα προέκυψε και στην εργασία των Phithakkitnukoon και Dantu (2011), με τη διαφορά ότι σε εκείνη την περίπτωση η βελτίωση από την αφαίρεση των επαφών χωρίς προηγούμενα ιστορικά δεδομένα ήταν κοντά στο 10% για τις 8 θέσεις πρόβλεψης, ενώ σε αυτή που παρουσιάζεται εδώ είναι περίπου 5%.

Για να εξερευνηθεί η ποιότητα των προβλέψεων (δηλαδή πόσο κοντά ήταν η επαφή που κλήθηκε τελικά στην κορυφή της λίστας των προτάσεων, επομένως πιο πιθανό

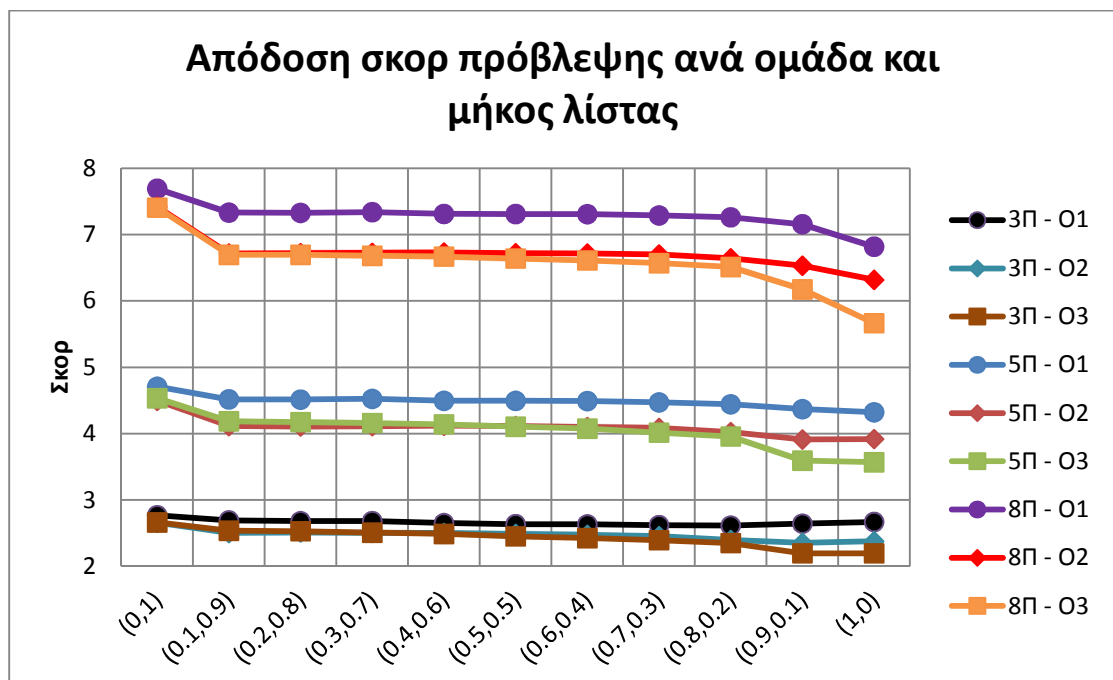
να εντοπιστεί γρηγορότερα από τον χρήστη), πραγματοποιήθηκε ένας δεύτερος γύρος πειραματικών προσομοιώσεων. Σε αυτές τις περιπτώσεις, καταγραφόταν η θέση στη λίστα όπου η επαφή βρισκόταν, όταν υπήρχε επιτυχία πρόβλεψης. Έτσι, ήταν δυνατό να αποδοθεί ένα σκορ που μπορεί να χρησιμοποιηθεί για την αξιολόγηση της ποιότητας της πρόβλεψης, το οποίο κυμαινόταν μεταξύ 1 και n, με μοναδιαίο βήμα αύξησης, όπου το n αναπαριστά το μήκος των θέσεων που περιελάμβανε η λίστα των προβλέψεων. Επειδή το σκορ αυτό επιλέχθηκε να αντιστοιχεί στην απόσταση από το κάτω μέρος της λίστας και όχι από την αρχή της, όσο μεγαλύτερη τιμή έχει τόσο καλύτερη είναι η πρόβλεψη. Με αυτόν τον τρόπο, πραγματοποιήθηκαν ξανά όλες οι προσομοιώσεις, εκτός από την περίπτωση της 1 θέσης πρόβλεψης, όπου σε κάθε πρόβλεψη υπάρχει είτε επιτυχία είτε αποτυχία και το σκορ δεν προσφέρει κάποια επιπλέον πληροφορία. Τα αποτελέσματα εμφανίζονται στα γραφήματα 29, 30 και 31. Μπορεί να παρατηρήσει κανείς ότι γενικά ο αλγόριθμος προσφέρει καλή τοποθέτηση των σωστών προβλέψεων στη λίστα με τις προτάσεις, η οποία είναι κατά μέσο όρο κοντά στην κορυφή σε κάθε περίπτωση. Και πάλι, μπορεί να παρατηρήσει κανείς ότι η διάσταση της προσφατότητας παρέχει καλύτερη απόδοση όταν λαμβάνει μεγαλύτερο βάρος σε σχέση με τη συχνότητα. Επιπλέον, παρατηρείται ότι οι χρήστες της ομάδας 1 έχουν τα καλύτερα αποτελέσματα, ενώ ακολουθούν αυτοί της ομάδας 2 και τέλος της ομάδας 3, επιβεβαιώνοντας ξανά την υπόθεση που αναπτύχθηκε νωρίτερα σχετικά με το ρόλο της «κοινωνικότητας» στην πρόβλεψη της επικοινωνιακής συμπεριφοράς.



Γράφημα 29. Μέση απόδοση σκορ πρόβλεψης για όλο τον πληθυσμό των χρηστών



Γράφημα 30. Μέση απόδοση σκορ πρόβλεψης για όλα τα μήκη λιστών. Με μωβ το θεωρητικά βέλτιστο σκορ (5,33)



Γράφημα 31. Απόδοση σκορ πρόβλεψης για όλα τα ζεύγη ομάδων και μήκους λίστας

Είναι εμφανές από τα προηγούμενα γραφήματα ότι στις περισσότερες περιπτώσεις υπάρχουν μετρήσιμες βελτιώσεις στην απόδοση από το συνδυασμό των διαστάσεων της συχνότητας και προσφατότητας της επικοινωνίας. Δεν παρατηρείται μετρήσιμη διαφορά γι' αυτούς τους συνδυασμούς, εκτός από τις περιπτώσεις των ομάδων 1 και 3, όπου η απόδοση φαίνεται να φτάνει στη μέγιστη τιμή της όταν δίνεται μεγαλύτερο

βάρος στην προσφατότητα. Η απόδοση του αλγορίθμου όταν λαμβάνει υπόψη μόνο τη συχνότητα ($w_f=1$) ή μόνο την προσφατότητα ($w_r=1$) συγκρίθηκε με την κατάσταση όπου δεν υπάρχει προηγούμενη γνώση για την «κοινωνικότητα» του χρήστη και επομένως τα βάρη μπορούν να τεθούν ίσα, χωρίς προκατάληψη υπέρ μιας διάστασης ($w_f=w_r=0,5$). Η ανάλυση ANOVA (πίνακας 14) δείχνει ότι για τις περισσότερες περιπτώσεις που προκύπτουν από τις 3 υπάρχουσες συνθήκες ($w_r=1$, $w_r=0,5$ and $w_r=0$), η μηδενική υπόθεση ($H_0: \mu_{w_r=0}=\mu_{w_r=0,5}=\mu_{w_r=1}$) απορρίπτεται. Οι μόνες περιπτώσεις για τις οποίες η μηδενική υπόθεση δεν απορρίπτεται ξεκάθαρα είναι για την ομάδα 1, η οποία όμως περιλαμβάνει μόνο 3 χρήστες και επομένως τα αναφερόμενα δεδομένα δε μπορεί να θεωρηθεί ότι είναι στατιστικά σημαντικά.

Πίνακας 14. Συγκεντρωτικά αποτελέσματα ανάλυσης ANOVA

Ομάδα	Μήκος λίστας	Απόδοση επιτυχίας		Απόδοση σκορ	
		F	p	F	p
1	1	0,08892	>0,05	-	-
1	3	3,46667	>0,05	1,78063	>0,05
1	5	3,90619	>0,05	2,55229	>0,05
1	8	7,14375	<0,05	3,61975	>0,05
2	1	4,27903	<0,05	-	-
2	3	9,33996	<0,01	7,13402	<0,01
2	5	18,71707	<0,01	8,04515	<0,01
2	8	34,19153	<0,01	16,73811	<0,01
3	1	37,18531	<0,01	-	-
3	3	23,82456	<0,01	32,42174	<0,01
3	5	18,28601	<0,01	25,97022	<0,01
3	8	28,22977	<0,01	18,40532	<0,01

4.3 Συζήτηση αποτελεσμάτων

Στην προηγούμενη ενότητα παρουσιάστηκαν αναλυτικά τα αποτελέσματα των πειραματικών προσομοιώσεων. Αρχικά, ένα σημαντικό εύρημα είναι ότι ο συνδυασμός των διαστάσεων της συχνότητας και της προσφατότητας της επικοινωνίας παρέχει καλύτερα αποτελέσματα σε σχέση με την εκτέλεση του αλγορίθμου που χρησιμοποιεί κάθε διάσταση ξεχωριστά, το οποίο είναι σύμφωνο με τα συμπεράσματα των Barzaiq και Loke (2011). Αυτό το συμπέρασμα είναι υποσχόμενο, καθώς ενισχύει την πεποίθηση ότι η προσθήκη ακόμα περισσότερων διαστάσεων στο διάλυμα αναπαράστασης του πλαισίου μπορεί να οδηγήσει σε ακόμα καλύτερα ποσοστά επιτυχών προβλέψεων, επιτρέποντας το σχεδιασμό πιο αποτελεσματικών διεπαφών ανάκτησης επαφών σε σχέση με τις υπάρχουσες (πρόσφατες κλήσεις, πιο συχνά καλούμενες επαφές). Επιπλέον, όπως αναμενόταν και σύμφωνα και με τα ευρήματα των Phithakkitnukoon, Dantu, Claxton και Eagle

(2011), όσο μεγαλύτερη είναι η λίστα με τις προτεινόμενες επαφές τόσο υψηλότερο είναι το ποσοστό επιτυχίας. Εντούτοις, η ένταση αυτής της θετικής επίπτωσης μειώνεται όσο η λίστα με τις προτάσεις μεγαλώνει. Έχοντας υπόψη ότι οι μικρές οθόνες των συσκευών κινητού υπολογισμού χωράνε έως 8 επαφές χωρίς να χρειάζεται κύλιση, δε φαίνεται να έχει νόημα ο σχεδιασμός λιστών μεγαλύτερου μήκους. Ακόμα μια ενδιαφέρουσα παρατήρηση είναι ότι καθώς το μέγεθος της λίστας προτάσεων αυξάνεται, ο ρόλος των βαρών των διαστάσεων γίνεται λιγότερο σημαντικός.

Επιπλέον, η παρατήρηση των Lee, Seo και Lee (2010), πως η ύπαρξη ομάδων χρηστών με διαφορετική επικοινωνιακή συμπεριφορά επηρεάζει την απόδοση των προβλέψεων, επιβεβαιώθηκε από τα αποτελέσματα των προσομοιώσεων. Η διακύμανση των αποτελεσμάτων εξαιτίας των διαφορετικών προτύπων επικοινωνίας για κάθε ομάδα είναι μια ισχυρή ένδειξη ότι τα βάρη των διαστάσεων δεν πρέπει να είναι στατικά, αλλά δυναμικά για κάθε χρήστη. Επιπρόσθετα, θα μπορούσε κανείς να υποστηρίξει ότι τα βάρη θα έπρεπε να είναι δυναμικά ακόμα και για τον ίδιο χρήστη, κάτω από διαφορετικά πλαίσια. Για παράδειγμα, η τοποθεσία θα μπορούσε να παίζει έναν πιο σημαντικό ρόλο απ' ότι η ώρα της ημέρας για την πρόβλεψη της επόμενης επαφής που θα κληθεί, όταν ο χρήστης ταξιδεύει σε μια άλλη πόλη.

Πίνακας 15. Σύγκριση αποτελεσμάτων πρόβλεψης με άλλες εργασίες

	Plessas, Stefanis, Komninos & Garofalakis (2013)	Phithakkitnukoon, Dantu, Claxton & Eagle (2011)	Lee, Seo & Lee (2010)	Srinivasan, Moghaddam, Mukherji, Rachuri, Xu & Tapia (2014)	Phithakkitnukoon & Dantu (2011)
1	44,6%	41,0%	39,3%	31,0%	31,0%
3	67,7%	57,0%		44,0%	52,0%
5	76,3%	70,0%	62,5%	53,0%	61,0%
8	83,0%	86,0%			71,0%

Τα αποτελέσματα του προτεινόμενου αλγορίθμου για την πρόβλεψη των εξερχόμενων κλήσεων μπορούν να συγκριθούν με αυτά που προκύπτουν από προσομοιώσεις άλλων αλγορίθμων που παρουσιάζονται σε σχετικές ερευνητικές εργασίες. Συνολικά, η σύγκριση παρουσιάζεται στον πίνακα 15. Δεν περιλαμβάνονται τα ευρήματα των Min, Kim και Cho (2008), καθώς αφορούσαν δεδομένα 12 ημερών από ένα μόνο χρήστη και επομένως οι τιμές που αναφέρονται δεν μπορούν να θεωρηθούν αξιόπιστες. Επιπλέον, δεν περιλαμβάνονται για τον ίδιο λόγο τα αποτελέσματα των Barzaiq και Loke (2011), καθώς αφορούσαν και πάλι έναν χρήστη και προβλέψεις μόνο για τις κλήσεις που πραγματοποιήθηκαν την πρώτη μέρα κάθε

μήνα που κάλυπτε το αρχείο κλήσεων. Είναι εμφανές ότι η προσέγγιση που προτάθηκε, παρότι χαρακτηρίζεται από απλότητα, παρέχει καλύτερα αποτελέσματα από την πλειοψηφία των προσεγγίσεων (οι οποίες χαρακτηρίζονται από αυξημένη πολυπλοκότητα) που παρουσιάζονται στην ερευνητική βιβλιογραφία. Είναι χαρακτηριστικό, ότι τα ποσοστά επιτυχίας του αλγορίθμου είναι τα υψηλότερα για την περίπτωση των 1, 3 και 5 θέσεων πρόβλεψης. Επιπλέον, για την περίπτωση της λίστας με 8 θέσεις πρόβλεψης, η απόδοση του αλγορίθμου είναι πολύ κοντά στη βέλτιστη προσέγγιση, αφού η διαφορά από τον αλγόριθμο των Phithakitnukoon, Dantu, Claxton και Eagle (2011) είναι 3%. Ακόμα, ο αλγόριθμος των Lee, Seo και Lee (2010) σημείωσε ποσοστά επιτυχίας (για λίστα 5 θέσεων) μεγαλύτερα από 75% μόνο για έναν τύπο χρηστών (τους πιο εύκολα προβλέψιμους), ενώ για τις άλλες δύο ομάδες που προέκυψαν τα ποσοστά επιτυχίας κυμάνθηκαν γύρω στο 50% και 30% κατά μέσο όρο, αντίστοιχα. Στην προτεινόμενη προσέγγιση, τα αντίστοιχα ποσοστά επιτυχούς πρόβλεψης του αλγορίθμου για τη λίστα 5 θέσεων ήταν για την κάθε ομάδα χρηστών, με φθίνουσα δυσκολία πρόβλεψης, περίπου 70%, 80% και 93% (γράφημα 27). Τέλος, ο δεύτερος γύρος πειραματικών προσομοιώσεων έδειξε ότι ο αλγόριθμος προσφέρει ικανοποιητική κατάταξη των προβλέψεων κοντά στην κορυφή της λίστας.

4.4 Μεγάλης κλίμακας πείραμα αξιολόγησης αλγορίθμου

Μετά την πραγματοποίηση του προκαταρκτικού πειράματος αξιολόγησης του αλγορίθμου πρόβλεψης των επαφών με τις οποίες είναι πιο πιθανό να πραγματοποιηθεί μια εξερχόμενη κλήση ανά πάσα στιγμή και καθώς έγινε δημόσια διαθέσιμο ένα πλήρες σύνολο δεδομένων από τη Nokia, το οποίο περιελάμβανε πολύ μεγαλύτερο αριθμό χρηστών και δεδομένα που κάλυπταν μεγαλύτερο χρονικό διάστημα, αποφασίστηκε η πραγματοποίηση επιπλέον πειραματικών προσομοιώσεων, ώστε να επικυρωθούν τα συμπεράσματα που προέκυψαν (Stefanis, Plessas, Komninos & Garofalakis, 2014).

4.4.1 Σύνολο δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε προέκυψε από μια ερευνητική πρωτοβουλία της Nokia με τίτλο Mobile Data Collection Campaign (Kiukkonen, Blom, Dousse, Gatica-Perez & Laurila, 2010; Laurila, Gatica-Perez, Aad, Bornet, Do, Dousse, ... & Miettinen, 2012) και έγινε δημόσια διαθέσιμο προς την ερευνητική κοινότητα. Η συλλογή των δεδομένων πραγματοποιήθηκε στην περιοχή της Γενεύης, στην Ελβετία, από τον Οκτώβριο του 2009 έως τον Μάρτιο του 2011. Συνολικά, ο αριθμός των συμμετεχόντων έφτασε τους 185 (38% γυναίκες και 62% άνδρες, με τα

2/3 του πληθυσμού να έχουν ηλικία μεταξύ 22 και 33 έτη). Οι συμμετέχοντες εξοπλίστηκαν με έξυπνα κινητά τηλέφωνα (Nokia N95), στα οποία ειδικό λογισμικό κατέγραφε δεδομένα που σχετίζονται με την τοποθεσία, την κίνηση, την εγγύτητα, την επικοινωνιακή συμπεριφορά, τη χρήση εφαρμογών κ.λπ. (πίνακας 16). Εκείνη την περίοδο καταγράφησαν περισσότερες από 240.000 κλήσεις. Τα δεδομένα στέλνονταν 2 φορές την ημέρα στους εξυπηρετητές της Nokia, όπου και γινόταν η επεξεργασία τους. Οι ερευνητές έλαβαν ιδιαίτερη μέριμνα για θέματα που αφορούν την ασφάλεια των δεδομένων καθώς και την προστασία της ιδιωτικότητας και της ανωνυμίας των συμμετεχόντων.

Πίνακας 16. Τύποι και ποσότητες δεδομένων που συλλέχθηκαν από την εκστρατεία της Nokia

Τύπος δεδομένων	Ποσότητα
Κλήσεις	240.227
SMS	175.832
Φωτογραφίες	37.151
Βίντεο	2.940
Γεγονότα εφαρμογών	8.096.870
Εγγραφές ημερολογίου	13.792
Επαφές	45.928
Σημεία τοποθεσίας	26.152.673
Μοναδικές κεραίες κινητής τηλεφωνίας	99.166
Εγγραφές επιταχυνσιόμετρου	1.273.333
Παρατηρήσεις Bluetooth	38.259.550
Μοναδικές Bluetooth συσκευές	498.593
Παρατηρήσεις WLAN	31.013.270
Μοναδικά σημεία πρόσβασης WLAN	560.441
Δείγματα ήχου	595.895

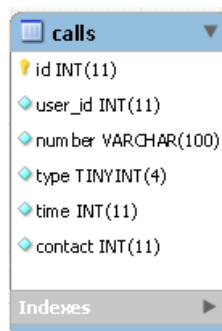
Η ανάλυση που ακολουθεί επικεντρώθηκε στα δεδομένα επικοινωνίας των χρηστών και συγκεκριμένα στο αρχείο κλήσεών τους. Εντούτοις, όπως και στο προηγούμενο πείραμα, κάποιοι από τους συμμετέχοντες στην εκστρατεία συλλογής δεδομένων χρειάστηκε να αποκλειστούν καθώς δε θεωρήθηκαν έγκυροι για ανάλυση του ιστορικού κλήσεων. Κάποιοι από αυτούς είχαν αρχεία κλήσεων που κάλυπταν εξαιρετικά μικρά χρονικά διαστήματα, ενώ άλλοι είχαν πολύ λίγες εγγραφές κλήσεων ή πολύ αραιή χρήση της συσκευής τους. Αυτό ήταν αναμενόμενο, αφού, όπως περιγράφεται στην τεκμηρίωση που παρείχε η Nokia, υπήρχαν χρήστες που εγκατέλειψαν νωρίς το πείραμα και άλλοι, οι οποίοι συχνά απενεργοποιούσαν το λογισμικό καταγραφής. Χρήστες με τα παρακάτω ακραία χαρακτηριστικά αποκλείστηκαν:

- χρονικό διάστημα που κάλυπτε το αρχείο κλήσεων μικρότερο από 30 ημέρες

- κλήσεις λιγότερες από 100
- κλήσεις ανά ημέρα λιγότερες από 2,25

Τελικά, προέκυψε ένα σύνολο που περιελάμβανε 106 χρήστες. Δεν μπορεί να γίνει αναφορά στο μέσο μέγεθος των αποθετηρίων επαφών των χρηστών, αφού το σύνολο δεδομένων της Nokia εξελισσόταν με το χρόνο (και επομένως τα περιεχόμενα των αποθετηρίων μεταβάλλονταν), σε αντίθεση με το σύνολο δεδομένων από το προκαταρκτικό πείραμα που μπορεί να θεωρηθεί ως ένα στιγμιότυπο στο χρόνο. Όπως και στο προκαταρκτικό πείραμα, τα αρχεία κλήσεων κάλυπταν διαφορετικές χρονικές περιόδους σε ημέρες (mean=374,98, sd=136,78, min=33,96, max=608,25). Κατά μέσο όρο κάθε χρήστης πραγματοποίησε 1.928,75 κλήσεις (sd=1.036,11, min=214, max=5.101).

Στην εικόνα 25 παρουσιάζεται το σχήμα της ΒΔ για την αποθήκευση των αρχείων κλήσεων των χρηστών. Στην ουσία, χρειάζεται ένας μόνο πίνακας, ο οποίος κρατάει τις απαραίτητες πληροφορίες για τις κλήσεις των χρηστών: το id του χρήστη που την πραγματοποίησε, το νούμερο (κρυπτογραφημένο) με το οποίο έγινε η επικοινωνία, το είδος της κλήσης (εισερχόμενη, εξερχόμενη, μη απαντημένη), τη χρονική στιγμή που έγινε και το id της επαφής με την οποία έλαβε χώρα.



Εικόνα 25. Σχήμα ΒΔ για την αποθήκευση των αρχείων κλήσεων των χρηστών

Όπως και στο προηγούμενο πείραμα, έτσι και σε αυτό υπολογίστηκε η τιμή της Κ-Μετρικής κάθε χρήστη και εκτελέστηκε ο αλγόριθμος k-means για τον εντοπισμό των ομάδων χρηστών ανάλογα με την «κοινωνικότητά» τους. Από την ανάλυση προέκυψε ότι η ομάδα 1 (λιγότερο «κοινωνικοί» χρήστες) περιλαμβάνει 4 χρήστες, η ομάδα 2 (μέτρια «κοινωνικοί» χρήστες) 20 χρήστες και η ομάδα 3 (περισσότερο «κοινωνικοί» χρήστες) 82 χρήστες. Στον πίνακα 17 παρουσιάζεται μια σύγκριση του συνόλου δεδομένων της Nokia και του συνόλου δεδομένων που χρησιμοποιήθηκε στο προκαταρκτικό πείραμα ως προς τα χαρακτηριστικά που σχετίζονται με την «κοινωνικότητα» των χρηστών.

Πίνακας 17. Σύγκριση των χαρακτηριστικών «κοινωνικότητας» των δύο συνόλων δεδομένων

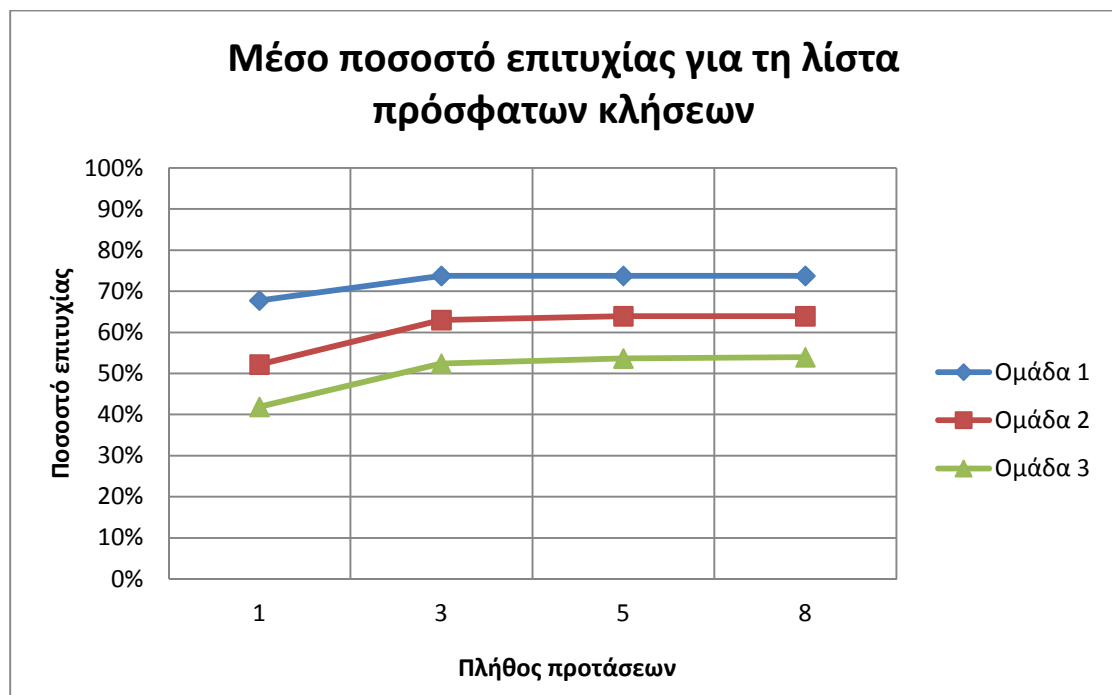
	Προκαταρκτικό Σύνολο Δεδομένων			Σύνολο Δεδομένων Nokia		
	Κλήσεις/ ημέρα	# χρήστες	Μέση τιμή Κ-Μετρικής	Κλήσεις/ Ημέρα	# χρήστες	Μέση τιμή Κ-Μετρικής
Ομάδα 1	13,73	3 (12%)	0,09983	5,74	4 (3.8%)	0,09916
Ομάδα 2	11,85	9 (36%)	0,03155	5,66	20 (18.9%)	0,01956
Ομάδα 3	10,7	13 (52%)	0,01232	5,33	82 (77.3%)	0,00579

4.4.2 Πειραματικά αποτελέσματα

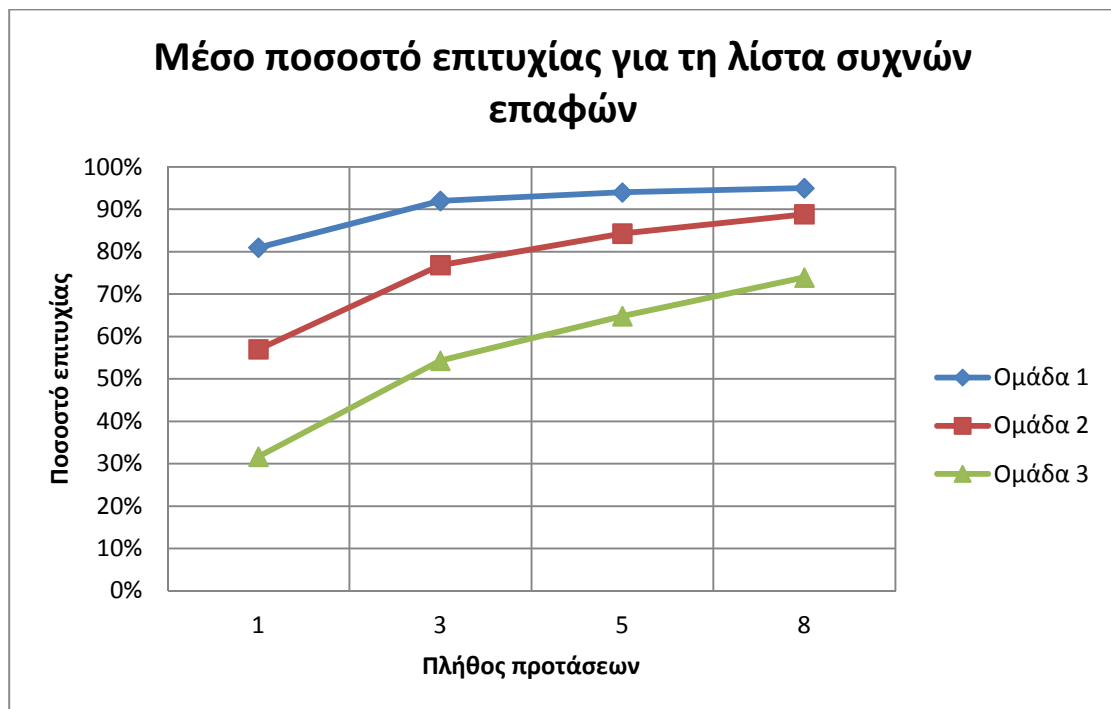
Για την πραγματοποίηση των προσομοιώσεων εφαρμόστηκε η ίδια διαδικασία και χρησιμοποιήθηκαν οι ίδιες τιμές των παραμέτρων, όπως και στο προκαταρκτικό πείραμα, δηλαδή παράθυρο εκπαίδευσης t_w 10 ημερών, παράθυρο προσφατότητας t_r 12 ωρών και λίστες πρόβλεψης μήκους 1,3, 5 και 8 θέσεων.

4.4.2.1 Πειράματα βάσης σύγκρισης

Όπως έγινε στο προκαταρκτικό πείραμα, έτσι και με τα δεδομένα της Nokia, πραγματοποιήθηκαν εκτελέσεις του αλγορίθμου που προσομοιώνουν τη λίστα πρόσφατων κλήσεων ($w_f=0$, $w_r=1$) και τη λίστα συχνών επαφών ($w_f=1$, $w_r=0$). Καθώς αυτές είναι οι πιο διαδεδομένες μέθοδοι ανάκτησης επαφών στις συσκευές κινητού υπολογισμού, τα αποτελέσματα αυτών των προσομοιώσεων επιτρέπουν τη σύγκριση με την προσέγγιση που προτείνεται, όπου οι δύο διαστάσεις συνδυάζονται.



Γράφημα 32. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της προσφατότητας ($w_f=0$, $w_r=1$, ισοδυναμεί με τη λίστα πρόσφατων κλήσεων)



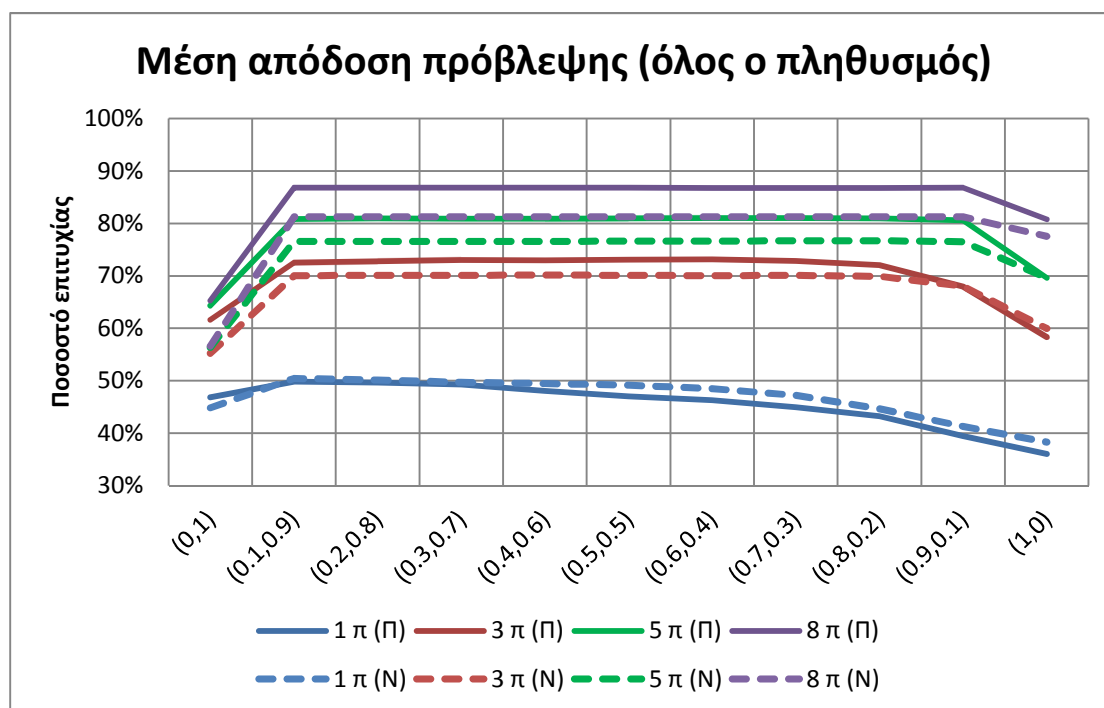
Γράφημα 33. Επιτυχία προβλέψεων χρησιμοποιώντας μόνο τη διάσταση της συχνότητας ($w_f=1$, $w_r=0$, ισοδυναμεί με τη λίστα συχνών επαφών)

Από τα αποτελέσματα των προσομοιώσεων (γραφήματα 32 και 33) παρατηρεί κανείς ότι αν και οι γραφικές παραστάσεις δεν είναι ίδιες με αυτές που προέκυψαν από το προκαταρκτικό πείραμα, η συνολική συμπεριφορά είναι παρόμοια και επιβεβαιώνει τα ευρήματα της ανάλυσης. Η λιγότερο «κοινωνική» ομάδα (ομάδα 1) εμφανίζει την καλύτερη απόδοση που προσεγγίζει το 74% και το 95% για την προσφατότητα και τη συχνότητα της επικοινωνίας αντίστοιχα, για 8 θέσεις πρόβλεψης. Παρομοίως, για την ομάδα 3 (περισσότερο «κοινωνική»), η οποία αποτελεί και τη μεγαλύτερη σε μέγεθος ομάδα, τα ποσοστά επιτυχίας είναι αρκετά χαμηλότερα και στις δύο περιπτώσεις. Τα αποτελέσματα της ανάλυσης είναι αποκαλυπτικά: αρχικά, όταν οι χρήστες χρησιμοποιούν την πιο οικεία μέθοδο ανάκτησης (ιστορικό κλήσεων), παρατηρεί κανείς ότι για την πολυπληθέστερη ομάδα χρηστών, ακόμα και αν το μέγεθος της λίστας προβλέψεων είναι μεγάλο (8 προτάσεις), υπάρχει στην καλύτερη περίπτωση μόνο 54% πιθανότητα να βρεθεί εκεί η επιθυμητή επαφή. Στην πραγματικότητα, φαίνεται να έχει μικρή σημασία εάν περισσότερες από 3 επαφές βρίσκονται στη λίστα, καθώς η επιτυχία πρόβλεψης δε βελτιώνεται ιδιαίτερα για περισσότερες από 3 προτάσεις. Μια ακόμα παρατήρηση είναι ότι η λίστα συχνών επαφών μπορεί να παράσχει καλύτερα αποτελέσματα, με την πιθανότητα εύρεσης της επιθυμητής επαφής να αυξάνεται παράλληλα με το μέγεθος της λίστας προβλέψεων. Ακόμα και για την ομάδα 3, το ποσοστό επιτυχίας πλησιάζει το 74%. Εντούτοις, είναι ενδιαφέρον ότι ενώ το ιστορικό κλήσεων είναι γρήγορα προσβάσιμο στις

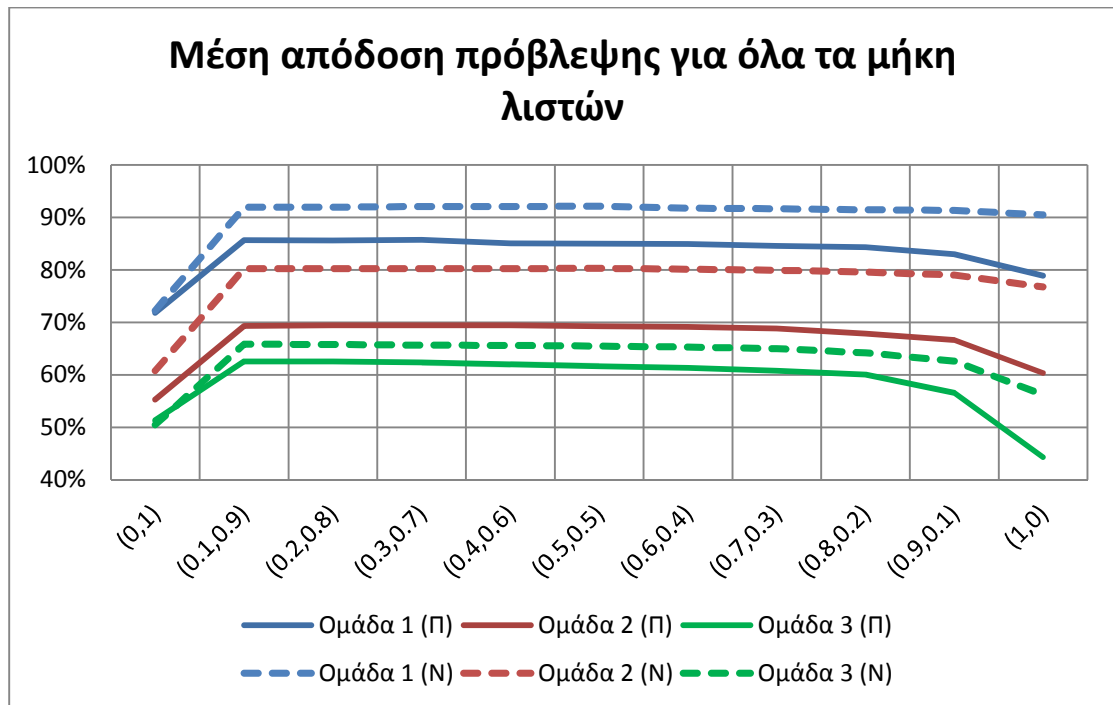
περισσότερες συσκευές, η λίστα με τις συχνές επαφές, η οποία φαίνεται να είναι πιο αποτελεσματική, είναι συχνά λιγότερο εύκολα προσβάσιμη, καθώς βρίσκεται βαθύτερα στη δομή της εφαρμογής της λίστας επαφών. Σε κάθε περίπτωση επιβεβαιώνεται πως οι υπάρχουσες διεπαφές ανάκτησης επαφών δεν είναι βέλτιστες, αλλά υπάρχει σημαντική δυνατότητα για βελτίωση.

4.4.2.2 Κυρίως πειράματα

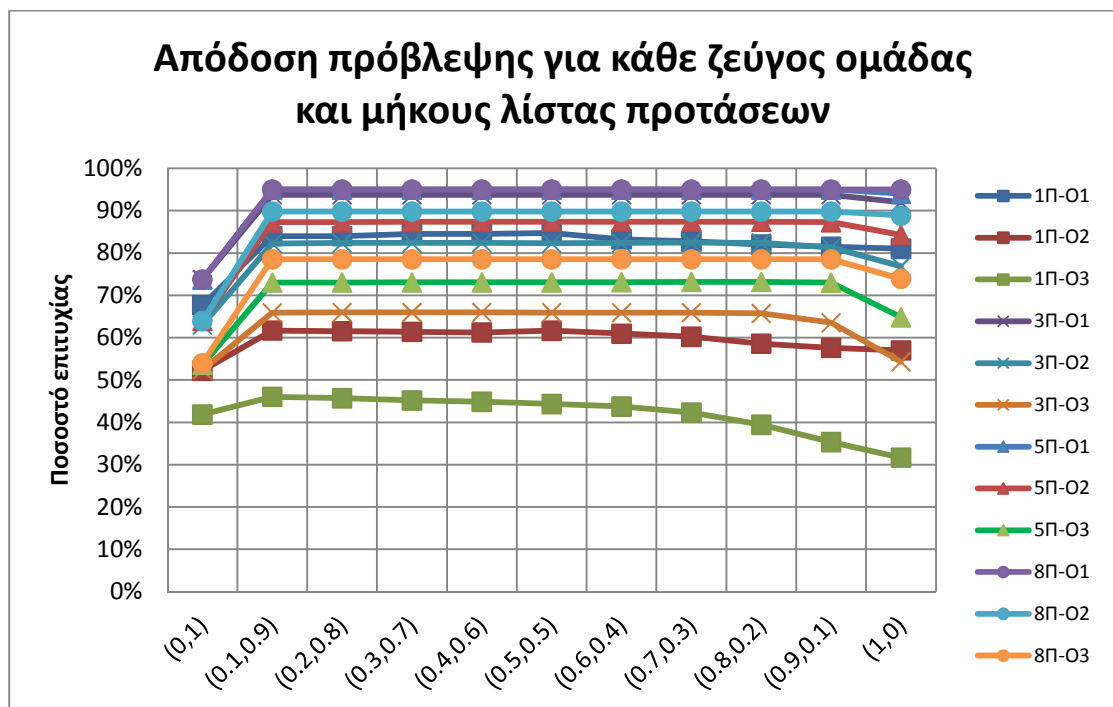
Στα γραφήματα 34 έως 41 παρουσιάζονται τα αποτελέσματα των πειραματικών προσομοιώσεων για το σύνολο δεδομένων της Nokia. Στα περισσότερα από αυτά εμφανίζονται και τα αποτελέσματα από το προκαταρκτικό πείραμα, για ευκολία σύγκρισης. Όπως μπορεί να παρατηρήσει κανείς, οι γραφικές παραστάσεις είναι παρόμοιες και για τα δύο σύνολα δεδομένων, τόσο για την ανάλυση που αφορά το ποσοστό επιτυχίας των προβλέψεων, όσο και για την ανάλυση που αφορά το σκορ, δηλαδή την τοποθέτηση των αποτελεσμάτων στη λίστα πρόβλεψης. Μια παρατήρηση είναι πως και για τις δύο περιπτώσεις, για το σύνολο δεδομένων της Nokia, ο αλγόριθμος πετυχαίνει ελαφρώς χειρότερες επιδόσεις.



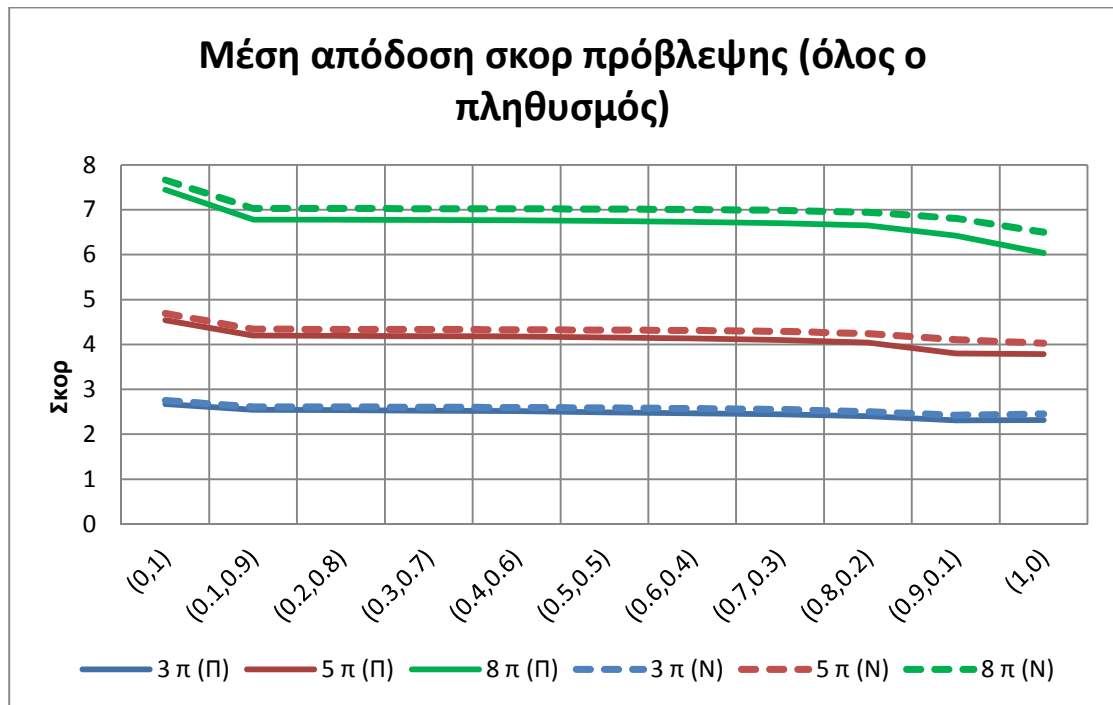
Γράφημα 34. Μέση απόδοση πρόβλεψης για όλο τον πληθυσμό ανά μήκος λίστας προτάσεων. Με συνεχή γραμμή οι παραστάσεις από το προκαταρκτικό πείραμα και με διακεκομμένη από το πείραμα με το σύνολο της Nokia



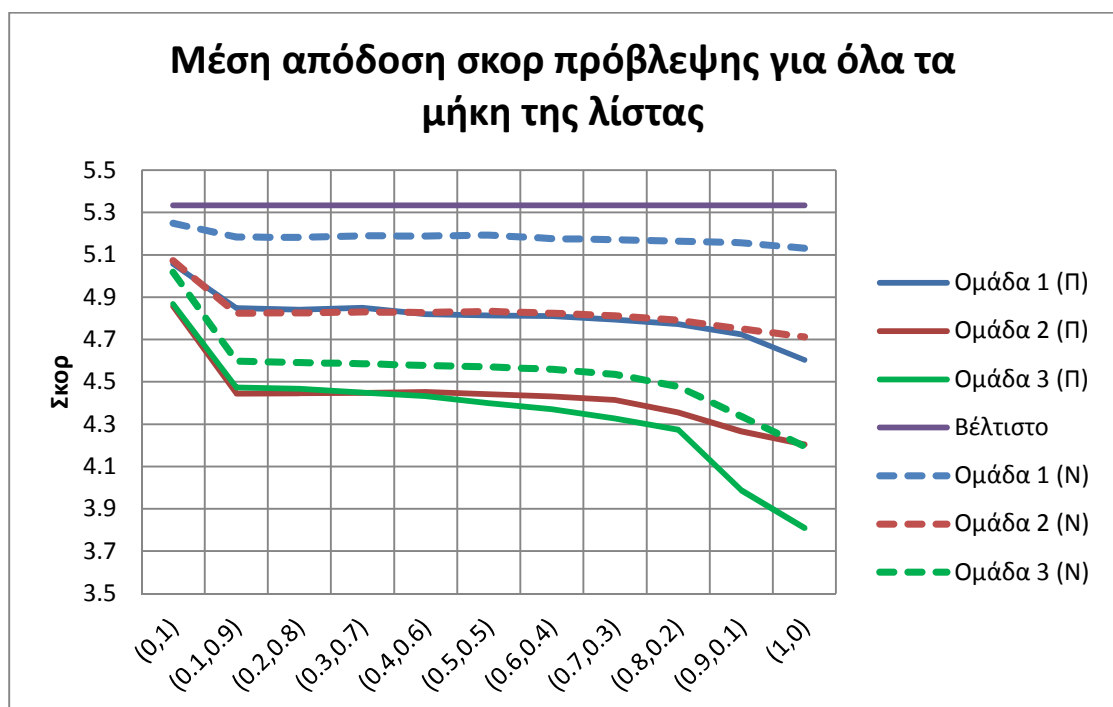
Γράφημα 35. Μέση απόδοση πρόβλεψης για όλα τα μήκη της λίστας προτάσεων ανά ομάδα χρηστών



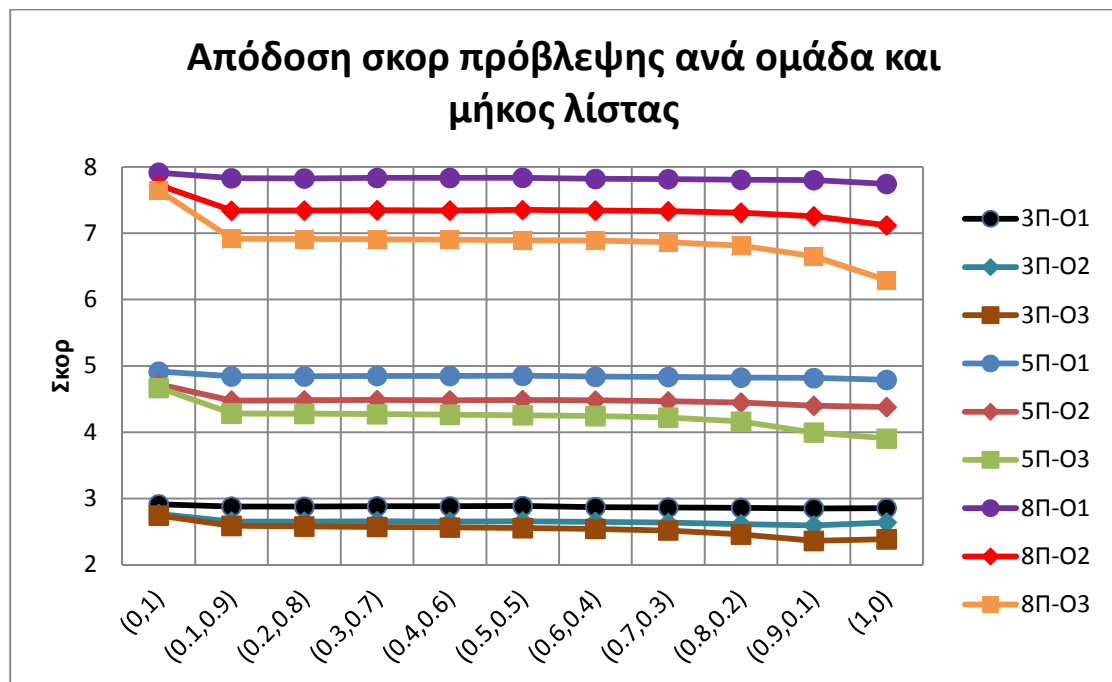
Γράφημα 36. Απόδοση πρόβλεψης για κάθε ζεύγος ομάδας και μήκους λίστας προτάσεων



Γράφημα 37. Μέση απόδοση σκορ πρόβλεψης για όλο τον πληθυσμό των χρηστών



Γράφημα 38. Μέση απόδοση σκορ πρόβλεψης για όλα τα μήκη λιστών. Με μωβ το θεωρητικά βέλτιστο σκορ (5,33)



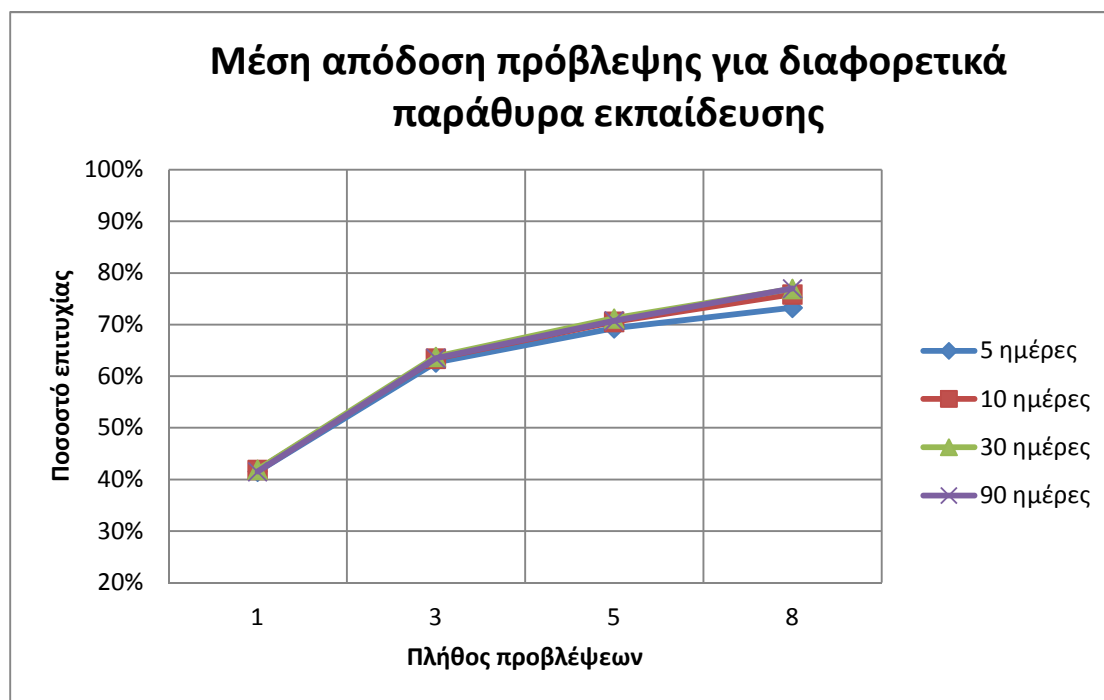
Γράφημα 39. Απόδοση σκορ πρόβλεψης για όλα τα ζεύγη ομάδων και μήκους λίστας

4.4.3 Πειραματισμός με διαφορετικά μήκη παραθύρων εκπαίδευσης

Καθώς τα αποτελέσματα των προβλέψεων για την ομάδα 3 (περισσότερο «κοινωνικοί» χρήστες) ήταν χαμηλότερα σε σύγκριση με τις άλλες δύο ομάδες και η ομάδα αυτή έχει με διαφορά τα περισσότερα μέλη, πραγματοποιήθηκαν μόνο γι' αυτή την ομάδα μερικές επιπλέον πειραματικές εκτελέσεις με διαφορετικά μήκη παραθύρων εκπαίδευσης. Ο αλγόριθμος εκτελέστηκε ξανά για την ομάδα 3, για τα ίδια μήκη λίστας πρόβλεψης (1, 3, 5 και 8 θέσεις) και το ίδιο παράθυρο προσφατότητας (12 ώρες), βάρη $w_i=0,1$ και $w_r=0,9$ (ο καλύτερος συνδυασμός βαρών για την ομάδα 3) ενώ το παράθυρο εκπαίδευσης ήταν μεταβλητό με τιμές 5, 10, 30 ή 90 ημέρες.

Το γράφημα 40 δείχνει πως μεταβάλλεται το ποσοστό επιτυχίας των προβλέψεων για διαφορετικά μήκη του παραθύρου εκπαίδευσης για κάθε μέγεθος της λίστας προβλέψεων. Παρατηρεί κανείς ότι η διακύμανση του ποσοστού επιτυχίας είναι αμελητέα. Το γράφημα 41 δείχνει πως μεταβάλλεται το μέσο σκορ πρόβλεψης για διαφορετικά μήκη του παραθύρου εκπαίδευσης για κάθε μέγεθος της λίστας προβλέψεων. Και πάλι, όπως αναμενόταν, δεν υπάρχουν σημαντικές διαφορές. Επομένως, φαίνεται ότι η επικοινωνιακή συμπεριφορά δεν αλλάζει σημαντικά με την πάροδο του χρόνου για τους χρήστες του συνόλου δεδομένων της Nokia και μπορεί να εξαχθεί από τα δεδομένα 5 ημερών. Αυτό έρχεται σε αντίθεση με την αρχική υπόθεση ότι χρειάζεται να περιλαμβάνει τουλάχιστον την περίοδο μιας εβδομάδας, όμως οι διαφορές στην απόδοση μπορούν να αποδοθούν στη φύση των χρηστών στο

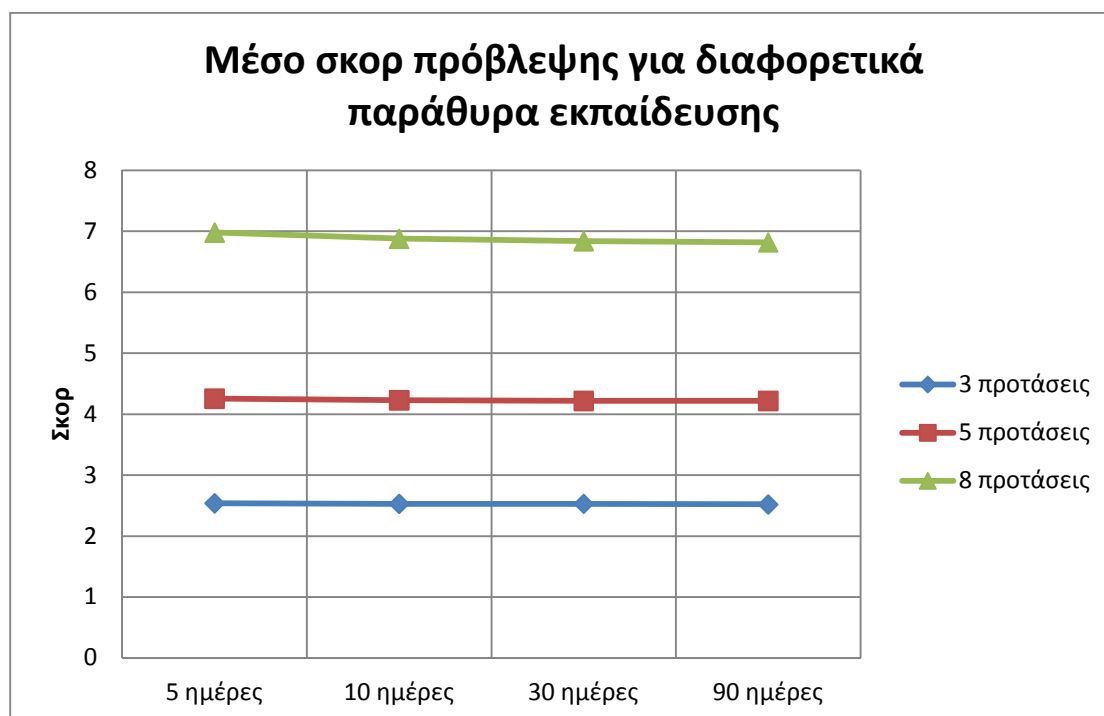
σύνολο δεδομένων. Συγκρίνοντας το μέσο αριθμό κλήσεων που πραγματοποιήθηκαν τα σαββατοκύριακα με όσες πραγματοποιήθηκαν καθημερινές, βρίσκει κανείς μια διαφορά μόνο 1,06 κλήσεις ανά ημέρα (4,65 κλήσεις ανά ημέρα και 5,71 αντίστοιχα). Σε κάθε περίπτωση, φαίνεται ότι ο υπολογιστικός φόρτος μπορεί να μειωθεί με ασφάλεια επιλέγοντας ένα μικρότερο παράθυρο εκπαίδευσης, χωρίς να επηρεαστεί σημαντικά η απόδοση, γεγονός σημαντικό για την εφαρμογή της τεχνικής σε συσκευές κινητού υπολογισμού που διαθέτουν περιορισμένους πόρους.



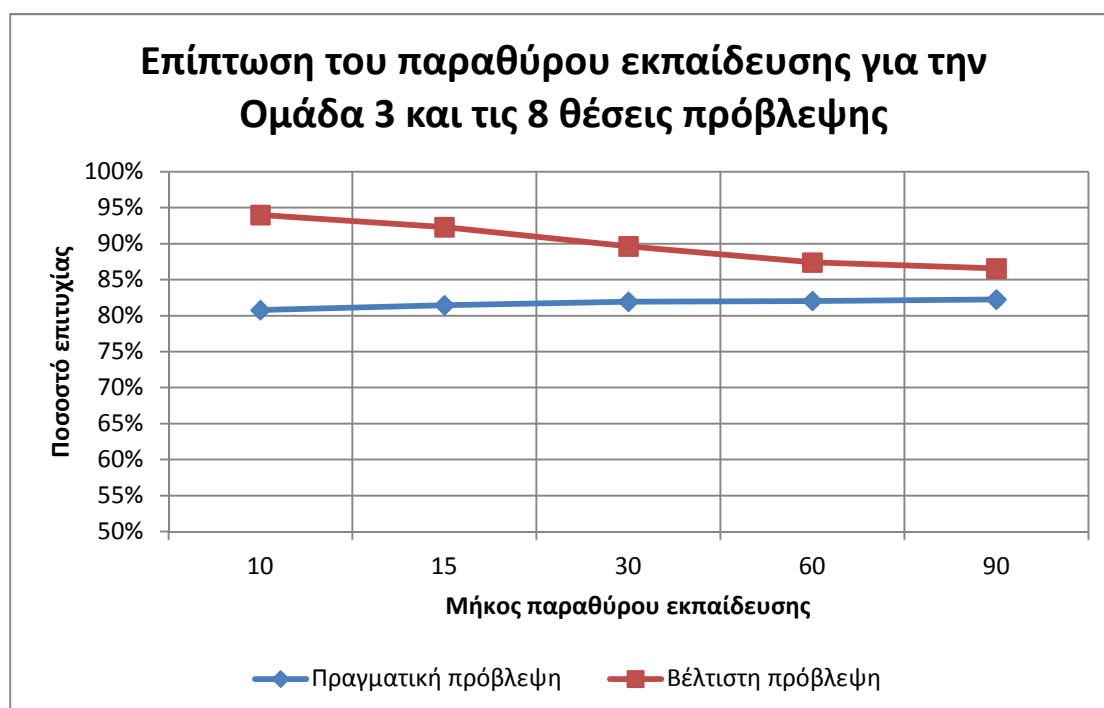
Γράφημα 40. Μέση απόδοση πρόβλεψης για το σύνολο των χρηστών για διαφορετικά μήκη παραθύρου εκπαίδευσης

Όσο μικρότερο είναι το παράθυρο εκπαίδευσης, τόσο περισσότερες είναι οι κλήσεις προς επαφές που δεν μπορούν να προβλεφθούν επειδή δεν υπάρχουν προηγούμενες επικοινωνίες με αυτές την περίοδο που καλύπτει. Αν αφαιρεθούν οι κλήσεις που δεν μπορούν να προβλεφθούν, προκύπτει ένα ακόμα υψηλότερο, πλασματικά βέλτιστο ποσοστό επιτυχίας. Στο γράφημα 42, παρουσιάζεται η εξέλιξη του πραγματικού και του πλασματικού ποσοστού επιτυχίας για την ομάδα 3 και την περίπτωση όπου η λίστα προβλέψεων περιλαμβάνει 8 επαφές (μια οθόνη). Ο αλγόριθμος προβλέπει σωστά το μεγαλύτερο ποσοστό κλήσεων που είναι εφικτό (σχεδόν 95%) για παράθυρο 10 ημερών. Το ποσοστό αυτό μειώνεται σταδιακά μέχρι το 85% για 90 ημέρες, καθώς αυξάνεται το παράθυρο εκπαίδευσης, αφού η μεγαλύτερη χρονική περίοδος που καλύπτει ενσωματώνει περισσότερες επαφές. Επειδή, όμως, πρόκειται για επαφές με μικρό αριθμό κλήσεων και επομένως δύσκολο να προβλεφθούν, η

αύξηση του μήκους του παραθύρου εκπαίδευσης δεν έχει αντίστοιχο αντίκτυπο στην πραγματική απόδοση πρόβλεψης του αλγορίθμου.



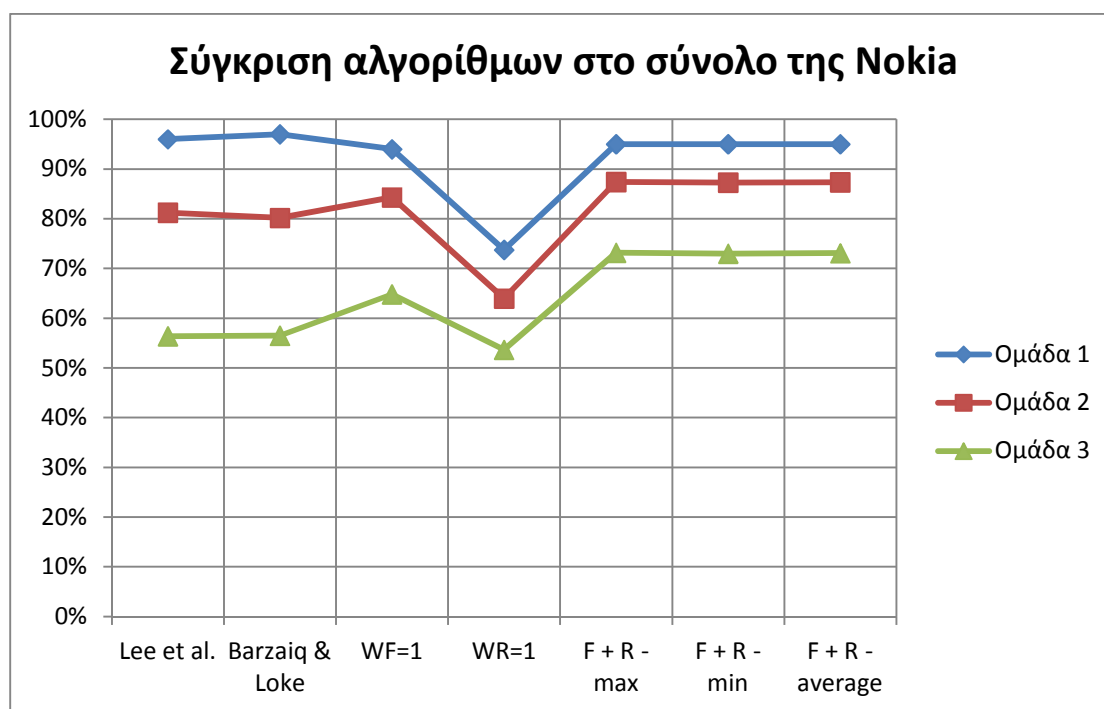
Γράφημα 41. Μέση απόδοση σκορ πρόβλεψης για το σύνολο των χρηστών για διαφορετικά μήκη παραθύρου εκπαίδευσης



Γράφημα 42. Επίπτωση του παραθύρου εκπαίδευσης για την Ομάδα 3 και μήκος λίστας πρόβλεψης 8 θέσεων

4.4.4 Απευθείας σύγκριση με άλλους αλγορίθμους

Το προκαταρκτικό πείραμα αξιολόγησης έδωσε τη δυνατότητα έμμεσης σύγκρισης (πίνακας 15) με τις αποδόσεις επιτυχίας άλλων προσεγγίσεων που προέκυψαν από αντίστοιχες προσομοιώσεις και παρουσιάζονται στη βιβλιογραφία. Επειδή, όμως, η απόδοση των προσεγγίσεων εξαρτάται από τα χαρακτηριστικά των υπό εξέταση χρηστών, όπως είναι η «κοινωνικότητα», για να πραγματοποιηθεί απευθείας σύγκριση, θα έπρεπε να εφαρμοστούν όλες οι προσεγγίσεις στο ίδιο σύνολο δεδομένων. Για το σκοπό αυτό, υλοποιήθηκαν οι αλγόριθμοι που παρουσιάζονται στα (Lee, Seo & Lee, 2010) και (Barzaiq & Loke, 2011). Για τις υπόλοιπες προσεγγίσεις είτε δεν υπήρχαν διαθέσιμες αρκετές λεπτομέρειες και λόγω της πολυπλοκότητάς τους δεν ήταν εφικτό να υλοποιηθούν, είτε δημοσιεύθηκαν σε μεταγενέστερο χρόνο από την πραγματοποίηση των πειραμάτων.



Γράφημα 43. Σύγκριση προτεινόμενης προσέγγισης με άλλες ερευνητικές εργασίες στα δεδομένα του συνόλου της Nokia

Οι αλγόριθμοι που υλοποιήθηκαν εφαρμόστηκαν στο σύνολο δεδομένων της Nokia και τα αποτελέσματα για την περίπτωση που το μήκος της λίστας προτάσεων είναι 5 θέσεις παρουσιάζονται στο γράφημα 43. Επιπλέον, παρουσιάζονται και τα αποτελέσματα για τη λίστα συχνών επαφών ($w_i=1$, $w_i=0$) και για τη λίστα πρόσφατων κλήσεων ($w_i=0$, $w_i=1$). Είναι εμφανές ότι εκτός από την ομάδα 1, όπου όλες οι προσεγγίσεις (με εξαίρεση τη λίστα πρόσφατων κλήσεων) αποδίδουν σχεδόν το ίδιο καλά, για τις ομάδες 2 και 3 που είναι πιο δύσκολο να προβλεφθούν, η προτεινόμενη

προσέγγιση παρέχει σημαντικά καλύτερα αποτελέσματα. Γι' αυτές τις ομάδες, οι αλγόριθμοι των Lee, Seo και Lee (2010) και των Barzaïq και Loke (2011) φαίνεται να είναι λιγότερο αποτελεσματικοί, ακόμα και συγκρινόμενοι με τις παραδοσιακές μεθόδους ανάκτησης.

4.4.5 Συζήτηση αποτελεσμάτων

Τα αποτελέσματα του πειράματος αξιολόγησης του αλγορίθμου στο σύνολο δεδομένων της Nokia επιβεβαίωσαν τα συμπεράσματα που προέκυψαν από το προκαταρκτικό πείραμα. Πρέπει να σημειωθεί ότι το σύνολο δεδομένων της Nokia είναι αυτή τη στιγμή το πληρέστερο δημόσια διαθέσιμο σύνολο δεδομένων που έχουν συλλεχθεί από κινητούς χρήστες, γεγονός που προσδίδει ιδιαίτερη σημασία στα πειραματικά αποτελέσματα.

Ο συνδυασμός διαστάσεων έδωσε καλύτερα αποτελέσματα πρόβλεψης σε σχέση με τη χρησιμοποίηση κάθε διάστασης ξεχωριστά, περιπτώσεις που αντιστοιχούν στις συνηθισμένες διεπαφές ανάκτησης των συσκευών κινητού υπολογισμού (λίστα πρόσφατων κλήσεων και λίστα συχνών επαφών). Επιπλέον, επιβεβαιώθηκε η εξάρτηση της επιτυχίας του αλγορίθμου πρόβλεψης από την «κοινωνικότητα» των χρηστών: χρήστες που επικοινωνούν κυρίως με έναν μικρό πυρήνα επαφών έχουν πιο εύκολα προβλέψιμη συμπεριφορά σε σχέση με όσους είναι περισσότερο «κοινωνικοί». Οι Phithakkitnukoon και Dantu (2011) βρήκαν ότι και για κάθε χρήστη ξεχωριστά, οι επαφές μπορούν να διαχωριστούν σε 3 ομάδες ανάλογα με την κοινωνική τους απόσταση από αυτόν. Οι κλήσεις που πραγματοποιούνται με επαφές που είναι κοινωνικά περισσότερο κοντά στον χρήστη είναι πιο εύκολο να προβλεφθούν σε σχέση με αυτές που πραγματοποιούνται με επαφές που είναι κοινωνικά πιο μακριά. Θα πρέπει να σημειωθεί ότι τα αποτελέσματα σε αυτό το πείραμα ήταν για όλες τις περιπτώσεις ελαφρώς χειρότερα σε σύγκριση με αυτά που προέκυψαν από το προκαταρκτικό πείραμα. Αυτό το γεγονός ήταν εν μέρει αναμενόμενο, εφόσον και για τις 3 ομάδες χρηστών, αλλά ιδιαίτερα για τις ομάδες 2 και 3 που αποτελούν σχεδόν το σύνολο του πληθυσμού, η μέση τιμή της K-Μετρικής ήταν χαμηλότερη σε σχέση με το προκαταρκτικό πείραμα (πίνακας 17). Κατά συνέπεια, οι χρήστες που περιλαμβάνονται στο σύνολο δεδομένων της Nokia εμφανίζουν περισσότερο «κοινωνική» συμπεριφορά σε σχέση με αυτούς που αναλύθηκαν στο προκαταρκτικό πείραμα.

Το γεγονός ότι η εκστρατεία συλλογής δεδομένων της Nokia είχε μεγάλη διάρκεια, είχε ως αποτέλεσμα τα αρχεία κλήσεων των χρηστών να καλύπτουν ένα πολύ μεγαλύτερο χρονικό διάστημα (κατά μέσο όρο μεγαλύτερο από ένα έτος) σε σχέση με

το σύνολο δεδομένων του προκαταρκτικού πειράματος. Αυτό επέτρεψε τον πειραματισμό με μεγαλύτερα μήκη για το παράθυρο εκπαίδευσης (έως 3 μήνες), χωρίς ωστόσο να βρεθούν σημαντικές αποκλίσεις στο ποσοστό επιτυχίας του αλγορίθμου. Μάλιστα, φαίνεται ότι τα πρότυπα συμπεριφοράς των χρηστών είναι επαναλαμβανόμενα και ακόμα κι ένα μικρό παράθυρο 5 ημερών είναι αρκετό για να τα καταγράψει αποτελεσματικά. Αυτό το εύρημα έρχεται σε αντίθεση με τα συμπεράσματα που παρουσιάζονται στα (Phithakkitnukoon & Dantu, 2008) και (Phithakkitnukoon, Dantu, Claxton & Eagle, 2011), όπου αναφέρεται ότι η διαδικασία πρόβλεψης απαιτεί ιστορικά δεδομένα τουλάχιστον ενός ή δύο μηνών ανάλογα με τον τύπο χρήστη. Επιπροσθέτως, αν αγνοηθούν οι κλήσεις προς επαφές για τις οποίες δεν υπάρχουν επικοινωνίες εντός του παραθύρου εκπαίδευσης και συνεπώς δεν μπορούν να προβλεφθούν από το πλαίσιο που προκύπτει από την ανάλυση του ιστορικού κλήσεων, φαίνεται ότι τα ποσοστά επιτυχίας του αλγορίθμου είναι πολύ κοντά στα μέγιστα δυνατά. Εδώ, προκύπτει το ερώτημα αν υπάρχουν διαστάσεις πλαισίου που θα επέτρεπαν την επιτυχή πρόβλεψη των κλήσεων προς αυτές τις επαφές, όπως ενδεχομένως το πλαίσιο που προκύπτει από την ανάλυση των εγγραφών του ημερολογίου ή της επικοινωνίας με άλλα μέσα (π.χ. email, αλληλεπιδράσεις σε υπηρεσίες κοινωνικής δικτύωσης κ.α.). Αυτό το ερώτημα αποκτά ιδιαίτερη σημασία για την περίπτωση της ομάδας 3, όπου φαίνεται να υπάρχει σημαντικό περιθώριο βελτίωσης των αποτελεσμάτων πρόβλεψης. Δυστυχώς, αυτές οι πληροφορίες δεν ήταν διαθέσιμες (για το ημερολόγιο συγκεκριμένα, οι εγγραφές δεν έχουν συσχετιστεί με επαφές ενώ το περιεχόμενό τους είναι κρυπτογραφημένο για λόγους προστασίας της ιδιωτικότητας των συμμετεχόντων). Στο επόμενο κεφάλαιο, εντούτοις, εξετάζονται οι διαστάσεις της φυσικής και σημασιολογικής τοποθεσίας, καθώς και το χρονικό πλαίσιο.

Τέλος, εκτός από τη θεωρητική σύγκριση της απόδοσης της προτεινόμενης προσέγγισης με άλλους αλγορίθμους πρόβλεψης εξερχόμενων κλήσεων από τη βιβλιογραφία, όπου όμως σε κάθε περίπτωση χρησιμοποιούνταν διαφορετικά σύνολα δεδομένων, πραγματοποιήθηκε απευθείας σύγκριση πάνω στα δεδομένα της Nokia. Υλοποιήθηκαν δύο αλγόριθμοι, οι οποίοι ακολουθούν την ίδια λογική αναπαράστασης του πλαισίου με διάνυμα διαστάσεων. Η πειραματική διαδικασία επιβεβαίωσε ξανά την ανωτερότητα της προτεινόμενης προσέγγισης, παρά την απλότητά της, ειδικά για τις ομάδες 2 και 3 των οποίων η συμπεριφορά είναι δυσκολότερο να προβλεφθεί και επομένως τα ποσοστά επιτυχίας του αλγορίθμου χαμηλότερα σε σύγκριση με την ομάδα 1.

4.5 Πείραμα πεδίου για την αξιολογηση του αλγορίθμου

Οι προσομοιώσεις της εκτέλεσης του αλγορίθμου στα διαθέσιμα σύνολα δεδομένων ήταν ένα πρώτο, αλλά σημαντικό βήμα για την αξιολόγησή του, καθώς έδειξαν ότι υπάρχει σημαντικό περιθώριο βελτίωσης για τις παραδοσιακές μεθόδους ανάκτησης των επαφών σε συσκευές κινητού υπολογισμού, δηλαδή για τις λίστες πρόσφατων κλήσεων και συχνών επαφών. Ένα μικρής κλίμακας εργαστηριακό πείραμα (Stefanis, Komninos, Plessas & Garofalakis, 2013) έδειξε ότι η προσαρμοζόμενη λίστα επαφών προσφέρει χαμηλότερους χρόνους ανάκτησης από τις παραδοσιακές λίστες όταν η ζητούμενη επαφή περιλαμβάνεται στις προτάσεις. Για την πληρέστερη τεκμηρίωση αυτών των συμπερασμάτων σε συνθήκες πραγματικής χρήσης, σχεδιάστηκε και εκτελέστηκε ένα πείραμα πεδίου (Plessas, Stefanis, Komninos & Garofalakis, under revision). Είναι σημαντικό να σημειωθεί ότι αντίστοιχο πείραμα πεδίου για μέθοδο πρόβλεψης εξερχόμενων κλήσεων, παρουσιάζεται μόνο στις εργασίες των Lee, Seo και Lee (2010).

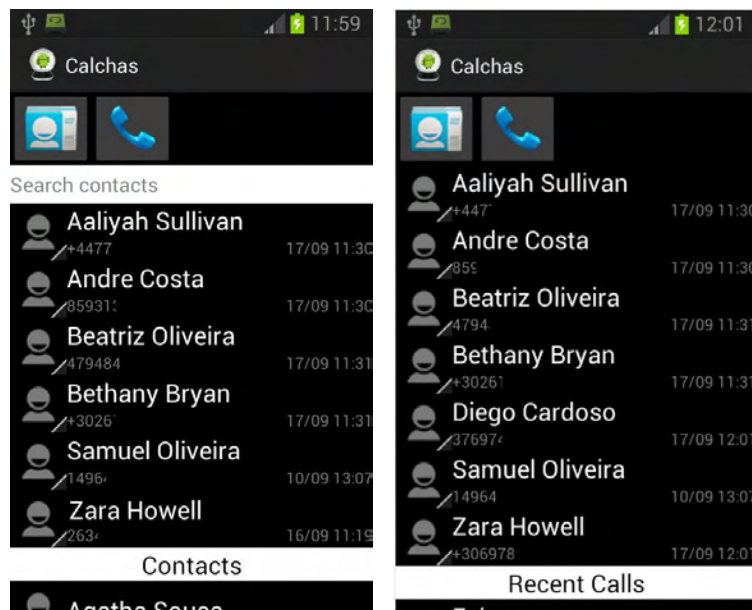
Ακόμα ένας στόχος του πειράματος, που δεν αποτελεί αντικείμενο της παρούσης διδακτορικής διατριβής, ήταν να εξεταστεί κατά πόσο υβριδικές διεπαφές, όπου κάτω από τη λίστα των προτάσεων ενσωματώνονται παραδοσιακές λίστες ως εφεδρικοί μηχανισμοί, προσφέρουν καλύτερες αποδόσεις. Οι χρήστες του πειράματος των Lee, Seo και Lee (2010) ανέφεραν ως σημαντική την ανάγκη ύπαρξης τέτοιων μηχανισμών στην περίπτωση αποτυχίας της λίστας με τις προβλέψεις.

4.5.1 Η εφαρμογή Calchas

Για την πραγματοποίηση του πειράματος πεδίου υλοποιήθηκε μια εφαρμογή με την ονομασία Calchas (εικόνα 26) για συσκευές με το λειτουργικό σύστημα Android. Η εφαρμογή είναι διαθέσιμη στο Google Play³.

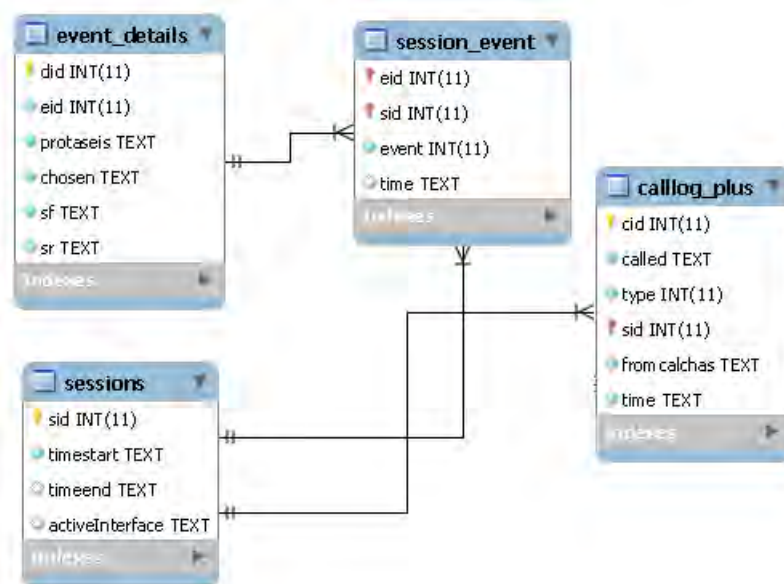
Η εφαρμογή αποτελείται από 3 υποσυστήματα: το υποσύστημα προβλέψεων, το γραφικό περιβάλλον και το υποσύστημα καταγραφής. Το υποσύστημα προβλέψεων υλοποιεί τον αλγόριθμο πρόβλεψης που παρουσιάστηκε στην αρχή του κεφαλαίου. Οι παράμετροι που χρησιμοποιεί είναι παράθυρο εκπαίδευσης 15 ημερών, παράθυρο προσφατότητας 12 ωρών και ίσα βάρη για τις δύο διαστάσεις ($w_t=0,5$ και $w_r=0,5$). Το υποσύστημα αυτό τροφοδοτεί το γραφικό περιβάλλον με τις k πιο πιθανές επαφές να κληθούν κάθε φορά που ο χρήστης ανοίγει την εφαρμογή.

³ <https://play.google.com/store/apps/details?id=com.upatras.ceid.calchasapp&hl=el>



Εικόνα 26. Η Android εφαρμογή Calchas

Το γραφικό περιβάλλον είναι ουσιαστικά το υποσύστημα παρουσίασης των προβλέψεων. Για το πείραμα, τροφοδοτείται με 7 προτάσεις ($k=7$) από το υποσύστημα προβλέψεων, τις οποίες και παρουσιάζει με τη μορφή λίστας στον χρήστη. Στο (Seipp & Devlin, 2013) οι ερευνητές βρήκαν ότι δεν υπάρχουν στατιστικά σημαντικά γρηγορότερες ζώνες σε μια οθόνη κινητής συσκευής για την εργασία της αναζήτησης, όταν ο χρήστης γνωρίζει το όνομα του στόχου. Επίσης, στο (Stefanis, Komninos, Plessas & Garofalakis, 2013) παρουσιάστηκε η υπόθεση ότι η σειρά παρουσίασης των προβλέψεων δεν παίζει σημαντικό ρόλο, καθώς δε βρέθηκαν στατιστικά σημαντικές διαφορές όσον αφορά το χρόνο ανάκτησης για επαφές που βρίσκονται στην κορυφή ή στο κάτω μέρος της προσαρμοζόμενης λίστας. Με αυτά τα δεδομένα και για να διατηρηθεί η αίσθηση οικειότητας των χρηστών με την αλφαβητική λίστα επαφών, η διεπαφή της προσαρμοζόμενης λίστας παρουσιάζει τις επαφές σε αλφαβητική σειρά. Κάτω από την προσαρμοζόμενη λίστα των προβλέψεων υπάρχει διαθέσιμη είτε η αλφαβητική λίστα επαφών είτε η λίστα πρόσφατων κλήσεων, που λειτουργούν ως εφεδρικοί μηχανισμοί ανάκτησης σε περίπτωση αποτυχίας του αλγορίθμου πρόβλεψης. Επιπλέον, υπάρχουν εικονίδια μετάβασης στην προκαθορισμένη εφαρμογή διαχείρισης των επαφών και στο ιστορικό πρόσφατων κλήσεων. Τα εικονίδια αυτά καταλαμβάνουν μια σειρά στην οθόνη και αυτός είναι ο λόγος που παρουσιάζονται 7 προτάσεις αντί για 8, χωρίς να χρειάζεται κάθετη κύλιση.



Εικόνα 27. Σχήμα της ΒΔ για την καταγραφή των δεδομένων

Τέλος, το υποσύστημα καταγραφής υλοποιεί ένα μηχανισμό που καταγράφει τη χρήση της εφαρμογής, κάτι που είναι απαραίτητο για την ποσοτική αξιολόγησή της. Τα ανώνυμα στατιστικά αποθηκεύονται σε μια SQLite ΒΔ (εικόνα 27) στη συσκευή του χρήστη και στέλνονται κατόπιν επιλογής σε έναν εξυπηρετητή. Όταν η εφαρμογή έρχεται στο προσκήνιο, μια νέα συνεδρία χρήσης καταγράφεται. Σε μια συνεδρία χρήσης, ο μηχανισμός καταγραφής αποθηκεύει το χρόνο έναρξης και λήξης της συνεδρίας και το γεγονός που έλαβε χώρα. Έχουν οριστεί 3 είδη γεγονότων: α) ο χρήστης κάνει κλικ πάνω σε μια επαφή και ξεκινά μια εξερχόμενη κλήση, β) ο χρήστης ανοίγει την προκαθορισμένη λίστα επαφών και γ) ο χρήστης ανοίγει τη λίστα πρόσφατων κλήσεων. Σε όλες τις περιπτώσεις, καταγράφεται ο χρόνος τη στιγμή που το γεγονός συνέβη. Επιπλέον, για το γεγονός (α) καταγράφεται η πλήρης λίστα των προβλέψεων (id κλήσης ή κρυπτογραφημένος αριθμός εάν δεν πρόκειται για επαφή), τα σκορ των διαστάσεων της συχνότητας και προσφατότητας επικοινωνίας για κάθε πρόβλεψη και η διεπαφή στην οποία η επαφή βρισκόταν (λίστα προτάσεων ή ενσωματωμένη εφεδρική λίστα). Τέλος, όταν ο χρήστης κάνει μια εξερχόμενη κλήση ή δέχεται μια εισερχόμενη, ένας broadcast receiver την αποθηκεύει στη Β.Δ., δημιουργώντας έτσι ένα αντίγραφο του αρχείου κλήσεων.

4.5.2 Σύνολο δεδομένων

Για την πραγματοποίηση του πειράματος πεδίου απευθύνθηκε μια δημόσια πρόσκληση στην πανεπιστημιακή κοινότητα για συμμετοχή. Αρχικά, ανταποκρίθηκαν 103 άτομα. Κάθε ενδιαφερόμενος ενημερώθηκε μέσω μηνύματος ηλεκτρονικού ταχυδρομείου για τη διαδικασία και τα απαραίτητα βήματα

συμμετοχής, καθώς και για τη συλλογή ανώνυμων δεδομένων. Επιπλέον, δημιουργήθηκε ένα ενημερωτικό βίντεο διάρκειας 3 λεπτών σχετικά με τη διαδικασία και αναρτήθηκε στο Youtube. Οι μισοί συμμετέχοντες θα ξεκινούσαν το πείραμα με τη διεπαφή που είχε ενσωματωμένη τη λίστα επαφών και οι άλλοι μισοί με τη διεπαφή που είχε ενσωματωμένη τη λίστα πρόσφατων κλήσεων. Αυτό ήταν απαραίτητο για την αξιολόγηση των εφεδρικών μηχανισμών σε περίπτωση αποτυχίας, που όπως έχει ήδη αναφερθεί δεν απασχολεί την παρούσα διδακτορική διατριβή. Στο τέλος του πειράματος, το οποίο διήρκεσε 30 ημέρες, οι χρήστες έλαβαν μια ειδοποίηση για την αποστολή των ανώνυμων δεδομένων που είχαν συλλεχθεί, ενώ κλήθηκαν να συμπληρώσουν και ένα ηλεκτρονικό ερωτηματολόγιο.

Από τα 103 άτομα που έδειξαν αρχικά ενδιαφέρον, παρελήφθησαν δεδομένα από 73. Για την ποσοτική και ποιοτική ανάλυση που ακολουθεί, αποκλείστηκαν εκείνοι οι χρήστες που δεν είχαν συμπληρώσει 30 ημέρες χρήσης, εκείνοι που χρησιμοποίησαν κυρίως τη μια από τις δύο διεπαφές (>75% χρήση της διεπαφής) και εκείνοι που άνοιξαν την εφαρμογή πολύ λίγες φορές (π.χ. 3 φορές σε 30 ημέρες). Τέλος, αποκλείστηκαν οι συμμετέχοντες που δεν συμπλήρωσαν το ερωτηματολόγιο. Τελικά, έμειναν 44 συμμετέχοντες με έγκυρα δεδομένα και συμπληρωμένα ερωτηματολόγια. Από αυτούς, 34,1% (15) ήταν γυναίκες. Όσον αφορά τις ηλικίες τους, 63,6% (28) ήταν μεταξύ 18-24 ετών, 31,8% (14) μεταξύ 25-34 και 4,5% (2) ήταν μεταξύ 35-44 ετών.

4.5.3 Πειραματικά αποτελέσματα

4.5.3.1 Ποσοτικά δεδομένα

Στον πίνακα 18 παρουσιάζεται μια σύνοψη των μετρήσεων για τις παραμέτρους που καταγράφονταν σε αυτό το πείραμα πεδίου. Διαπιστώνεται ότι κατά μέσο όρο το 59,05% των εξερχόμενων κλήσεων των συμμετεχόντων κατά τη διάρκεια του πειράματος πραγματοποιήθηκαν από την εφαρμογή Calchas. Όπως αναμενόταν, φαίνεται να υπάρχουν καταστάσεις στις οποίες είναι ευκολότερο για τον χρήστη να υιοθετήσει άλλες μεθόδους για την έναρξη μιας εξερχόμενης κλήσης, όπως για παράδειγμα στην περίπτωση μιας ειδοποίησης μετά από μη απαντημένη κλήση ή όταν λαμβάνει ένα γραπτό μήνυμα. Για όλες τις περιπτώσεις που η εφαρμογή δε χρησιμοποιήθηκε δεν μπορεί να είναι κανείς σίγουρος εάν αυτό συνέβη διότι οι συμμετέχοντες προτίμησαν να χρησιμοποιήσουν κάποια άλλη μέθοδο ανάκτησης (π.χ. λίστα αγαπημένων επαφών, λίστα πρόσφατων κλήσεων, λίστα επαφών κ.α.) καθώς δεν ήταν συνηθισμένοι σε εναλλακτικές μεθόδους ανάκτησης από το αποθετήριο των επαφών ή επειδή απλά δεν κατάφεραν να βρουν την επαφή που έψαχναν στην εφαρμογή. Για να αποκτηθεί μια καλύτερη εκτίμηση της πραγματικής

χρήσης της εφαρμογής, λήφθηκαν υπόψη κλήσεις που έγιναν εντός ενός λεπτού από τη στιγμή που μια συνεδρία χρήσης της εφαρμογής (που δεν οδήγησε σε κλήση) ολοκληρώθηκε, θεωρώντας αυτές τις κλήσεις ως πραγματικές αποτυχίες της εφαρμογής να εξυπηρετήσει τους χρήστες (π.χ. ο χρήστης άνοιξε την εφαρμογή, δεν κατάφερε να εντοπίσει τη ζητούμενη επαφή, στη συνέχεια άνοιξε μια άλλη εφαρμογή που παρέχει πρόσβαση στις επαφές και πραγματοποίησε την κλήση). Όταν αφαιρούνται οι κλήσεις για τις οποίες η εφαρμογή δε χρησιμοποιήθηκε σε αυτό το σύντομο χρονικό διάστημα, το ποσοστό των κλήσεων που πραγματοποιήθηκαν από την εφαρμογή αυξάνεται στο 85%. Αυτό το γεγονός δείχνει ότι όταν η εφαρμογή χρησιμοποιόταν σκόπιμα, οι συμμετέχοντες κατάφερναν συνήθως να ανακτήσουν τη ζητούμενη επαφή. Αυτό, επίσης, εξηγεί και τους μέσους αριθμούς από κλικ στα εικονίδια μετάβασης στη λίστα επαφών και τη λίστα πρόσφατων κλήσεων, που ήταν ιδιαίτερα χαμηλοί (mean=8,32, sd=10,61 και mean=8,73, sd=9,85 αντίστοιχα). Η μέση διάρκεια των συνεδριών που κατέληξαν σε εξερχόμενες κλήσεις αντιστοιχεί στο μέσο χρόνο ανάκτησης των επαφών, ο οποίος ήταν 5,61 δευτερόλεπτα (sd=1,51). Επιπλέον, μπορεί να παρατηρήσει κανείς ότι ο μέσος αριθμός από κλήσεις για την ενσωματωμένη λίστα (είτε τη λίστα επαφών είτε τη λίστα πρόσφατων κλήσεων) που βρίσκεται κάτω από τη λίστα των προτάσεων ήταν σημαντικά χαμηλότερος (αναλογία 1:14) από τον αριθμό κλήσεων από τη λίστα των προβλέψεων. Αυτή η παρατήρηση υπονοεί ότι το ποσοστό επιτυχίας του αλγορίθμου ήταν τέτοιο που επέτρεπε συνήθως στους χρήστες να ανακτήσουν τη ζητούμενη επαφή από την πρώτη οθόνη της εφαρμογής, χωρίς να υπάρχει ανάγκη κύλισης ή αναζήτησης.

Πίνακας 18. Σύνοψη πειραματικών μετρήσεων

	Ελάχιστο	Μέγιστο	Μέση τιμή	Τυπική απόκλιση
<i>Πλήθος συνεδριών χρήσης</i>	58	1248	401,66	245,21
<i>Μέση διάρκεια συνεδριών (δευτερόλεπτα)</i>	4,09	32,98	8,70	5,71
<i>Μέση διάρκεια συνεδριών (δευτερόλεπτα) που κατέληξαν σε κλήση</i>	3,15	9,19	5,61	1,51
<i>Συνολικές εξερχόμενες κλήσεις</i>	65	651	224,73	103,12
<i>Εξερχόμενες κλήσεις από την εφαρμογή Calchas</i>	20	408	132,23	74,27
<i>Ποσοστό κλήσεων από την εφαρμογή Calchas</i>	12,43%	92,41%	59,05%	19,73%

<i>Εκτιμώμενο ποσοστό κλήσεων από την εφαρμογή Calchas</i>	63,28%	100%	84,93%	7,97%
<i>Κλήσεις από τη λίστα προτάσεων (στην εφαρμογή Calchas)</i>	17	363	115,57	68,88
<i>Κλήσεις από τη λίστα πρόσφατων κλήσεων (στην εφαρμογή Calchas)</i>	0	31	8,64	6,94
<i>Κλήσεις από τη λίστα επαφών (στην εφαρμογή Calchas)</i>	0	25	8,07	5,84
<i>Αριθμός κλικ στο εικονίδιο μετάβασης στο αρχείο κλήσεων</i>	0	43	8,73	9,85
<i>Αριθμός κλικ στο εικονίδιο μετάβασης στη λίστα επαφών</i>	0	57	8,32	10,61

Στη συνέχεια, πραγματοποιήθηκε ανάλυση των δεδομένων ανά ομάδα «κοινωνικότητας». Για το σκοπό αυτό, υπολογίστηκε η Κ-Μετρική για κάθε συμμετέχοντα, για τις κλήσεις που καταγράφησαν κατά τη διάρκεια του πειράματος. Βρέθηκε ότι η λιγότερη «κοινωνική» ομάδα είχε 1 χρήστη, η μέτρια «κοινωνική» ομάδα είχε 12 χρήστες, ενώ η περισσότερο «κοινωνική» ομάδα 31 χρήστες. Η χρήση αυτής της μετρικής για την περαιτέρω υποστήριξη των αποτελεσμάτων προσφέρει περισσότερο σημαίνουσες ερμηνείες των δεδομένων και επιτρέπει την καλύτερη κατανόηση της συμπεριφοράς των συμμετεχόντων.

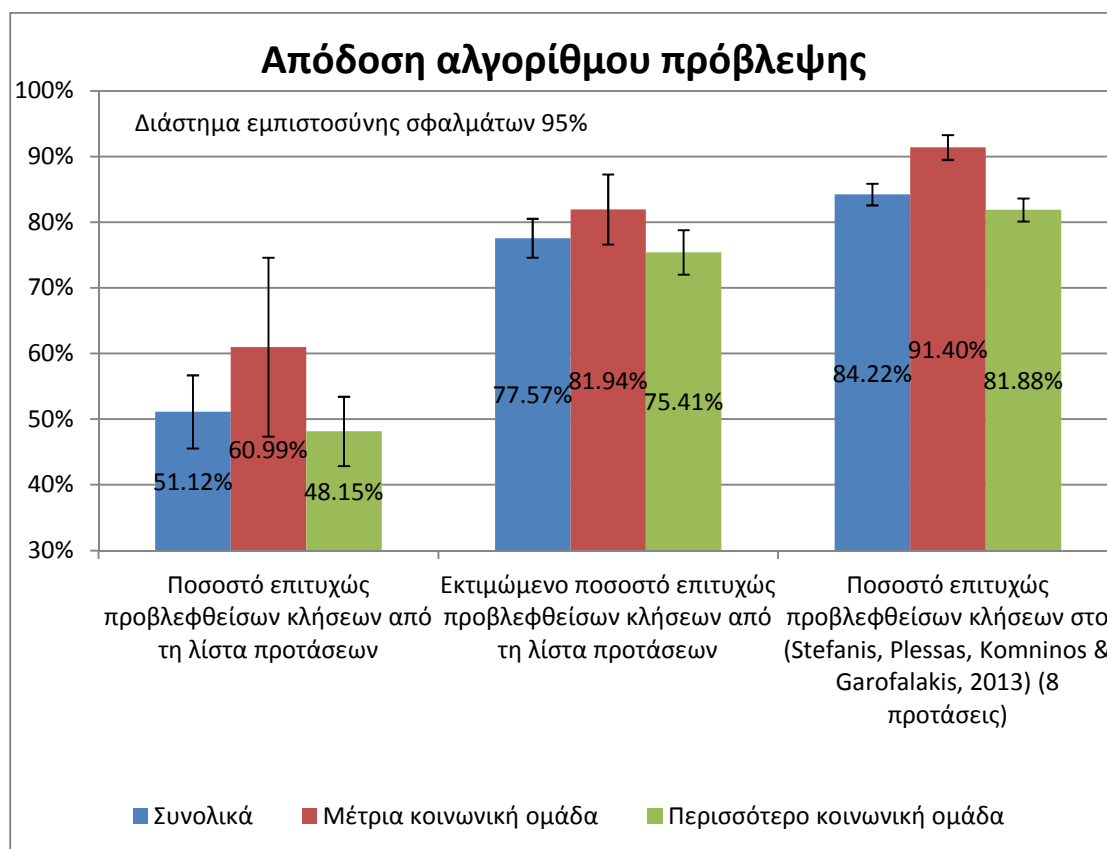
Ο πίνακας 19 δείχνει τα ποσοτικά αποτελέσματα του πειράματος ομαδοποιημένα με βάση την «κοινωνικότητα» (η λιγότερο κοινωνική ομάδα περιελάμβανε μόνο έναν χρήστη και αποκλείστηκε από την ανάλυση). Για τις περισσότερες παραμέτρους δεν προέκυψαν στατιστικά σημαντικές διαφορές. Εντούτοις, όσον αφορά το ποσοστό εξερχόμενων κλήσεων από την εφαρμογή, ένα Mann-Whitney U Test αποκάλυψε μια στατιστικά σημαντική διαφορά ($U=109$, $p<0.05$), δείχνοντας ότι η λιγότερο «κοινωνική» ομάδα πραγματοποίησε σχεδόν 9% περισσότερες κλήσεις μέσω της εφαρμογής Calchas. Αυτό το αποτέλεσμα ήταν αναμενόμενο, καθώς η συμπεριφορά των χρηστών που ανήκουν στην περισσότερο «κοινωνική» ομάδα έχει αποδειχθεί ότι είναι πιο δύσκολο να προβλεφθεί. Ενδιαφέρον, επίσης, έχει το γεγονός ότι ο μέσος αριθμός κλήσεων είναι ελαφρώς μεγαλύτερος για τους χρήστες της ομάδας 2 σε σχέση με την ομάδα 3, ένα εύρημα που επιβεβαιώνει την υπόθεσή μας ότι η ένταση της επικοινωνίας δεν είναι από μόνη της δείκτης «κοινωνικότητας», σε αντίθεση με την Κ-Μετρική που εφαρμόστηκε.

Πίνακας 19. Ανάλυση των μεταβλητών ανά ομάδα «κοινωνικότητας» (2 = μέτρια «κοινωνικοί», 3 = περισσότερο «κοινωνικοί»)

Μεταβλητή	Πλήθος		Μέση τιμή		Τυπική απόκλιση	
	2	3	2	3	2	3
<i>Μέση διάρκεια συνεδριών (δευτερόλεπτα) που κατέληξαν σε κλήση</i>	12	31	4,95	5,90	1,14	1,57
<i>Συνολικές εξερχόμενες κλήσεις</i>	12	31	231,08	227,06	154,31	75,84
<i>Εξερχόμενες κλήσεις από την εφαρμογή Calchas</i>	12	31	152,00	128,19	111,09	52,80
<i>Ποσοστό κλήσεων από την εφαρμογή Calchas</i>	12	31	66,64%	57,17%	24,94%	16,35%
<i>Εκτιμώμενο ποσοστό κλήσεων από την εφαρμογή Calchas</i>	12	31	86,77%	83,88%	8,6%	7,61%
<i>Κλήσεις από τη λίστα προτάσεων (στην εφαρμογή Calchas)</i>	12	31	138,92	109,68	100,75	50,08
<i>Κλήσεις από τη λίστα πρόσφατων κλήσεων (στην εφαρμογή Calchas)</i>	12	31	6,58	9,71	8,93	5,86
<i>Κλήσεις από τη λίστα επαφών (στην εφαρμογή Calchas)</i>	12	31	6,58	8,84	5,77	5,83

Τα αποτελέσματα του πειράματος πεδίου επιβεβαιώνουν τα ευρήματα των προηγούμενων πειραμάτων προσομοίωσης, όπου εκτελέστηκε ο αλγόριθμος κάτω από ιδανικές συνθήκες, θεωρώντας ότι η προτεινόμενη μέθοδος ανάκτησης υιοθετήθηκε για κάθε εξερχόμενη κλήση. Συγκεκριμένα, πραγματοποιήθηκε σύγκριση με τα αποτελέσματα της προσομοίωση στο σύνολο δεδομένων της Nokia (Stefanis, Plessas, Komninos & Garofalakis). Στο γράφημα 44 παρουσιάζονται τα ποσοστά επιτυχίας του αλγορίθμου πρόβλεψης. Για τον υπολογισμό της απόδοσης πρόβλεψης στο πείραμα πεδίου, θεωρούνται ως επιτυχείς προβλέψεις μόνο κλήσεις που πραγματοποιήθηκαν από τη λίστα των προτάσεων και όχι από τις ενσωματωμένες λίστες, ώστε να μπορεί να γίνει απευθείας σύγκριση με τα προηγούμενα ευρήματα. Παρατηρείται ότι το εκτιμώμενο ποσοστό επιτυχίας του αλγορίθμου είναι μόνο ελαφρώς χαμηλότερο από αυτό που υπολογίστηκε όταν εφαρμόστηκε στο σύνολο δεδομένων της Nokia, όπου χρησιμοποιήθηκε μια λίστα προτάσεων με 8 επαφές αντί για 7. Επιπλέον, επιβεβαιώνεται ότι το ποσοστό επιτυχίας είναι υψηλότερο για χρήστες της ομάδας μέτριας «κοινωνικότητας» σε

σχέση με τους περισσότερο «κοινωνικούς» χρήστες, των οποίων οι αλληλεπιδράσεις επικοινωνίας είναι δυσκολότερο να προβλεφθούν.



Γράφημα 44. Απόδοση πρόβλεψης του αλγορίθμου σε σύγκριση με το (Stefanis, Plessas, Komninos & Garofalakis, 2013)

Επιπλέον, μπορεί να γίνει σύγκριση των ευρημάτων με αυτά που αναφέρονται στο Lee, Seo και Lee (2010), όπου επίσης πραγματοποιήθηκε αξιολόγηση της προσαρμοζόμενης λίστας επαφών στο πεδίο με ένα πείραμα που αφορούσε 10 συμμετέχοντες και είχε διάρκεια 3 μήνες. Οι ερευνητές έλαβαν υπόψη τους μόνο κλήσεις που πραγματοποιήθηκαν από την εφαρμογή τους και ως εκ τούτου το μέτρο του ποσοστού επιτυχίας αντιστοιχεί στην εκτιμώμενη απόδοση του αλγορίθμου πρόβλεψης. Στο εν λόγω πείραμα, αναφέρουν έναν αριθμό επιτυχώς προβλεφθεισών κλήσεων που αντιστοιχεί σε ποσοστό επιτυχίας 72,62%, ενώ όπως φαίνεται από το γράφημα 44 στο πείραμα που παρουσιάζεται το ποσοστό αυτό ήταν 77,57%. Εντούτοις, η λίστα προτάσεων περιείχε 5 επαφές. Από την άλλη, δεν υπάρχει κάποια ένδειξη για την «κοινωνικότητα» των χρηστών, που αποτελεί ένα σημαντικό παράγοντα στην πρόβλεψη των εξερχόμενων κλήσεων. Επιπλέον, το πλήθος των κλήσεων που πραγματοποιήθηκε από την εφαρμογή τους ήταν σχεδόν 12 φορές μικρότερο σε σχέση με αυτό που παρουσιάζεται σε αυτή την ενότητα (μόλις 427 έναντι 5.085)

4.5.3.2 Ποιοτικά δεδομένα

Οι συμμετέχοντες κλήθηκαν μετά το τέλος του πειράματος να απαντήσουν ένα ηλεκτρονικό ερωτηματολόγιο. Το ερωτηματολόγιο περιελάμβανε 3 μέρη: το πρώτο, το οποίο δεν παρουσιάζεται, αφορούσε ερωτήσεις για τη σύγκριση των δύο διαφορετικών υβριδικών διεπαφών. Το δεύτερο στόχευε στη λήψη γενικής πληροφόρησης για τη σύγκριση της εφαρμογής με τις παραδοσιακές μεθόδους ανάκτησης, ενώ το τρίτο επέτρεπε στους συμμετέχοντες να προσφέρουν ανατροφοδότηση με τη μορφή ελεύθερου κειμένου.

Ο πίνακας 20 δείχνει, σε φθίνουσα σειρά, πόσο συχνά (1=σπάνια, 5=πολύ συχνά) οι συμμετέχοντες δήλωσαν ότι χρησιμοποιούσαν τις παραδοσιακές μεθόδους που προσφέρουν οι συσκευές για την έναρξη μιας εξερχόμενης κλήση πριν από το πείραμα. Η λίστα πρόσφατων κλήσεων (δεύτερη πιο δημοφιλής επιλογή στο (Lee, Seo & Lee, 2010) είναι το μέρος το οποίο οι χρήστες συνηθίζουν να επισκέπτονται πιο συχνά προκειμένου να ανακτήσουν μια επαφή, ενώ τα popups των ειδοποιήσεων φαίνεται ότι διευκολύνουν αλληλεπιδράσεις με τις επαφές και η λίστα επαφών ακολουθεί σε δημοφιλία.

Πίνακας 20. Συχνότητα χρησιμοποίησης παραδοσιακών μεθόδων ανάκτησης επαφών για έναρξη εξερχόμενων κλήσεων

	Μέση τιμή	Τυπική απόκλιση
<i>Λίστα πρόσφατων κλήσεων</i>	4,57	0,873
<i>Ειδοποιήσεις</i>	3,95	0,963
<i>Λίστα επαφών</i>	3,43	1,021
<i>Λίστα αγαπημένων επαφών</i>	2,18	1,483
<i>SMS</i>	1,66	0,888
<i>Ομάδες επαφών</i>	1,36	0,810

Ο πίνακας 21 παρουσιάζει τις απαντήσεις των συμμετεχόντων στην ερώτηση «πόσο βοηθητική βρήκατε την εφαρμογή Calchas σε σχέση με τις παρακάτω εναλλακτικές μεθόδους για την πραγματοποίηση μιας εξερχόμενης κλήσης;» (1=πολύ λιγότερο, 5=πολύ περισσότερο). Μπορεί να παρατηρήσει κανείς ότι οι συμμετέχοντες βρήκαν την εφαρμογή κατά μέσο όρο ελαφρώς πιο βοηθητική σε σχέση με όλες τις εναλλακτικές μεθόδους ανάκτησης επαφών. Μια ενδιαφέρουσα παρατήρηση είναι ότι ο Calchas αξιολογείται κατά μέσο όρο περισσότερο βοηθητικός από τη λίστα πρόσφατων κλήσεων και τη λίστα επαφών, ενώ αξιολογείται σχεδόν το ίδιο με τις ειδοποιήσεις. Αυτό το εύρημα υπογραμμίζει τη χρησιμότητα κάθε εργαλείου

ανάκτησης εντός του πλαισίου στο οποίο χρησιμοποιείται και ότι οι συμμετέχοντες είναι πιθανό να εξυπηρετούνται καλύτερα από πολλαπλά εργαλεία και όχι μόνο από ένα.

Πίνακας 21. Σύγκριση της προτίμησης υπέρ της εφαρμογής σε σχέση με τις παραδοσιακές μεθόδους ανάκτησης επαφών για την έναρξη εξερχόμενων κλήσεων

	Μέση τιμή	Τυπική απόκλιση
<i>Calchas - Λίστα πρόσφατων κλήσεων</i>	3,73	0,973
<i>Calchas - Ειδοποιήσεις</i>	3,05	1,033
<i>Calchas - Λίστα επαφών</i>	3,64	1,014
<i>Calchas - Λίστα αγαπημένων επαφών</i>	3,50	1,355
<i>Calchas - SMS</i>	3,32	1,343
<i>Calchas - Ομάδες επαφών</i>	3,25	1,366

Αναλύοντας τα σχόλια των συμμετεχόντων σχετικά με την εμπειρία τους με την εφαρμογή, επιβεβαιώνεται ότι σε γενικές γραμμές τη βρήκαν χρήσιμη («πολύ χρήσιμη εφαρμογή, γρήγορη και εύκολη στη χρήση», «μου άρεσε πάρα πολύ» κ.λπ.). Κάποιοι συμμετέχοντες πρότειναν έμμεσα την ενσωμάτωση και άλλων διαστάσεων πλαισίου για τον αλγόριθμο πρόβλεψης («θα ήταν χρήσιμο αν υπήρχε διασύνδεση με το ημερολόγιό μου και οι προτάσεις της λίστας περιελάμβαναν επαφές την ημέρα των γενεθλίων τους») και εξέφρασαν την επιθυμία τους για παραμετροποίηση («θα ήθελα να μπορούσα να αλλάζω το μέγεθος της λίστας προτάσεων», «θα προτιμούσα να θέτω χειροκίνητα τα βάρη των κριτηρίων πρόβλεψης»). Οι περισσότεροι από τους χρήστες βρήκαν τον αλγόριθμο πρόβλεψης επαρκώς ακριβή («στις περισσότερες περιπτώσεις ήμουν πολύ ικανοποιημένος με τις προτεινόμενες επαφές», «η εφαρμογή προέβλεψε ικανοποιητικά τις περισσότερες από τις κλήσεις που ήθελα να κάνω»). Τα περισσότερα αρνητικά σχόλια αφορούσαν το γραφικό περιβάλλον («δε μου άρεσε το γραφικό περιβάλλον», «η εμπειρία μου με το γραφικό περιβάλλον ήταν αρκετά άσχημη»). Καθώς ο σχεδιασμός του γραφικού περιβάλλοντος δεν ήταν πολύ διαφορετικός από τις υπάρχουσες διεπαφές, αυτή είναι μια ένδειξη ότι οι χρήστες δεν είναι ιδιαίτερα ικανοποιημένοι με τις παραδοσιακές διεπαφές ανάκτησης επαφών.

4.5.4 Συζήτηση αποτελεσμάτων

Παρότι αλγόριθμοι πρόβλεψης για προσαρμοζόμενες λίστες επαφών έχουν προταθεί και αξιολογηθεί μέσω προσομοίωσης στη βιβλιογραφία, με την εξαίρεση του (Lee, Seo και Lee, 2010) όπου μόνο 10 συμμετέχοντες έλαβαν μέρος στο πείραμα πεδίου και παρότι η διάρκεια του ήταν 3 μήνες, πραγματοποίησαν ένα σημαντικά μικρότερο αριθμό κλήσεων σε σχέση με αυτόν που προέκυψε από το πείραμα που

παρουσιάστηκε νωρίτερα, στις υπόλοιπες περιπτώσεις δεν έχουν αξιολογηθεί στο πεδίο. Σκοπός του πειράματος πεδίου που παρουσιάστηκε ήταν να αντιμετωπίσει αυτή την έλλειψη στη βιβλιογραφία, εξετάζοντας την απόδοση του προτεινόμενου αλγορίθμου σε πραγματικές συνθήκες χρήσης.

Αυτή η αξιολόγηση της χρήσης του αλγορίθμου εκτός του εργαστηρίου, στον πραγματικό κόσμο, είναι μια σημαντική συνεισφορά του πειράματος. Οι συμμετέχοντες πραγματοποίησαν, κατά μέσο όρο, σχεδόν το 60% των συνολικών εξερχόμενων κλήσεων από την εφαρμογή. Όταν λαμβάνονται υπόψη εκείνες οι κλήσεις που πραγματοποιήθηκαν σε ένα σύντομο χρονικό διάστημα, αφού μια συνεδρία χρήσης της εφαρμογής ολοκληρώθηκε, είναι εφικτό να διαχωριστούν οι περιπτώσεις στις οποίες οι χρήστες δεν κατάφεραν να ανακτήσουν την επιθυμητή επαφή (σε όλες τις άλλες περιπτώσεις θεωρήθηκε ότι η εφαρμογή δε χρησιμοποιήθηκε καθόλου). Αποκλείοντας όλες τις άλλες κλήσεις από την ανάλυση, προέκυψε μια εκτίμηση της πραγματικής απόδοσης του αλγορίθμου πρόβλεψης (περίπου 78% κατά μέσο όρο), η οποία είναι σημαντικά υψηλότερη από εκείνη που προέκυψε όταν υπολογίστηκαν μόνο οι συνεδρίες της εφαρμογής που κατέληξαν σε κλήση. Η απόδοση αυτή είναι πολύ κοντά σε εκείνη που προέκυψε όταν εφαρμόστηκε ο αλγόριθμος μέσω προσομοίωσης στα διαθέσιμα σύνολα δεδομένων, επιβεβαιώνοντας τόσο την ακρίβεια όσο και την αποτελεσματικότητα του αλγορίθμου στο πεδίο, καθώς και την εγκυρότητα των αποτελεσμάτων της προσομοίωσης. Επομένως, στις περισσότερες περιπτώσεις, η ζητούμενη επαφή είναι εφικτό να βρεθεί στην πρώτη οθόνη μιας προσαρμοζόμενης διεπαφής επαφών. Τα πειραματικά αποτελέσματα επιβεβαίωσαν, επίσης, τα προηγούμενα ευρήματα σχετικά με το πώς η «κοινωνικότητα» των χρηστών σε επίπεδο επικοινωνιακής συμπεριφοράς επηρεάζει την επιτυχία των προβλέψεων. Διαπιστώθηκε και σε αυτή την περίπτωση, ότι η συμπεριφορά χρηστών με πιο πολύπλοκα πρότυπα επικοινωνίας είναι πιο δύσκολο να προβλεφθεί, μια υπόθεση που όπως αναφέρθηκε και νωρίτερα, είχε προταθεί και από τους Lee, Seo και Lee (2010).

Μια ενδιαφέρουσα παρατήρηση ήταν πως υπήρχαν μερικές συνεδρίες χρήσης της εφαρμογής κοντά στις οποίες δεν υπήρχαν καθόλου εξερχόμενες κλήσεις, γεγονός το οποίο παρατηρήθηκε και κατά τη δοκιμή στο πεδίο του πρωτοτύπου που παρουσιάστηκε στο (Jung, Anttila & Blom, 2008). Φαίνεται ότι, τότε τότε, οι συμμετέχοντες συνήθιζαν να ανοίγουν την εφαρμογή απλά για να ελέγξουν τις προτάσεις και να πάρουν μια ιδέα για τις επαφές που ήταν πιο πιθανό να καλέσουν. Επιπλέον, για το 40% των κλήσεων που πραγματοποιήθηκαν εκτός της εφαρμογής, δεν μπορεί να είναι κανείς σίγουρος ποια άλλη μέθοδο της συσκευής

χρησιμοποιήθηκε για την έναρξή τους (π.χ. ειδοποιήσεις, λίστα μηνυμάτων). Εντούτοις, φαίνεται ότι κάτω από διαφορετικά πλαίσια, οι κινητοί χρήστες υιοθετούν διαφορετικά μέσα ανάκτησης της ζητούμενης επαφής για την έναρξη εξερχόμενης επικοινωνίας. Για παράδειγμα, όταν ένας χρήστης βρίσκει μια ειδοποίηση για μια αναπάντητη κλήση, είναι πολύ πιθανό να καλέσει πίσω αυτή την επαφή κάνοντας κατευθείαν κλικ πάνω στην ειδοποίηση. Ακόμα μια τέτοια περίπτωση είναι όταν μια εξερχόμενη κλήση γίνεται μετά τη λήψη ενός μηνύματος ή όταν ο χρήστης θυμάται να έχει μιλήσει με μια επαφή πρόσφατα και την αναζητά στη λίστα πρόσφατων κλήσεων. Ειδικά για την τελευταία περίπτωση, ο Whittaker (2011) έχει σημειώσει ότι αυτού του είδους η προηγούμενη γνώση μπορεί να κάνει την ανάκτηση της προσωπικής πληροφορίας πιο εύκολη. Προς το παρόν, δεν υπάρχει βιβλιογραφία που να εξετάζει τη συμπεριφορά των χρηστών και τις μεθόδους επικοινωνίας που χρησιμοποιούν σε διαφορετικά πλαίσια. Επιπλέον έρευνα προς αυτή την κατεύθυνση θα μπορούσε να αποκαλύψει και άλλες διαστάσεις πλαισίου που θα μπορούσαν να βελτιώσουν τα αποτελέσματα των προβλέψεων.

Η ποιοτική ανάλυση των απαντήσεων των συμμετεχόντων στα ερωτηματολόγια επιβεβαίωσε την επιτυχία του αλγορίθμου πρόβλεψης, καθώς οι χρήστες δήλωσαν ότι προτιμούν την εφαρμογή σε σχέση με όλους τους παραδοσιακούς τρόπους ανάκτησης, ενώ αντίστοιχα ήταν και τα ελεύθερα σχόλια που κατέθεσαν. Μια ενδιαφέρουσα παρατήρηση είναι πως οι συμμετέχοντες δήλωσαν ότι χρησιμοποιούσαν κυρίως τη λίστα πρόσφατων κλήσεων για την ανάκτηση επαφών, η οποία όμως, όπως έχει αποδειχθεί από τις πειραματικές προσομοιώσεις, έχει τη χειρότερη απόδοση σε σχέση με τη λίστα συχνών επαφών και τις προσαρμοζόμενες διεπαφές που παρουσιάστηκαν και βασίζονται σε αλγορίθμους πρόβλεψης.

Όπως έχει ήδη σημειωθεί η εφαρμογή Calchas είναι διαθέσιμη στο Google Play για χρήστες συσκευών Android. Δύο μήνες μετά τη διανομή της είχε περίπου 1500 λήψεις, με περίπου 300 χρήστες να τη χρησιμοποιούν σε μόνιμη βάση, ενώ μετά από 1,5 χρόνο περίπου, όπου γίνονταν μόνο αναβαθμίσεις για διορθώσεις σφαλμάτων, είχαν πραγματοποιηθεί 1900 λήψεις και καταγράφησαν περίπου 160 ενεργές εγκαταστάσεις. Αυτοί οι χρήστες θα μπορούσαν να αποτελέσουν μια σημαντική βάση για μελλοντική έρευνα.

4.6 Σύνοψη

Σε αυτό το κεφάλαιο παρουσιάστηκε ένας αλγόριθμος που χρησιμοποιεί το κοινωνικό πλαίσιο που εξάγεται από το ιστορικό κλήσεων των συσκευών κινητού υπολογισμού και διευκολύνει την ανάκτηση των επαφών μέσω της πρόβλεψης εξερχόμενων

κλήσεων. Ο αλγόριθμος δεν περιορίζεται μόνο στην περίπτωση των επαφών, αλλά μπορεί να προσαρμοστεί και για την περίπτωση άλλων στοιχείων προσωπικής πληροφορίας. Οι διαστάσεις πλαισίου που ενσωματώνει είναι αυτές που προέκυψαν ως σημαντικές από την ανάλυση των προτύπων επικοινωνίας των χρηστών που παρουσιάστηκε στο προηγούμενο κεφάλαιο, δηλαδή η συχνότητα και η προσφατότητα της επικοινωνίας, μπορούν όμως με τον ίδιο τρόπο να χρησιμοποιηθούν και άλλες διαστάσεις που ίσως προκύψουν μετά από μελλοντικές αναλύσεις.

Οι πειραματικές προσομοιώσεις που πραγματοποιήθηκαν σε ένα προκαταρκτικό σύνολο δεδομένων και στο σύνολο δεδομένων της Nokia, έδειξαν ότι ο συνδυασμός των δύο διαστάσεων οδηγεί σε πολύ καλύτερα αποτελέσματα από τη χρήση κάθε διάστασης ξεχωριστά, επιλογές που αντιστοιχούν στις παραδοσιακές μεθόδους ανάκτησης (λίστα πρόσφατων κλήσεων και λίστα συχνών επαφών) και επομένως οι χρήστες δεν εξυπηρετούνται από βέλτιστες διεπαφές. Επιπλέον, η σύγκριση με τα αποτελέσματα άλλων προσεγγίσεων που εφαρμόστηκαν σε διαφορετικά σύνολα δεδομένων, έδειξε ότι η προτεινόμενη προσέγγιση, αν και ιδιαίτερα απλή, υπερτερεί ως προς τα ποσοστά επιτυχίας των προβλέψεων.

Επιπροσθέτως, μελετήθηκε ο ρόλος που παίζει η επικοινωνιακή συμπεριφορά των χρηστών στην επιτυχία των προβλέψεων. Συγκεκριμένα, προτάθηκε μια μετρική για την κατάταξη των χρηστών σε ομάδες ανάλογα με την «κοινωνικότητά» τους. Όπως αναμενόταν, όσο περισσότερο «κοινωνικός» είναι ένας χρήστης, τόσο πιο δύσκολο είναι να προβλεφθεί η συμπεριφορά του. Μάλιστα, για την ομάδα των περισσότερο «κοινωνικών» χρηστών φαίνεται να υπάρχει περιθώριο για βελτίωση των αποτελεσμάτων, ενδεχομένως με την προσθήκη επιπλέον διαστάσεων πλαισίου.

Τέλος, οργανώθηκε και εκτελέστηκε ένα πείραμα πεδίου, που περιελάμβανε μεγαλύτερο δείγμα χρηστών από το μοναδικό αντίστοιχο που αναφέρεται στη βιβλιογραφία, για την αξιολόγηση του αλγορίθμου σε πραγματικές συνθήκες. Το πείραμα στήριξε τα συμπεράσματα από τις προσομοιώσεις, ενώ τόσο τα ποσοτικά, όσο και τα ποιοτικά δεδομένα επιβεβαίωσαν την αποτελεσματικότητα του αλγορίθμου πρόβλεψης.

5. Οι Επιπτώσεις της Τοποθεσίας στη Διαχείριση Προσωπικής Πληροφορίας Συσκευών Κινητού Υπολογισμού

Στην επισκόπηση της σχετικής ερευνητικής δραστηριότητας παρουσιάστηκε αναλυτικά η αντίληψη που επικρατεί στην ερευνητική κοινότητα για τη σημασία της τοποθεσίας ως παράγοντα πλαισίου για την πρόβλεψη των κλήσεων και κατά συνέπεια για την αξιοποίηση των επαφών, αλλά και γενικότερα των στοιχείων προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού. Ταυτόχρονα, δόθηκε έμφαση στο γεγονός πως δεν έχει επικυρωθεί πειραματικά η επίπτωση που έχει αυτή η διάσταση στο υπό μελέτη πρόβλημα. Εντούτοις, σε πολλές από τις προτεινόμενες προσεγγίσεις η τοποθεσία ενσωματώνεται στο μοντέλο πρόβλεψης, χωρίς περαιτέρω διερεύνηση. Σε αυτό το κεφάλαιο μελετάται η επίπτωση της τοποθεσίας, είτε έμμεσα (μέσω του χρονικού πλαισίου) είτε άμεσα (χρησιμοποιώντας τη φυσική και τη σημασιολογική τοποθεσία) και παρουσιάζονται τα πειραματικά ευρήματα.

5.1 Προκαταρκτική διερεύνηση του χρονικού πλαισίου

Σύμφωνα με τη βιβλιογραφία, το χρονικό πλαίσιο μπορεί να χρησιμοποιηθεί για τον εντοπισμό μοτίβων κίνησης (Bicocchi, Castelli, Mamei, Rosi & Zambonelli, 2008) και την εξαγωγή σημασιολογικής πληροφορίας για την τοποθεσία του χρήστη (Eagle, Clauset & Quinn, 2009). Πιο συγκεκριμένα, καθώς οι άνθρωποι τείνουν να επισκέπτονται διάφορα μέρη σε συγκεκριμένες ώρες της ημέρας (π.χ. το σπίτι κατά τη διάρκεια της νύχτας, το γραφείο τις πρωινές ώρες, εστιατόρια την ώρα του μεσημεριανού γεύματος κ.λπ.) (Montoliu, Martínez-Uso, Martínez-Sotoca & McInerney, 2012), η σημασία των τοποθεσιών μπορεί να ποικίλει ανάλογα με την ώρα της ημέρας (Cao, Cong & Jensen, 2010).

Με βάση τις παραπάνω παρατηρήσεις, ο αλγόριθμος που προτάθηκε στο προηγούμενο κεφάλαιο επεκτάθηκε ώστε να ενσωματώνει μια τρίτη διάσταση πλαισίου: το χρονικό πλαίσιο. Στο (Lee, Seo & Lee, 2010) και στο (Phithakkitnukoon & Dantu, 2011) το χρονικό πλαίσιο αποκτήθηκε είτε σε ωριαία βάση είτε χωρίζοντας την ημέρα σε τέσσερα άνισα μέρη, ανάλογα με τους ρυθμούς ζωής των ανθρώπων. Επομένως, δε φαίνεται να υπάρχει μια κοινή προσέγγιση σχετικά με τη βέλτιστη αναπαράσταση αυτού του είδους του πλαισίου. Στην προσέγγιση που ακολουθήθηκε στα πλαίσια της διδακτορικής έρευνας, η ημέρα χωρίστηκε σε τρία λογικά τμήματα που θεωρήθηκε ότι αποτυπώνουν καλύτερα τα γενικά μοτίβα που ακολουθούν οι

χρήστες κατά τη διάρκειά της: πρωί (08:00-16:00), απόγευμα (16:00-24:00) και νύχτα (00:00-08:00). Για τον υπολογισμό του σκορ της χρονικής διάστασης για μια επαφή i τη χρονική στιγμή dt , αθροίζονται όλες οι επικοινωνίες που πραγματοποιήθηκαν με αυτή την επαφή στο τμήμα της ημέρας που αντιστοιχεί στη χρονική στιγμή dt κατά τη διάρκεια του παραθύρου εκπαίδευσης και το πλήθος αυτό διαιρείται με το συνολικό πλήθος των επικοινωνιών που πραγματοποιήθηκαν σε αυτό το τμήμα της ημέρας κατά τη διάρκεια του παραθύρου εκπαίδευσης. Η συνάρτηση σκορ σε αυτή την περίπτωση είναι η

$$FToD_t: C \rightarrow [0,1]$$

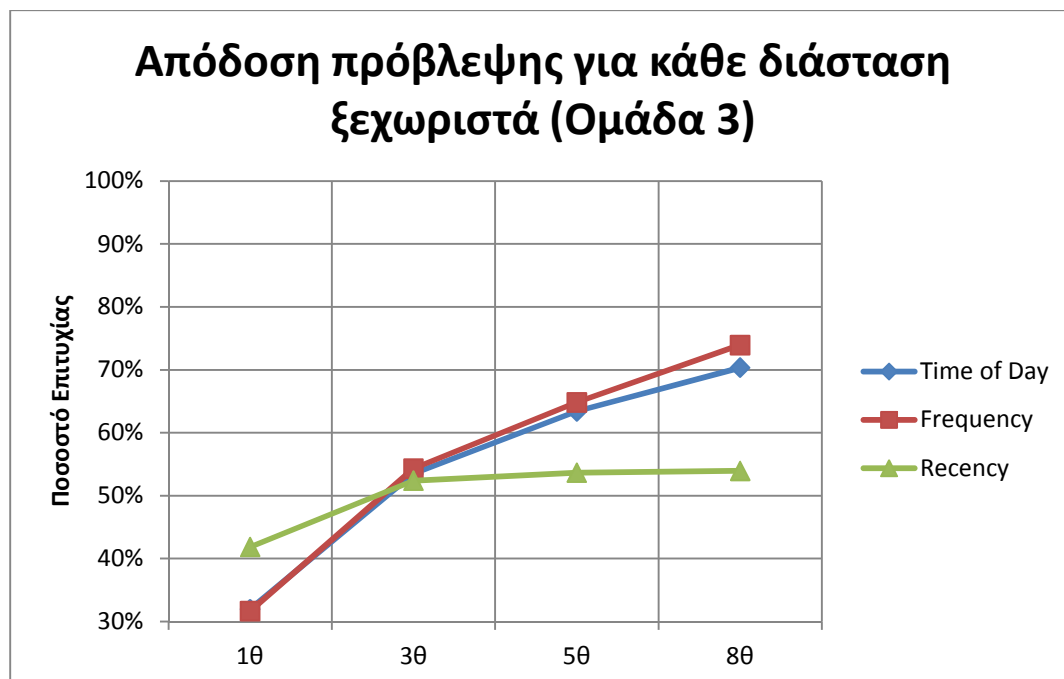
$$FToD_t(c_k) = \frac{\sum \text{κλήσεων στο αντίστοιχο τμήμα της ημέρας με τη } c_k \text{ στο } t_w}{\sum \text{κλήσεων στο αντίστοιχο τμήμα της ημέρας στο } t_w}$$

Όταν η διάσταση του χρονικού πλαισίου ενσωματώνεται στο μοντέλο αναπαράστασης του πλαισίου, το διάνυσμα που αντιπροσωπεύει το πλαίσιο της επαφής c_k γίνεται

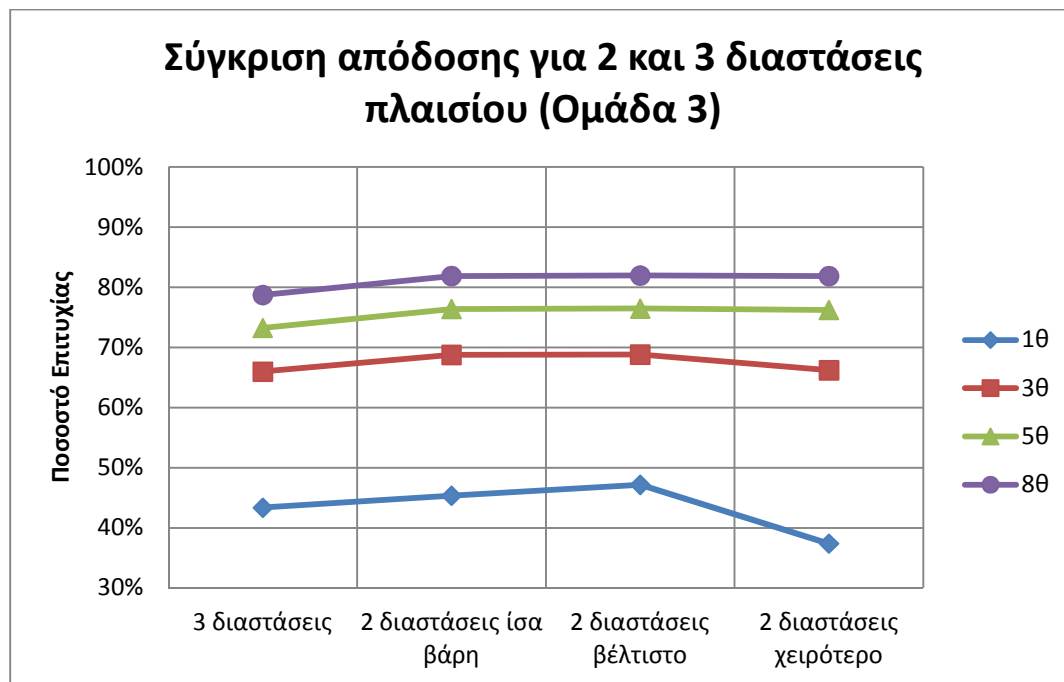
$$c_k = (F_t(c_k), R_t(c_k), FToD_t(c_k))$$

ενώ ο αλγόριθμος πρόβλεψης υπολογίζει το συνολικό σκορ κάθε επαφής ως

$$\Pi_t(c_k) = w_f F_t(c_k) + w_r \times R_t(c_k) + w_{ftod} FToD_t(c_k)$$



Γράφημα 45. Απόδοση πρόβλεψης για κάθε διάσταση ξεχωριστά για την ομάδα με τους χρήστες μεγαλύτερης κοινωνικότητας



Γράφημα 46. Σύγκριση της απόδοσης για τις τρεις συνδυασμένες διαστάσεις πλαisiού (συχνότητα, προσφατότητα, ώρα της ημέρας) για την ομάδα με τους χρήστες μεγαλύτερης κοινωνικότητας

Όπως φαίνεται στο γράφημα 45, η διάσταση του χρονικού πλαisiού (λογικό τμήμα της ημέρας) όταν χρησιμοποιείται μόνη της στο μοντέλο πρόβλεψης παρέχει αποτελέσματα συγκρίσιμα με αυτά που προκύπτουν όταν χρησιμοποιείται μόνο η συχνότητα επικοινωνίας. Το γράφημα 46 παρουσιάζει την απόδοση του αλγορίθμου όταν προστίθεται η διάσταση της ώρας της ημέρας για την Ομάδα 3 (περισσότερο κοινωνικοί χρήστες), η οποία έχει και το μεγαλύτερο περιθώριο βελτίωσης της πρόβλεψης. Γι' αυτά τα αποτελέσματα, πραγματοποιήθηκαν διάφοροι συνδυασμοί τιμών για τα βάρη των διαστάσεων, χωρίς να βρεθούν σημαντικές διαφοροποιήσεις, ενώ το παράθυρο εκπαίδευσης ήταν και πάλι 15 ημέρες. Μπορεί κανείς να διακρίνει πως η προσθήκη της τρίτης διάστασης οδηγεί σε ελαφριά μείωση της μέσης απόδοσης, παρά τις προσδοκίες πως θα μπορούσε να βελτιώσει το ποσοστό επιτυχούς πρόβλεψης.

5.2 Φυσική vs σημασιολογική τοποθεσία

Η φυσική τοποθεσία αναφέρεται σε μια απόλυτη γεωγραφική τοποθεσία στην επιφάνεια της γης, η οποία εκφράζεται συνήθως με ένα ζεύγος γεωγραφικών συντεταγμένων. Μια διεύθυνση όπως «Πανεπιστημίου 40, Αθήνα» μπορεί να εξαχθεί από συντεταγμένες τοποθεσίας και από αυτή κανείς είναι σε θέση να κατανοήσει σε ποια χώρα, πόλη και συγκεκριμένη γεωγραφική περιοχή βρίσκεται η τοποθεσία. Από

την άλλη, η σημασιολογική τοποθεσία υπονοεί την επισύναψη μιας σημασιολογικής ετικέτας σε μια τοποθεσία, όπως «σπίτι», «δουλειά», «πανεπιστήμιο» κ.λπ.

Η διάσταση της φυσικής τοποθεσίας θα μπορούσε ενδεχομένως να αποδειχθεί χρήσιμη σε σενάρια όπως τα παρακάτω:

«Ο Γιώργος μεγάλωσε σε μια μικρή πόλη της Δυτικής Ελλάδας. Μετά την αποφοίτησή του από το λύκειο, μετακόμισε στην Αθήνα καθώς εκεί βρήκε μια καλή ευκαιρία επαγγελματικής αποκατάστασης. Από καιρού εις καιρόν επισκέπτεται τη γενέτειρά του όπου ζουν οι γονείς του. Όταν βρίσκεται εκεί, συνήθως καλεί τους παιδικούς του φίλους και κανονίζουν να συναντηθούν στην αγαπημένη τους μπυραρία. Παρότι δεν επικοινωνούν συχνά όταν βρίσκεται στην Αθήνα, αισθάνονται ακόμα κοντά μεταξύ τους.»

«Η Μαρία εργάζεται σε μια πολυεθνική εταιρεία πληροφορικής που εδρεύει στο Λονδίνο. Εντούτοις, συχνά ταξιδεύει στο εξωτερικό, στα γραφεία της εταιρείας στη Μαδρίτη. Όταν βρίσκεται στη Μαδρίτη, καλεί συχνά τους τοπικούς συναδέλφους της από το τηλέφωνό της, ενώ όταν βρίσκεται στο Λονδίνο επικοινωνεί μαζί τους χρησιμοποιώντας το Skype ή το ηλεκτρονικό ταχυδρομείο.»

Μια παρόμοια κατάσταση περιγράφεται στο (Barzaiq & Loke, 2011), όπου η τοποθεσία προτείνεται ως μια χρήσιμη διάσταση πλαισίου για την πρόβλεψη των εξερχόμενων κλήσεων, όταν πρόκειται για χρήστες που περνούν περιόδους της ζωής τους μετακινούμενοι ανάμεσα σε διαφορετικές χώρες. Αυτά τα σενάρια δίνουν έμφαση στο γεγονός πως η επικοινωνιακή συμπεριφορά μπορεί να αλλάξει όταν οι χρήστες μετακινούνται σε μια διαφορετική γεωγραφική τοποθεσία σε επίπεδο πόλης, στην ίδια ή διαφορετική χώρα.

Αντίστοιχα, η σημασιολογική τοποθεσία σχετίζεται περισσότερο με τις δραστηριότητες που πραγματοποιεί ο χρήστης και οι οποίες συνήθως εξαρτώνται από τη σημασιολογική ετικέτα που έχει αποδοθεί σε αυτή. Για παράδειγμα, μπορεί κανείς να υποθέσει πως τις πρωινές ώρες που περνάει κάποιος στο γραφείο, επικοινωνεί κυρίως με πελάτες ή συναδέλφους για επαγγελματικά ζητήματα, ενώ τα απογεύματα στο σπίτι επικοινωνεί κυρίως με φίλους για να κανονίσουν κάποια συνάντηση για διασκέδαση.

5.3 Μελέτη της επίπτωσης της φυσικής τοποθεσίας

Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα της μελέτης της επίπτωσης της φυσικής τοποθεσίας (Plessas, Georgiadou, Stefanis, Komninos & Garofalakis,

2015), σε επίπεδο πόλης, στην αξιοποίηση των επαφών για την έναρξη εξερχόμενης επικοινωνίας.

5.3.1 Σύνολο δεδομένων

Για τη διερεύνηση της επίπτωσης της φυσικής τοποθεσίας στην πρόβλεψη εξερχόμενων κλήσεων χρησιμοποιήθηκε και πάλι το Nokia Mobile Data Challenge Dataset (Laurila, Gatica-Perez, Aad, Bornet, Do, Dousse, ... & Miettinen, 2012), που όπως έχει ήδη αναφερθεί είναι το πληρέστερο δημόσια διαθέσιμο σύνολο δεδομένων χρήσης συσκευών κινητού υπολογισμού και ειδικά όσον αφορά την καταγραφή στοιχείων τοποθεσίας (Kiukkonen, Blom, Dousse, Gatica-Perez & Laurila, 2010), καθώς έχουν καταγραφεί συνολικά περισσότερα από 26.000.000 σημεία τοποθεσίας. Οι πηγές καταγραφής περιλαμβάνουν ίχνη GPS (εικόνα 28) που παρέχουν γεωγραφικές συντεταγμένες, σημεία πρόσβασης σε ασύρματα δίκτυα των οποίων η τοποθεσία εκτιμήθηκε από τα διαθέσιμα GPS δεδομένα (εικόνα 29) και κεραίες GSM δικτύων (εικόνα 30). Εντούτοις, για την προστασία της ιδιωτικότητας των συμμετεχόντων η ακρίβεια των GPS δεδομένων έχει μειωθεί σκόπιμα. Για την εύρεση της τοποθεσίας των GSM κεραιών χρησιμοποιήθηκε ένα ανεπίσημο API⁴ της Google, καθώς δεν περιλαμβάνεται στο σύνολο δεδομένων, όμως η εύρεση της τοποθεσίας με βάση το αναγνωριστικό της κεραίας (Cell-ID) είναι μια μέθοδος χαμηλής ακρίβειας (Eagle & Pentland, 2006; Smit, Stander & Ophoff, 2012) σε σχέση με τις άλλες διαθέσιμες μεθόδους. Συνεπώς, για τους παραπάνω λόγους δεν είχε νόημα η μελέτη της φυσικής τοποθεσίας σε επίπεδο μικρότερο από αυτό της πόλης.

	db_key [PK] integer	time integer	vertical_accuracy numeric	horizontal_accuracy numeric	altitude numeric	longitude numeric(5,2)	latitude numeric(5,2)
1	29632357	1234365498	22.5	193.5503692629999	387.5	6.56	46.52
2	29632410	1234366831	19.5	293.3594360349999	389.5	6.56	46.52
3	29664346	1234370586	111.0	43.78769302370000	455.0	6.56	46.52
4	29664349	1234370596	47.0	40.57267379759999	459.0	6.56	46.52

Εικόνα 28. Παράδειγμα εγγραφών από τα δεδομένα τοποθεσίας GPS

	db_key [PK] integer	latitude numeric(5,2)	longitude numeric(5,2)	mac character varying(75)
1	44508325	46.52	6.57	0017f2-7a2d81acf76709a953286b90a4cfd810aeb954e4efd4069e5
2	44888433	46.52	6.57	0017f2-7a2d81acf76709a953286b90a4cfd810aeb954e4efd4069e5
3	44888448	46.52	6.57	0017f2-7a2d81acf76709a953286b90a4cfd810aeb954e4efd4069e5
4	44933564	46.52	6.57	0017f2-7a2d81acf76709a953286b90a4cfd810aeb954e4efd4069e5

Εικόνα 29. Παράδειγμα εγγραφών από τα δεδομένα τοποθεσίας WLAN

⁴ <http://www.google.com/glm/mmap>

	db_key [PK] integer	cellid integer	signaldbm integer	signal integer
1	29502513	9449	100	1
2	29511134	9449	87	6
3	29511138	9449	101	1
4	29511141	4700	94	3

	id [PK] integer	cell_id integer	network_cod integer	area_code integer	country_code integer
1	560	31043	12	1920	244
2	561	31038	12	1920	244
3	562	102704	1	1	228

Εικόνα 30. Παράδειγμα εγγραφών από τα δεδομένα τοποθεσίας GSM (αριστερά) και εγγραφών με λεπτομέρειες που αφορούν τις GSM κεραίες (δεξιά)

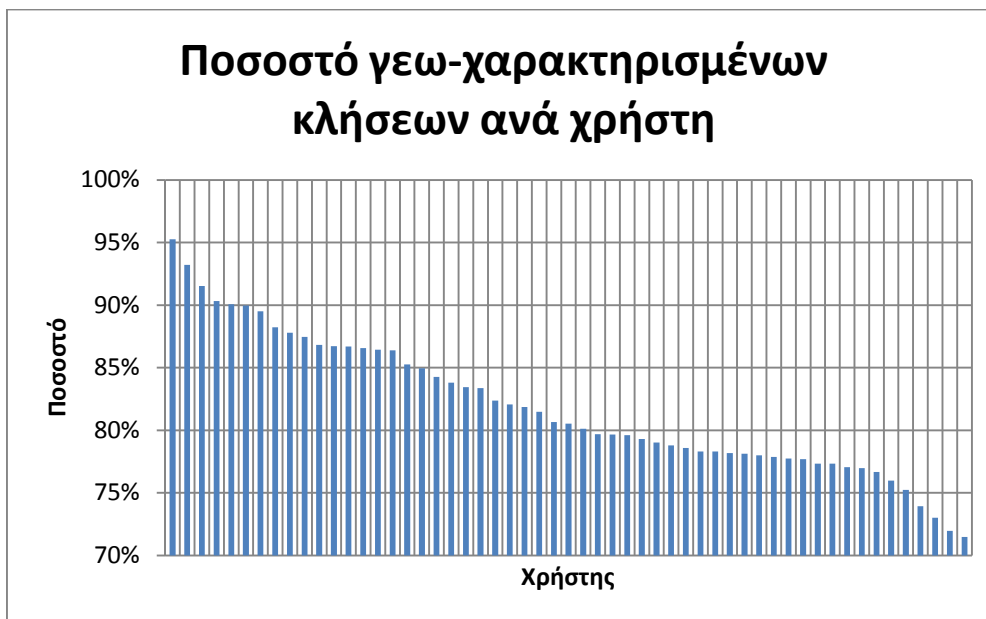
Επιπλέον, καθώς οι χρήστες μπορούσαν να απενεργοποιήσουν το GPS ή το WiFi της συσκευής τους κατά τη διάρκεια της συμμετοχής τους στο πείραμα, πολλά από τα διαθέσιμα καταγεγραμμένα γεγονότα του συνόλου δεδομένων δεν είναι γεωγραφικά χαρακτηρισμένα. Για να ξεπεραστεί αυτό το εμπόδιο και καθώς έπρεπε σε κάθε κλήση να αντιστοιχηθεί μια τοποθεσία (πληροφορία που δεν ήταν άμεσα διαθέσιμη, αφού οι χρονικές στιγμές καταγραφής της τοποθεσίας δε συμπίπτουν απαραίτητα με τις χρονικές στιγμές πραγματοποίησης κλήσεων), αποφασίστηκε να θεωρηθεί ως αντιπροσωπευτική τοποθεσία κάθε κλήσης η κοντινότερη διαθέσιμη εγγραφή τοποθεσίας εντός εύρους x λεπτών πριν ή μετά τη χρονική στιγμή που πραγματοποιήθηκε η κλήση. Για το σκοπό αυτό, υπολογίστηκε το συνολικό ποσοστό κλήσεων στις οποίες ανατέθηκε τοποθεσία για διαφορετικές τιμές του εύρους x (πίνακας 22). Όταν το εύρος ισούται με 15 λεπτά, το ποσοστό αυτό είναι μεγαλύτερο από 70%, ένα φαινομενικά ικανοποιητικό κατώφλι. Όπως μπορεί να παρατηρήσει κανείς, η επέκταση του εύρους κατά 15 λεπτά έχει ως αποτέλεσμα μια αύξηση μόνο κατά 1% των γεω-χαρακτηρισμένων κλήσεων, ενώ αναμφισβήτητα η αβεβαιότητα για την πραγματική τοποθεσία του χρήστη τη χρονική στιγμή της κλήσης αυξάνεται. Συνεπώς, υιοθετήθηκε η τιμή των 15 λεπτών για το εύρος, καθώς πρόκειται για μια σχετικά σύντομη περίοδο ώστε να θεωρήσει κανείς πως η τοποθεσία του χρήστη δεν έχει πρακτικά αλλάξει, τουλάχιστον σε επίπεδο πόλης. Προφανώς, για κάποιες κλήσεις δεν ήταν εφικτό να αντιστοιχηθεί τοποθεσία, λόγω της έλλειψης γεωγραφικών εγγραφών μέσα σε αυτό το χρονικό εύρος. Ένα μικρότερο εύρος θα είχε ως αποτέλεσμα ακόμα λιγότερες κλήσεις χωρίς αντιστοίχιση τοποθεσίας, ενώ μεγαλύτερες τιμές θα οδηγούσαν σε μεγαλύτερο βαθμό αβεβαιότητας όσον αφορά την ακρίβεια της τοποθεσίας, χωρίς σημαντικό κέρδος σε σχέση με το πλήθος των γεω-χαρακτηρισμένων κλήσεων.

Πίνακας 22. Συνολικό ποσοστό γεω-χαρακτηρισμένων κλήσεων για διάφορες τιμές του εύρους

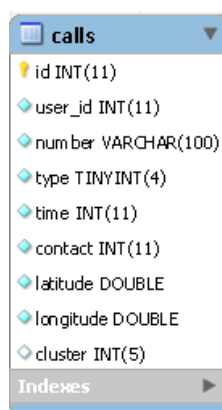
Εύρος x (λεπτά)	Συνολικό % κλήσεων με τοποθεσία
5	69,4%
10	69,8%
15	71.3%

30	72.4%
45	73.4%
60	74.5%

Μετά από αυτή τη διαδικασία, παρατηρήθηκε ότι υπήρχαν σημαντικές διαφοροποιήσεις στο πλήθος των κλήσεων που αντιστοιχήθηκαν σε τοποθεσία ανάμεσα στους διάφορους χρήστες. Υπήρχαν χρήστες με ποσοστό ταυτοποίησης της τοποθεσίας για τις κλήσεις τους μεγαλύτερο από 90%, αλλά επίσης και χρήστες (που πιθανότατα απενεργοποιούσαν συχνά το λογισμικό καταγραφής) για τους οποίους μόνο το 20% των κλήσεων κατέστη δυνατό να χαρακτηριστούν με γεωγραφικές συντεταγμένες. Ως εκ τούτου, αποφασίστηκε να απορριφθούν από περαιτέρω ανάλυση χρήστες με ποσοστό γεω-χαρακτηρισμένων κλήσεων μικρότερο από 70%, καθώς η διατήρηση χρηστών με ποσοστό μικρότερο από αυτό το κατώφλι θα μπορούσε να επηρεάσει αρνητικά την αξιοπιστία των αποτελεσμάτων πρόβλεψης. Επιπροσθέτως, αποκλείστηκαν χρήστες που δεν μπορούσαν να θεωρηθούν έγκυροι για ανάλυση του αρχείου κλήσεων, καθώς κάποιοι είχαν αρχείο κλήσεων που κάλυπτε ένα πολύ σύντομο χρονικό διάστημα (μικρότερο από 30 ημέρες), ενώ άλλοι είχαν πραγματοποιήσει πολύ λίγες κλήσεις (μικρότερες από 100) ή χαρακτηρίζονταν από πολύ αραιή χρήση της συσκευής για την πραγματοποίηση κλήσεων (λιγότερες από 2,25 κλήσεις ανά ημέρα). Πρόκειται για μια διαδικασία που πραγματοποιήθηκε και κατά την ανάλυση που παρουσιάστηκε στο προηγούμενο κεφάλαιο, με τις ίδιες τιμές αποκλεισμού χρηστών για τα παραπάνω χαρακτηριστικά. Αυτά τα ακραία χαρακτηριστικά μπορούν να εξηγηθούν, καθώς υπήρχαν χρήστες που εγκατέλειψαν το πείραμα νωρίς ή άλλοι που αποφάσιζαν συχνά να απενεργοποιήσουν το λογισμικό καταγραφής, σύμφωνα με την τεκμηρίωση του συνόλου δεδομένων. Εν τέλει, έμειναν 54 χρήστες, ένας αριθμός που δεν είναι πολύ μακριά από τον αριθμό των 80 χρηστών που ξεχώρισαν ως αυτοί με την καλύτερη ποιότητα δεδομένων τοποθεσίας για το διαγωνισμό της NOKIA (Laurila, Gatica-Perez, Aad, Bornet, Do, Dousse, ... & Miettinen, 2012), λαμβάνοντας υπόψη ότι στην περίπτωση που παρουσιάζεται σε αυτή την έρευνα χρησιμοποιήθηκε και η ποιότητα των χαρακτηριστικών επικοινωνίας ως παράγοντας αποκλεισμού. Το μέσο ποσοστό κλήσεων που χαρακτηρίστηκαν με συντεταγμένες τοποθεσίας ήταν 82,05% (sd=5,42%), ενώ στο γράφημα 47 παρουσιάζεται η κατανομή των γεω-χαρακτηρισμένων κλήσεων για όλους τους χρήστες.



Γράφημα 47. Ποσοστό γεω-χαρακτηρισμένων κλήσεων ανά χρήστη σε φθίνουσα κατάταξη



Εικόνα 31. Προσαρμογή του πίνακα της ΒΔ για την αποθήκευση των πληροφοριών τοποθεσίας για κάθε κλήση.

Για την αποθήκευση της επιπλέον πληροφορίας (συντεταγμένες, συστάδα που προέκυψε από την ανάλυση της επόμενης παραγράφου) για κάθε κλήση, ο πίνακας της ΒΔ που χρησιμοποιούνταν για την αποθήκευση του ιστορικού κλήσεων για το σύνολο δεδομένων της Nokia επεκτάθηκε όπως φαίνεται στην εικόνα 31.

5.3.2 Συσταδοποίηση (clustering) δεδομένων τοποθεσίας

Για την πραγματοποίηση της ανάλυσης, ήταν απαραίτητη η συσταδοποίηση όλων των γεω-χαρακτηρισμένων κλήσεων κάθε χρήστη σε επίπεδο πόλης. Στη βιβλιογραφία υπάρχουν διαθέσιμοι διάφοροι αλγόριθμοι γεωγραφικής συσταδοποίησης, ενώ ένας ευρέως χρησιμοποιούμενος είναι ο DBSCAN (Ester, Kriegel, Sander & Xu, 1996). Ο αλγόριθμος DBSCAN, για το σχηματισμό των συστάδων (clusters) βασίζεται στην έννοια της πυκνότητας των σημείων, η οποία

είναι μεγαλύτερη μέσα στη συστάδα απ' ότi έξω από αυτή. Ο αλγόριθμος διακρίνει τα σημεία ενός συνόλου D σε αυτά που ανήκουν σε κάποια συστάδα και αυτά που δεν ανήκουν, τα οποία αποτελούν «θόρυβο». Η βασική ιδέα είναι πως για κάθε σημείο μιας συστάδας, η γειτονιά του μιας δοσμένης ακτίνας πρέπει να περιέχει τουλάχιστον ένα ελάχιστο πλήθος σημείων, δηλαδή η πυκνότητα της γειτονιάς πρέπει να υπερβαίνει κάποιο κατώφλι. Το σχήμα της γειτονιάς καθορίζεται από την επιλογή μιας συνάρτησης υπολογισμού της απόστασης ανάμεσα σε δύο σημεία p και q , που συμβολίζεται ως $\text{dist}(p,q)$. Για παράδειγμα, αν χρησιμοποιηθεί η απόσταση Manhattan στον χώρο των δύο διαστάσεων, το σχήμα της γειτονιάς είναι ορθογώνιο. Ο αλγόριθμος μπορεί να λειτουργήσει με οποιαδήποτε συνάρτηση απόστασης είναι κατάλληλη για το υπό εξέταση πρόβλημα και στη συνέχεια θεωρείται πως αυτή είναι η Ευκλείδεια απόσταση. Στη συνέχεια, ακολουθούν κάποιοι απαραίτητοι ορισμοί και ο αλγόριθμος:

Ορισμός 1: (ε -γειτονιά ενός σημείου) Η ε -γειτονιά ενός σημείου p , που συμβολίζεται ως $N_\varepsilon(p)$, ορίζεται ως $N_\varepsilon(p) = \{q \in D \mid \text{dist}(p,q) \leq \varepsilon\}$

Ορισμός 2: (απευθείας density-reachable) Ένα σημείο p είναι απευθείας density-reachable από ένα σημείο q , δεδομένων των ε και $MinPts$, εάν

1. $p \in N_\varepsilon(q)$
2. $|N_\varepsilon(q)| \geq MinPts$ (προϋπόθεση βασικού σημείου)

Ορισμός 3: (density-reachable) Ένα σημείο p είναι density-reachable από ένα σημείο q , δεδομένων των ε και $MinPts$, εάν υπάρχει μια ακολουθεία σημείων $p_1, \dots, p_n$ με $p_1=p$ και $p_n=q$ όπου κάθε p_{i+1} είναι απευθείας density-reachable από το p_i .

Ορισμός 4: (density-connected) Ένα σημείο p είναι density-connected με ένα σημείο q , δεδομένων των ε και $MinPts$, εάν υπάρχει ένα σημείο o , τέτοιο ώστε τόσο το p όσο και το q να είναι density-reachable από το o .

Ορισμός 5: (συστάδα) Έστω D ένα σύνολο σημείων. Μια συστάδα C , δεδομένων των ε και $MinPts$, είναι ένα μη κενό υποσύνολο του D που ικανοποιεί τις παρακάτω συνθήκες

1. $\forall p,q$: εάν το $p \in C$ και το q είναι density-reachable από το p , τότε το $q \in C$
2. $\forall p,q \in C$: το p είναι density-connected με το q

Ορισμός 6: (θόρυβος) Έστω $C_1, \dots, C_k$ οι συστάδες ενός συνόλου σημείων D , δεδομένων των παραμέτρων ε και $MinPts$. Ως θόρυβος ορίζεται το σύνολο των

σημείων του D που δεν ανήκουν σε κανένα C_i , $i=1\dots k$, δηλαδή θόρυβος = $\{p \in D \mid \forall i: p \notin C_i\}$

Λήμμα 1: Έστω p ένα σημείο του D και $|N_\varepsilon(p)| \geq MinPts$. Τότε το σύνολο $O = \{o \mid o \in D \text{ και } o \text{ είναι density-reachable από το } p\}$ είναι μια συστάδα, δεδομένων των ε και $MinPts$.

Λήμμα 2: Έστω C μια συστάδα, δεδομένων των ε και $MinPts$, και έστω p οποιοδήποτε σημείο στο C με $|N_\varepsilon(p)| \geq MinPts$. Τότε το C είναι ίσο με το σύνολο $O = \{o \mid o \text{ είναι density-reachable από το } p\}$.

Τα παραπάνω λήμματα δηλώνουν πως, δεδομένων των παραμέτρων ε και $MinPts$, ο αλγόριθμος μπορεί να εφαρμοστεί σε δύο βήματα: το πρώτο αφορά την εύρεση ενός αυθαίρετου σημείου που ικανοποιεί την προϋπόθεση βασικού σημείου, ενώ το δεύτερο την ανάκτηση όλων των σημείων που είναι density-reachable από αυτό και αποτελούν τη συστάδα στην οποία ανήκει.

Πίνακας 23. Ο αλγόριθμος DBSCAN

```

DBSCAN (SetOfPoints, Eps, MinPts)
// SetOfPoints is UNCLASSIFIED
ClusterId := nextId(NOISE);
FOR i FROM 1 TO SetOfPoints.size DO
  Point := SetOfPoints.get(i);
  IF Point.CiId = UNCLASSIFIED THEN
    IF ExpandCluster(SetOfPoints, Point, ClusterId, Eps, MinPts) THEN
      ClusterId := nextId(ClusterId);
    END IF;
  END IF;
END FOR;
END; //DBSCAN

ExpandCluster(SetOfPoints, Point, CiId, Eps, MinPts) : Boolean;
seeds := SetOfPoints.regionQuery (Point, Eps)
IF seeds.size < MinPts THEN // no core point
  SetOfPoint.changeCiId(Point, NOISE);
  RETURN False;
ELSE //all points in seeds are density-reachable from Point
  SetOfpoints.changeCiIds(seeds, CiId);
  seeds.delete(Point);
  WHILE seeds <> Empty DO
    currentP := seeds.first();
    result := setofPoints.regionQuery(currentP, Eps)
    IF result.size >= MinPts THEN
      FOR i FROM 1 TO result.size DO

```

```

resultP := result.get(i);
IF resultP.CiId IN {UNCLASSIFIED,NOISE} THEN
  IF resultP.CiId = UNCLASSIFIED THEN
    seeds.append(resultP);
  END IF;
  SetOfPoints.changeCiId(resultP,CiId);
END IF; //UNCLASSIFIED or NOISE
END FOR;
END IF; // result.size >= MinPts
seeds.delete(currentP);
END WHILE; // seeds <> Empty
RETURN True;
END IF;
END; //ExpandCluste

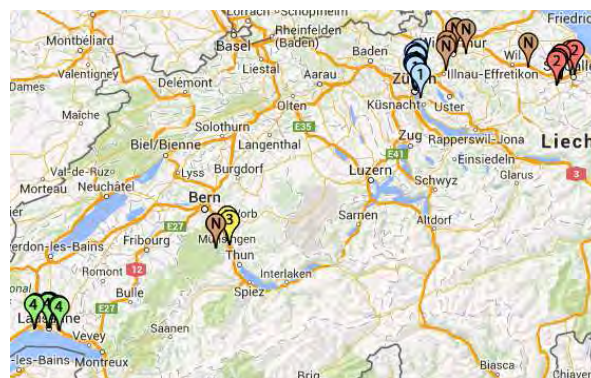
```

Ο αλγόριθμος απαιτεί δύο παραμέτρους: την ελάχιστη απόσταση ε για να θεωρηθούν δύο σημεία απευθείας density-reachable και το ελάχιστο πλήθος σημείων που απαιτούνται για να σχηματιστεί μια συστάδα *MinPts*. Οι κατάλληλες τιμές γι' αυτές τις παραμέτρους εξαρτώνται από το σύνολο δεδομένων και το σκοπό για τον οποίο πραγματοποιείται η συσταδοποίηση και συνήθως ορίζονται με βάση την εμπειρία και την παρατήρηση. Μετά τον πειραματισμό με διάφορες τιμές για τις παραμέτρους και έχοντας υπόψη ότι σκοπός ήταν συσταδοποίηση στο γεωγραφικό επίπεδο της πόλης, ενώ συστάδες με πολύ μικρό αριθμό κλήσεων δεν είχαν νόημα, αποφασίστηκε η χρησιμοποίηση για την παράμετρο ε της τιμής των 5 χιλιομέτρων και η τιμή της παραμέτρου *MinPts* τέθηκε στο 2,5% του συνολικού πλήθους κλήσεων του χρήστη. Υψηλότερες τιμές της παραμέτρου ε είχαν ως αποτέλεσμα συστάδες που κάλυπταν πολύ μεγαλύτερες περιοχές (ένα παράδειγμα παρουσιάζεται στην εικόνα 32 αριστερά, όπου $\varepsilon=10\chi\lambda\mu$), ενώ μικρότερες τιμές είχαν ως αποτέλεσμα το σχηματισμό διακριτών συστάδων εντός της ίδιας πόλης (όπως φαίνεται στην εικόνα 32 δεξιά, όπου $\varepsilon=3\chi\lambda\mu$). Από την άλλη, μεγαλύτερες τιμές για την παράμετρο *MinPts* οδηγούσαν σε μεγαλύτερο ποσοστό κλήσεων-θορύβου (που όμως οπτικά φαίνονταν να σχηματίζουν συστάδες σε συγκεκριμένες περιοχές) και μικρότερες τιμές είχαν ως αποτέλεσμα το σχηματισμό συστάδων που αποτελούνταν από πολύ μικρό αριθμό κλήσεων (συνήθως λιγότερες από 10). Ένα παράδειγμα των συστάδων που σχηματίστηκαν για έναν από τους χρήστες εμφανίζεται στην εικόνα 33. Στην εικόνα 34 παρουσιάζεται ένα άλλο παράδειγμα συσταδοποίησης με μεγάλο ποσοστό δεδομένων θορύβου. Σε αυτή την περίπτωση, μπορεί κανείς να παρατηρήσει πως οι κλήσεις-θόρυβος φαίνεται να είναι είτε απομονωμένες κλήσεις σε μέρη όπου ο χρήστης δεν έχει συχνή παρουσία, είτε κλήσεις ανάμεσα σε δύο συστάδες, που

πραγματοποιήθηκαν πιθανότατα καθώς ο χρήστης μετακινούνταν από τη μια τοποθεσία στην άλλη.



Εικόνα 32. Συσταδοποίηση με $\varepsilon=10$ χλμ (αριστερά) και $\varepsilon=3$ χλμ (δεξιά)



Εικόνα 33. Χάρτης με τις συσταδοποιημένες κλήσεις (4 συστάδες) ενός χρήστη. Πινέζες με το σύμβολο «N» αναπαριστούν θόρυβο



Εικόνα 34. Χάρτης με τις συσταδοποιημένες κλήσεις ενός χρήστη με υψηλό ποσοστό θορύβου

Μετά την εφαρμογή του αλγορίθμου DBSCAN για τη γεωγραφική συσταδοποίηση των κλήσεων κάθε χρήστη, παρατηρήθηκε πως 20 χρήστες είχαν τηλεφωνικές επικοινωνίες ενώ βρίσκονταν μόνο σε μια τοποθεσία (συστάδα). Αυτοί οι χρήστες δεν παρουσιάζουν σημαντική κινητικότητα κατά τη διάρκεια συλλογής των δεδομένων. Οι υπόλοιποι έχουν πραγματοποιήσει κλήσεις που κατανέμονται σε από ένα έως πέντε συστάδες, όπως φαίνεται στον πίνακα 24.

Πίνακας 24. Κατανομή των κλήσεων στις γεωγραφικές συστάδες που σχηματίστηκαν ανά χρήστη

ID Χρήστη	Σ1%	Σ2%	Σ3%	Σ4%	Σ5%	Θόρυβος %
5927	76,03	2,32	17,33			4,32
5928	31,94	61,84				6,22
5940	27,46	62,47				10,07
5943	86,96	4,35				8,69
5947	72,63	2,05	5,75	3,32	2,56	13,69
5948	68,45	24,47				7,08
5953	70,92	10,46				18,62
5964	54,09	37,74				8,17
5972	79,28	2,62	12,07			6,03
5975	32,26	57,66	4,03	4,44		1,61
5976	61,77	26,77	4,11			7,35
5987	14,37	72,75				12,88
5988	33,17	32,19	9,09	9,09		16,46
5989	35,77	37,91	8,44			17,88
6012	88,3	9,55				2,15
6014	60,43	16,53	6,5	2,98	3,79	9,77
6020	66,15	9,73	7,39	4,67		12,06
6026	37,5	28,26	21,74			12,5
6028	49,47	37,46	5,12			7,95
6032	86,15	5,66				8,19
6035	67,15	30,69				2,16
6038	19,76	35,87	3,44	27,42	2,66	10,85
6045	49,2	35,35	3,37			12,08
6063	88,03	4,21				7,76
6082	5,8	91,67				2,53
6106	44,86	39,51				15,63
6109	62,39	16,28				21,33
6168	13,76	82,07				4,17
6175	82,35	6,62	3,68			7,35
6176	39,75	40,75	11,75			7,75
6178	65,22	24,71	3,98			6,09
6192	65,6	32,48				1,92
6194	60,77	22,88	5,96	7,69		2,7
6197	79,1	19,73				1,17

Στον πίνακα 25 παρουσιάζεται μια σύνοψη των ευρημάτων σε σχέση με το πλήθος των διακριτών συστάδων που εντοπίστηκαν. Παρατηρείται πως οι χρήστες συνήθως

επικοινωνούν από ένα μικρό πλήθος γεωγραφικών τοποθεσιών, το οποίο είναι σύμφωνο με τα αποτελέσματα που παρουσιάστηκαν στο (Subramanya, 2012). Μια ακόμα σημαντική παρατήρηση από τον πίνακα 24 είναι πως για την πλειοψηφία των χρηστών με δύο και περισσότερες συστάδες υπάρχει μια συστάδα με υψηλό ποσοστό κλήσεων, που κυμαίνεται από 60% έως 90%. Μπορεί να υποτεθεί ότι αυτή η συστάδα αναπαριστά το μέρος όπου ζουν και περνούν το μεγαλύτερο μέρος του χρόνου τους. Για κάποιους χρήστες υπάρχει μια δεύτερη συστάδα με ένα σημαντικό ποσοστό κλήσεων (γύρω στο 30%), που θα μπορούσε να είναι μια τοποθεσία όπου ξοδεύουν ένα σημαντικό μέρος του χρόνου τους. Μια πιθανή εξήγηση γι' αυτούς τους χρήστες θα μπορούσε να είναι πως μετακινούνται καθημερινά σε κάποια άλλη πόλη για δουλειά ή σπουδές. Για παράδειγμα, βρέθηκε ότι ο χρήστης 6028 έχει πραγματοποιήσει το 99% των επικοινωνιών της συστάδας 2 (37,46% των συνολικών επικοινωνιών) κατά τη διάρκεια καθημερινών. Άλλα παραδείγματα περιλαμβάνουν τους χρήστες 5976 και 5989 με 98% και 100% των κλήσεων της δεύτερης μεγαλύτερης συστάδας να πραγματοποιούνται καθημερινές. Τέλος, συστάδες με πολύ μικρό ποσοστό κλήσεων αντιστοιχούν σε μέρη που οι χρήστες δεν επισκέπτονται συχνά και θα μπορούσαν να είναι δείκτες ταξιδιών για διασκέδαση ή εργασία.

Πίνακας 25. Σύνοψη αποτελεσμάτων συσταδοποίησης γεω-χαρακτηρισμένων κλήσεων

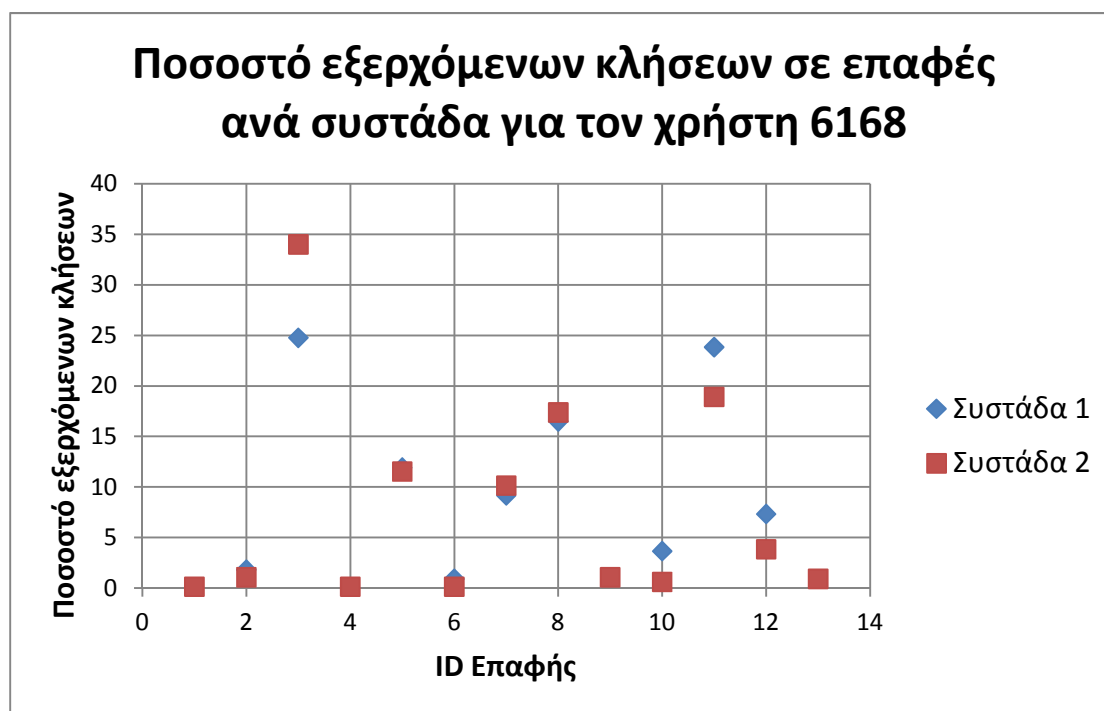
Πλήθος Συστάδων	Πλήθος Χρηστών
1	20
2	17
3	10
4	4
5	3
Σύνολο	54

5.3.3 Τύποι χρηστών

Το επόμενο βήμα μετά τη συσταδοποίηση της επικοινωνίας κάθε χρήστη με βάση την τοποθεσία ήταν η ανάλυση του ποσοστού των εξερχόμενων κλήσεων με τις επαφές που κάθε χρήστης καλεί ανά συστάδα. Μετά από αυτή την ανάλυση παρατηρήθηκε ότι υπάρχουν διαφορετικοί τύποι χρηστών ανάλογα με τα πρότυπα εξερχόμενης επικοινωνίας ανά τοποθεσία. Σε αυτή την ενότητα, παρουσιάζονται αυτοί οι τύποι χρηστών σε συνδυασμό με γραφήματα που απεικονίζουν τα μοτίβα της επικοινωνίας.

Όπως έχει ήδη αναφερθεί, υπάρχουν χρήστες που όλη τους η επικοινωνία συσταδοποιήθηκε σε μια μόνο τοποθεσία. Αυτοί οι χρήστες φαίνεται ότι δεν είχαν

ταξιδέψει (ή ότι δεν είχαν τηλεφωνική επικοινωνία όταν ταξίδευαν ή ακόμα δεν είχαν αρκετή επικοινωνία σε άλλη τοποθεσία ώστε να σχηματιστεί μια συστάδα) κατά την χρονική περίοδο της συμμετοχής τους στο πείραμα της Nokia, επομένως θα μπορούσε να πει κανείς ότι χαρακτηρίζονται από χαμηλή κινητικότητα για εκείνο το διάστημα. Η χρήση της τοποθεσίας ως διάσταση πλαισίου για την πρόβλεψη των επαφών που είναι πιο πιθανό να καλέσει ο χρήστης δε θα έχει καμία επίπτωση στην ακρίβεια της πρόβλεψης γι' αυτόν τον τύπο χρηστών, επομένως παραλείφθηκαν από περαιτέρω ανάλυση.

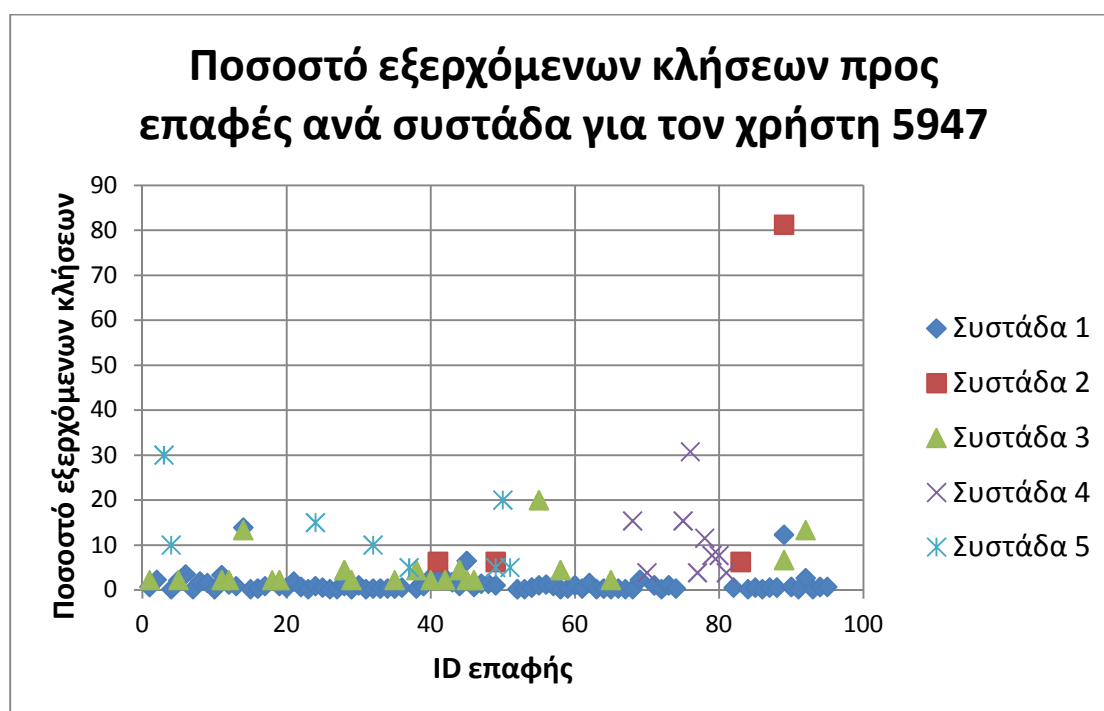


Γράφημα 48. Ποσοστό εξερχόμενων κλήσεων σε επαφές ανά συστάδα για τον χρήστη με ID 6168

Ένας άλλος παρατηρούμενος τύπος περιλαμβάνει τους χρήστες που ενώ πραγματοποιούν κλήσεις από διαφορετικές τοποθεσίες, τα πρότυπα της επικοινωνίας τους παραμένουν ανεπηρέαστα όταν αλλάζει η φυσική τους τοποθεσία, δηλαδή καλούν τις ίδιες επαφές ανεξάρτητα από το που βρίσκονται. Ένα χαρακτηριστικό παράδειγμα παρουσιάζεται στο γράφημα 48, όπου φαίνεται το ποσοστό των εξερχόμενων κλήσεων προς κάθε επαφή ανά συστάδα για τον χρήστη με ID 6168. Αυτός ο χρήστης πραγματοποιεί κλήσεις από δύο τοποθεσίες, ενώ από τον πίνακα 24 διαπιστώνεται πως η πλειοψηφία αυτής της επικοινωνίας πραγματοποιείται στη συστάδα 2. Για όλες τις επαφές που καλούνται και στις δύο συστάδες παρατηρείται ότι τα ποσοστά των εξερχόμενων κλήσεων είναι αρκετά κοντά, ανεξάρτητα από τη συστάδα της τοποθεσίας. Βεβαίως, στη συστάδα από την οποία πραγματοποιείται το μεγαλύτερο μέρος της επικοινωνίας υπάρχουν κλήσεις προς επαφές που δεν

καλούνται από την άλλη τοποθεσία. Εντούτοις, γι' αυτόν τον χρήστη παρατηρείται ότι το ποσοστό των κλήσεων προς αυτές τις επαφές είναι χαμηλό, κάτω από 5%. Γι' αυτόν τον τύπο χρηστών η διάσταση πλαισίου της φυσικής τοποθεσίας αναμένεται επίσης να μη συνεισφέρει ιδιαίτερα στη διαδικασία πρόβλεψης.

Από την άλλη, φαίνεται ότι υπάρχει ένας τρίτος τύπος χρηστών: χρήστες για τους οποίους παρατηρούνται σημαντικές αποκλίσεις όσον αφορά τα πρότυπα εξερχόμενης επικοινωνίας ανάμεσα στις διάφορες συστάδες. Μια τέτοια περίπτωση παρουσιάζεται στο γράφημα 49, όπου απεικονίζεται το ποσοστό εξερχόμενων κλήσεων προς τις επαφές ανά συστάδα για τον χρήστη με ID 5947. Είναι εμφανές ότι αυτός ο χρήστης παρουσιάζει σε κάθε συστάδα μια εντελώς διαφορετική συμπεριφορά ως προς την επικοινωνία. Εντούτοις, αυτός ο τύπος χρήστη θα μπορούσε να περιλαμβάνει και χρήστες για τους οποίους παρατηρούνται αποκλίσεις σε μερικές μόνο από τις συστάδες-τοποθεσίες, το οποίο σημαίνει πως η συμπεριφορά τους διαφοροποιείται σε μερικά μόνο από τα μέρη από τα οποία πραγματοποιούν κλήσεις. Συνεπώς, αναμένεται πως η ακρίβεια της πρόβλεψης θα μπορούσε να βελτιωθεί γι' αυτόν τον τύπο χρηστών εάν το πλαίσιο της φυσικής τοποθεσίας ενσωματωνόταν στον αλγόριθμο πρόβλεψης.



Γράφημα 49. Ποσοστό εξερχόμενων κλήσεων σε επαφές ανά συστάδα για τον χρήστη με ID 5947

Για την επικύρωση των οπτικών παρατηρήσεων, ορίστηκε μια μετρική ενδεικτική της διακύμανσης της επικοινωνίας των χρηστών ανάμεσα στις γεωγραφικές συστάδες. Καθώς οι επαφές με τις οποίες υπάρχει μικρό πλήθος επικοινωνιών είναι ουσιαστικά

ασήμαντες για τη διαδικασία της πρόβλεψης, αποκλείστηκαν αυτές με ποσοστό επικοινωνίας μικρότερο από 10% σε όλες τις συστάδες. Για τις υπόλοιπες επαφές υπολογίστηκαν για όλα τα ζεύγη συστάδων οι Ευκλείδειες αποστάσεις ανάμεσα στα διανύσματα συχνότητας επικοινωνίας των επαφών και η μετρική τέθηκε ίση με το μέσο όρο των τιμών αυτών των ζευγών. Χαμηλότερες τιμές αυτής της μετρικής είναι αντιπροσωπευτικές μικρότερων μεταβολών της επικοινωνίας ανάμεσα στις διάφορες συστάδες. Στη συνέχεια, εφαρμόστηκε ο αλγόριθμος k-means για τις τιμές αυτής της μετρικής (με εξαίρεση τους χρήστες που δεν παρουσίασαν κινητικότητα) για $k=1,2,3$ και 4. Η εκτέλεση του αλγορίθμου για $k=3$ έδωσε τα καλύτερα αποτελέσματα, καθώς για τα διαθέσιμα δεδομένα μείωσε το άθροισμα των τετραγώνων (sum of squares) εντός των συστάδων που σχηματίστηκαν. Συνεπώς, υπάρχει ακόμα ένας τύπος χρήστη ανάμεσα σε αυτούς τους περιγράφησαν στις προηγούμενες δύο παραγράφους. Αυτός ο τύπος περιγράφει χρήστες σε μια ενδιάμεση κατάσταση όσον αφορά την απόκλιση της επικοινωνίας τους στις διάφορες συστάδες.

Τελικά, τα αποτελέσματα του αλγορίθμου k-means παρουσιάζονται στον πίνακα 26. Συνολικά, από τους 54 χρήστες που αναλύθηκαν προέκυψαν οι παρακάτω 4 τύποι: 20 χρήστες χωρίς καθόλου κινητικότητα, 13 χρήστες που παρουσιάζουν χαμηλή απόκλιση επικοινωνίας ανάμεσα στις συστάδες, 15 χρήστες με μέτρια απόκλιση επικοινωνίας και 6 χρήστες με υψηλή απόκλιση επικοινωνίας.

Πίνακας 26. Πλήθος χρηστών ανά τύπο χρήστη με βάση την απόκλιση της επικοινωνίας

Τύπος Χρήστη	Πλήθος Χρηστών
Χωρίς κινητικότητα	20
Χαμηλή απόκλιση επικοινωνίας	13
Μέτρια απόκλιση επικοινωνίας	15
Υψηλή απόκλιση επικοινωνίας	6
Σύνολο	54

5.3.4 Αποτελέσματα πρόβλεψης

Καθώς, σκοπός της έρευνας που παρουσιάζεται σε αυτή την ενότητα ήταν η μελέτη της επίπτωσης του πλαισίου της φυσικής τοποθεσίας αποκλειστικά, υλοποιήθηκε μια παραλλαγή του αλγορίθμου πρόβλεψης, στην οποία χρησιμοποιήθηκε μόνο η διάσταση της συχνότητας επικοινωνίας. Για τη δημιουργία μιας βάσης σύγκρισης της απόδοσης, ορίστηκε η μετρική πρόβλεψης να είναι ίση με τη συνολική συχνότητα εξερχόμενων κλήσεων για κάθε επαφή. Μια τέτοια επιλογή προσομοιώνει την οθόνη με τη λίστα συχνά χρησιμοποιούμενων επαφών (frequently called) της εφαρμογής επαφών σε κάποιες συσκευές Android. Για τη σύγκριση με την περίπτωση όπου

χρησιμοποιείται η διάσταση της φυσικής τοποθεσίας, πραγματοποιήθηκε μια δεύτερη εκτέλεση του αλγορίθμου, όπου αντικαταστάθηκε η διάσταση της συνολικής συχνότητας κλήσης των επαφών με τη συχνότητα εξερχόμενων κλήσεων κάθε επαφής από τη συστάδα στην οποία βρίσκεται ο χρήστης κατά τη χρονική στιγμή που πραγματοποιήθηκε η κλήση. Με άλλα λόγια, σε κάθε πρόβλεψη λήφθηκαν υπόψη μόνο οι κλήσεις που πραγματοποιήθηκαν από την τοποθεσία στην οποία βρισκόταν ο χρήστης. Η συνάρτηση σκορ, επομένως, γίνεται σε αυτή την περίπτωση

$$FL_t: C \rightarrow [0,1]$$

$$FL_t(c_k) = \frac{\sum \text{κλήσεων στη συστάδα που βρίσκεται ο χρήστης με τη } c_k \text{ στο } t_w}{\sum \text{κλήσεων στη συστάδα που βρίσκεται ο χρήστης στο } t_w}$$

Όπως εξηγήθηκε στην προηγούμενη ενότητα, θα ανέμενε κανείς μια αξιοσημείωτη επίπτωση αυτής της διάστασης πλαισίου κυρίως για χρήστες που παρουσίαζαν υψηλή απόκλιση της κατανομής της συχνότητας επικοινωνίας ανάμεσα στις διάφορες γεωγραφικές συστάδες.

Στον αλγόριθμο πρόβλεψης που παρουσιάστηκε στο προηγούμενο κεφάλαιο, χρησιμοποιήθηκε ένα χρονικό παράθυρο εκπαίδευσης για τη διευκόλυνση της πρόβλεψης. Αυτή η προσέγγιση θα ήταν προβληματική εδώ, αφού θα οδηγούσε σε άδεια σύνολα δεδομένων εκπαίδευσης για συστάδες με μικρή δραστηριότητα κλήσεων, τις οποίες ο χρήστης δεν είχε επισκεφθεί κατά τη διάρκεια του παραθύρου εκπαίδευσης. Για την αντιμετώπιση αυτού του εμποδίου, πραγματοποιήθηκε μια μικρή διαφοροποίηση στον αλγόριθμο, ώστε αντί για ένα χρονικό παράθυρο εκπαίδευσης να χρησιμοποιείται ένα αριθμητικό, υπολογίζοντας για κάθε χρήστη το πλήθος των κλήσεων ανά ημέρα που κάλυπτε το αρχείο κλήσεων και πολλαπλασιάζοντας με το πλήθος των ημερών του αντίστοιχου χρονικού παραθύρου (15 ημέρες). Ακολουθώντας την ίδια διαδικασία προσομοίωσης που περιγράφηκε στο προηγούμενο κεφάλαιο, για 1, 3 και 5 θέσεις πρόβλεψης προέκυψαν τα αποτελέσματα που παρουσιάζονται στον πίνακα 27. Για κάθε πλήθος προτάσεων, η αριστερή στήλη δείχνει το ποσοστό επιτυχίας της πρόβλεψης για το πείραμα της βάσης σύγκρισης (B1, B3 και B5), όπου χρησιμοποιήθηκε η συχνότητα κλήσεων με την κλασική της έννοια, ενώ η δεξιά στήλη παρουσιάζει την επίπτωση της φυσικής τοποθεσίας ως τη διαφορά του ποσοστού επιτυχίας από την εκτέλεση του αλγορίθμου με την εξαρτώμενη από την τοποθεσία συχνότητα κλήσεων σε σχέση με το πείραμα βάσης (ΔL1, ΔL3 και ΔL5). Οι χρήστες που δεν παρουσίαζαν καθόλου κινητικότητα αποκλείστηκαν από την ανάλυση, αφού δε θα υπήρχε διαφορά στην ακρίβεια της πρόβλεψης από την εκτέλεση των δύο πειραματικών προσομοιώσεων.

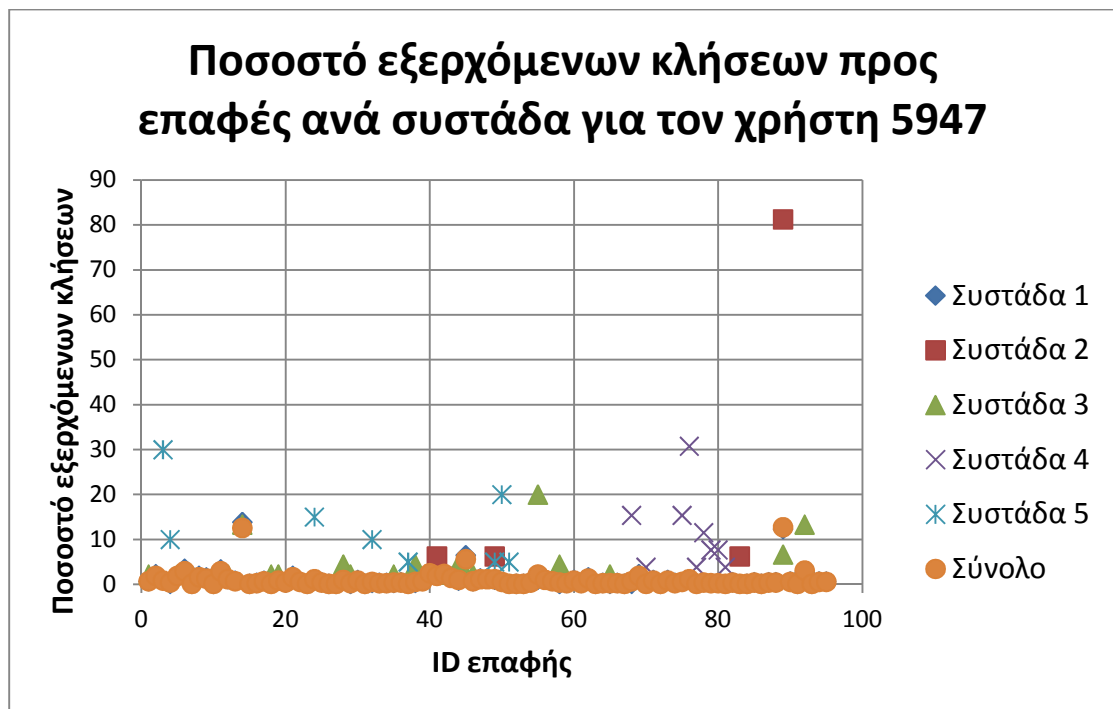
Πίνακας 27. Επίπτωση της φυσικής τοποθεσίας στην ακρίβεια πρόβλεψης (ΔLx) στα αποτελέσματα της βασικής απόδοσης (Bx) για 1, 3 και 5 θέσεις προτάσεων

User ID	B1%	$\Delta L1\%$	B3%	$\Delta L3\%$	B5%	$\Delta L5\%$
<i>Χαμηλή απόκλιση επικοινωνίας</i>						
5927	91	0	96	-1	97	0
5940	37	2	66	2	74	-1
5948	57	0	77	-3	81	-1
5953	19	-1	34	0	44	2
5964	12	0	28	0	34	-2
5988	61	-5	78	-1	83	-4
6012	71	-1	91	0	95	-1
6082	57	0	81	0	86	-1
6106	19	-4	42	-3	52	-2
6168	32	-1	70	0	89	-1
6176	53	-4	71	-7	77	-5
6192	14	-1	33	0	47	0
6197	37	2	68	-2	79	-1
<i>Μέτρια απόκλιση επικοινωνίας</i>						
5928	77	-2	93	-1	95	0
5943	11	0	29	0	41	0
5972	19	2	42	0	51	0
5976	32	6	66	2	77	-2
5987	16	2	40	0	56	-1
5989	52	1	71	0	78	0
6014	18	-1	35	2	49	1
6026	28	0	60	-5	75	-7
6035	22	-1	42	2	58	-1
6038	43	-2	82	0	96	0
6045	51	-1	63	1	71	1
6063	35	1	53	0	62	-1
6109	35	0	51	0	60	-1
6178	21	-2	45	-3	58	-4
6194	27	-3	56	-1	70	-3
<i>Υψηλή απόκλιση επικοινωνίας</i>						
5947	25	1	51	0	60	1
5975	51	11	67	-3	75	13
6020	22	1	42	3	52	8
6028	49	4	74	3	82	0
6032	58	0	76	0	84	0
6175	15	-1	38	-1	47	-1

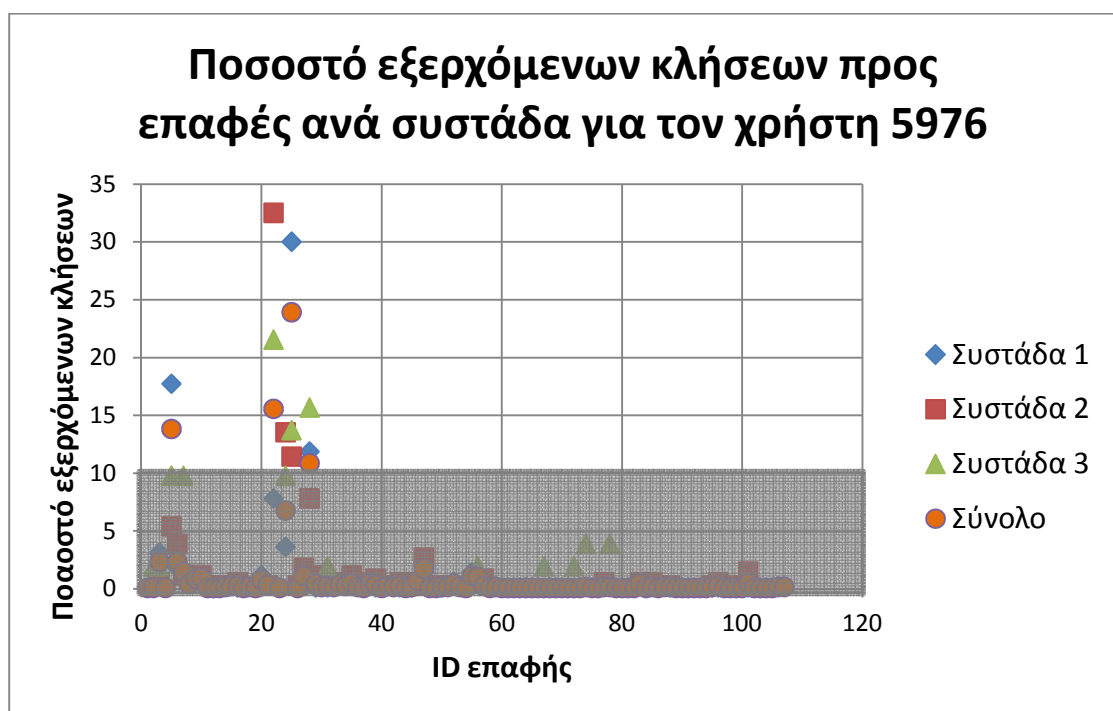
Συνολικά, η ενσωμάτωση της φυσικής τοποθεσίας είχε μια επίπτωση 0,09% ($sd=2,90\%$) στους χρήστες για τη 1 θέση πρόβλεψης, -0,47% ($sd=2,10\%$) για τις 3 θέσεις πρόβλεψης και -0,41% ($sd=3,35\%$) για τις 5 θέσεις πρόβλεψης. Εντούτοις, η στατιστική ανάλυση έδειξε ότι αυτές οι διαφορές δεν είναι στατιστικά σημαντικές (Wilcoxon Signed-Ranked Test για την περίπτωση της 1 θέσης πρόβλεψης με $Z=-0,322$ και $p=0,747$, Paired Samples T-Test για τις 3 και 5 θέσεις πρόβλεψης με $t(33)=1,302$ $p=0,202$ και $t(33)=0,715$ $p=0,480$ αντίστοιχα).

Αναλύοντας τα δεδομένα ανά τύπο χρηστών προκύπτουν τα παρακάτω αποτελέσματα: Για τους χρήστες με χαμηλή απόκλιση επικοινωνίας παρουσιάστηκε μια μέση επίπτωση -1,00% ($sd=2,16\%$) για 1 θέση πρόβλεψης, -1,15% ($sd=2,23\%$) για 3 θέσεις πρόβλεψης και -1,30% ($sd=1,75\%$) για 5 θέσεις πρόβλεψης. Οι διαφορές είναι στατιστικά σημαντικές μόνο για την περίπτωση των 5 προτάσεων (Paired Samples T-Test με $t(12)=1,669$ $p=0,121$ για 1 θέση πρόβλεψης, $t(12)=1,865$ $p=0,087$ για 3 θέσεις και $t(12)=2,694$ $p=0,02$ για 5 θέσεις). Για τους χρήστες με μέτρια απόκλιση επικοινωνίας προέκυψε μια μέση διαφορά 0% ($sd=2,24\%$) για τη μια θέση πρόβλεψης, -0,20% ($sd=1,86\%$) για τις 3 θέσεις και -1,20% ($sd=2,11\%$) για τις 5 θέσεις. Οι διαφορές είναι στατιστικά σημαντικές μόνο για την περίπτωση των 5 προτάσεων (Paired Samples T-Test με $t(14)=0$ $p=1$ για 1 θέση πρόβλεψης, $t(14)=0,417$ $p=0,683$ για 3 θέσεις και $t(14)=2,201$ $p=0,045$ για 5 θέσεις). Τέλος, για τους χρήστες που ανήκουν στην κατηγορία υψηλής απόκλισης επικοινωνίας υπάρχει μια μέση διαφορά 2,67% ($sd=4,41\%$) για τη 1 θέση, 0,33% ($sd=2,34\%$) για τις 3 θέσεις και 3,50% ($sd=5,68\%$) για τις 5 θέσεις. Οι διαφορές δεν είναι στατιστικά σημαντικές για κανένα μήκος της λίστας προτάσεων (Paired Samples T-Test με $t(5)=-1,480$ $p=0,199$ για τη 1 θέση πρόβλεψης, $t(5)=-0,349$ $p=0,741$ για τις 3 θέσεις και $t(5)=-1,508$ $p=0,192$ για τις 5 θέσεις).

Τα αποτελέσματα αυτά καταδεικνύουν το γεγονός πως η ενσωμάτωση της φυσικής τοποθεσίας στον αλγόριθμο πρόβλεψης δεν οδηγεί απαραίτητα σε βελτίωση της ακρίβειας πρόβλεψης. Όταν οι διαφορές δεν είναι στατιστικά σημαντικές, τα αποτελέσματα πρόβλεψης δεν επηρεάζονται από την τοποθεσία. Στις περισσότερες περιπτώσεις που τα αποτελέσματα είναι στατιστικά σημαντικά, οι διαφορές είναι οριακές, ενώ υπάρχουν χρήστες για τους οποίους ο αλγόριθμος οδήγησε σε ελαφρώς χειρότερα ή καλύτερα αποτελέσματα. Για τους χρήστες που ανήκουν στην κατηγορία της χαμηλής απόκλισης της επικοινωνίας, με παρόμοια πρότυπα επικοινωνίας σε όλες τις γεωγραφικές συστάδες, κάτι τέτοιο φαίνεται φυσιολογικό.



Γράφημα 50. Κατανομή συχνότητας εξερχόμενης επικοινωνίας για τον χρήστη με ID 5947



Γράφημα 51. Κατανομή συχνότητας εξερχόμενης επικοινωνίας για τον χρήστη με ID 5976

Στη συνέχεια, αναλύονται μερικά τυπικά παραδείγματα, αντιπροσωπευτικά της πλειοψηφίας των χρηστών που υπάρχουν στο υπό ανάλυση σύνολο δεδομένων, ώστε να εξηγηθούν τα ευρήματα για τους χρήστες που εμφανίζουν διαφορετικά πρότυπα επικοινωνίας σε διαφορετικές συστάδες (χρήστες μέτριας και υψηλής απόκλισης). Μια αρχική παρατήρηση είναι πως κάποιοι από αυτούς πραγματοποιούσαν την

πλειοψηφία των εξερχόμενων κλήσεων κυρίως από μια συστάδα, επομένως οι αποκλίσεις στα πρότυπα επικοινωνίας στις υπόλοιπες συστάδες διαδραματίζουν μικρό ρόλο στη συνολική ακρίβεια των προβλέψεων. Μια τέτοια τυπική περίπτωση, είναι ο χρήστης 5947 (υψηλή απόκλιση), ο οποίος επικοινωνούσε κυρίως από τη συστάδα 1, όπως φαίνεται στον πίνακα 24. Στο γράφημα 50, παρατηρεί κανείς ότι γι' αυτόν τον χρήστη η κατανομή της συχνότητας επικοινωνίας με τις επαφές για τη συστάδα 1 σχεδόν συμπίπτει με την κατανομή συχνότητας της συνολικής εξερχόμενης επικοινωνίας.

Μια διαφορετική εξήγηση αντιστοιχεί σε χρήστες όπως ο 5976 (μέτρια απόκλιση). Αυτός ο χρήστης πραγματοποιούσε κλήσεις από 3 συστάδες, όμως μόνο 2 συστάδες έχουν σημαντική επίπτωση στην επικοινωνιακή του συμπεριφορά, καθώς η επικοινωνία από τη συστάδα 3 αντιστοιχεί μόνο στο 4% της συνολικής επικοινωνίας (πίνακας 24). Αγνοώντας επαφές με συχνότητα κλήσεων κάτω από 10% (γκρι περιοχή στο γράφημα 51), οι οποίες είναι συνήθως δύσκολο να προβλεφθούν από την ανάλυση του ιστορικού κλήσεων, παρατηρείται ότι απομένουν μόνο 3-4 επαφές. Κλήσεις προς μία τόσο ολιγομελή ομάδα επαφών είναι εύκολο να προβλεφθούν ακόμα και χωρίς να χρησιμοποιηθεί πλαίσιο όπως η τοποθεσία, όταν ο αλγόριθμος προσφέρει 3 ή 5 προβλέψεις. Στο εν λόγω παράδειγμα, αυτό είναι ακόμα πιο ξεκάθαρο, καθώς οι πιο συχνά κληθέντες επαφές για τη συστάδα 1, τη συστάδα 2 και τη συνολική επικοινωνία, ήταν οι ίδιες με διαφορετική κατάταξη.

Μια παρατήρηση των αρχείων κλήσεων για τους υπό εξέταση χρήστες και μια εξέταση των αποτυχιών της διαδικασίας πρόβλεψης όταν η εξαρτώμενη από την τοποθεσία συχνότητα χρησιμοποιείται αντί για τη συνηθισμένη διάσταση της συχνότητας, αποκάλυψε ότι αυτές οι αποτυχίες προκύπτουν συνήθως όταν γινόταν μετάβαση από μια τοποθεσία σε μια άλλη και κυρίως όταν η τοποθεσία – προορισμός δεν ήταν αυτή με τη μεγαλύτερη δραστηριότητα επικοινωνίας. Όταν η τοποθεσία χρησιμοποιήθηκε, παρατηρήθηκε ότι το παράθυρο εκπαίδευσης για τις «δευτερεύουσες» συστάδες, τις τοποθεσίες δηλαδή που οι χρήστες δεν επισκέπτονταν συχνά ή στις οποίες περνούν σύντομης διάρκειας χρονικά διαστήματα, μπορεί να περιέχουν παλαιότερες κλήσεις από προγενέστερες επισκέψεις των χρηστών. Συνεπώς, αυτό το παράθυρο εκπαίδευσης αποτύγχανε να αναπαραστήσει κατάλληλα το πρόσφατο πρότυπο επικοινωνίας των χρηστών, το οποίο παρέχει καλύτερα αποτελέσματα μελλοντικών προβλέψεων (Phithakkitnukoon, Dantu, Claxton & Eagle, 2011). Όταν οι χρήστες έμεναν για μεγάλα χρονικά διαστήματα (περισσότερο από 15 ημέρες) στην ίδια συστάδα, οι δύο μέθοδοι σταδιακά παρείχαν τα ίδια αποτελέσματα (καθώς και οι δύο ανέλυναν τα ίδια σύνολα δεδομένων εκπαίδευσης για

την παραγωγή των προβλέψεων). Συμπερασματικά, αυτές οι αποτυχίες μετάβασης είναι κρίσιμες για την απόδοση του αλγορίθμου πρόβλεψης όταν η τοποθεσία λαμβάνεται υπόψη.

5.4 Μελέτη της επίπτωσης της σημασιολογικής τοποθεσίας

Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα της μελέτης της επίπτωσης της σημασιολογικής τοποθεσίας στην αξιοποίηση των επαφών για την έναρξη εξερχόμενης επικοινωνίας.

5.4.1 Σύνολο δεδομένων

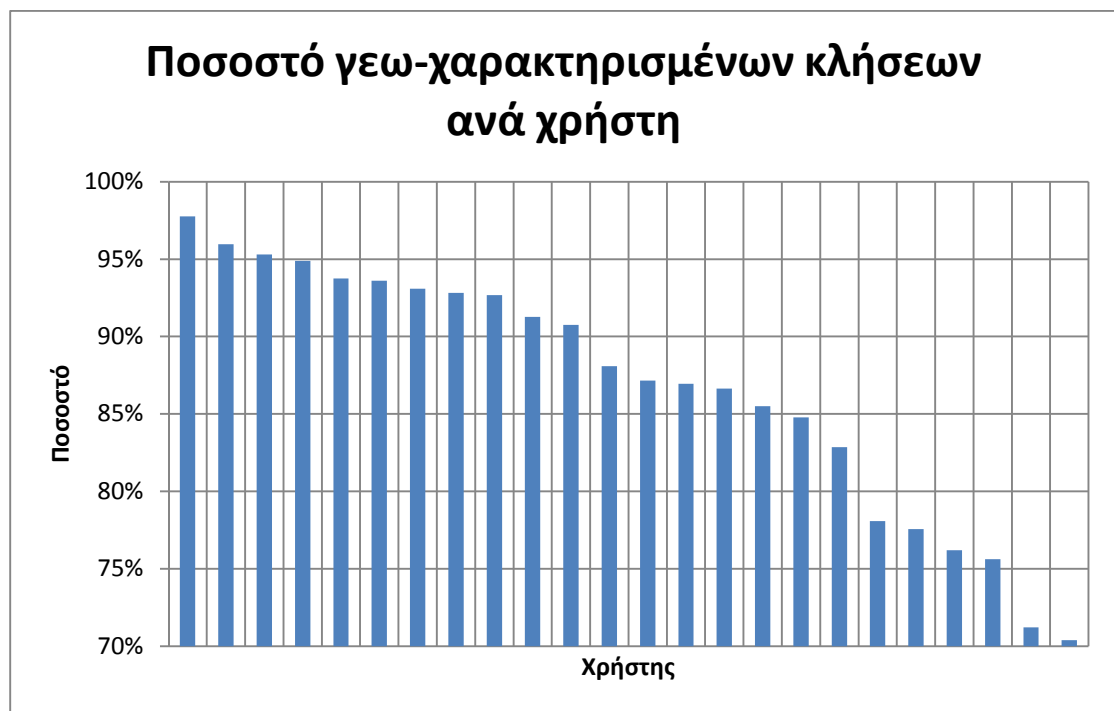
Για τη διερεύνηση της επίπτωσης της σημασιολογικής τοποθεσίας ήταν απαραίτητη η χρησιμοποίηση ενός συνόλου δεδομένων στο οποίο υπάρχει διαθέσιμη αυτή η πληροφορία. Στο σύνολο δεδομένων της Nokia, αυτή η πληροφορία δεν είναι διαθέσιμη, αλλά, όπως παρουσιάστηκε στην προηγούμενη ενότητα, υπάρχουν μόνο εγγραφές που αφορούν τη φυσική τοποθεσία των χρηστών, χωρίς να τους έχει ανατεθεί κάποια σημασιολογική ετικέτα. Προφανώς, η εξαγωγή σημασιολογικών ετικετών από τις φυσικές τοποθεσίες είναι εφικτή σε συνδυασμό με δημογραφικά δεδομένα (τα οποία πάντως δεν είναι δημόσια διαθέσιμα γι' αυτό το σύνολο δεδομένων) για τους χρήστες και μάλιστα ήταν ένα από τα βασικά θέματα του διαγωνισμού που οργάνωσε η Nokia για την αξιοποίηση των δεδομένων από την επιστημονική κοινότητα (Laurila, Gatica-Perez, Aad, Borner, Do, Dousse, ... & Miettinen, 2012). Εντούτοις, μια τέτοια ανάλυση ήταν εκτός του σκοπού της συγκεκριμένης διδακτορικής έρευνας, ενώ και οι προσεγγίσεις που έχουν προταθεί και βασίζονται στο συγκεκριμένο σύνολο δεδομένων πετυχαίνουν ποσοστό ακρίβειας μεταξύ 65% και 75% (Krumm & Rouhana, 2013).

Για τους παραπάνω λόγους αποφασίστηκε να αξιοποιηθεί ένα άλλο διαθέσιμο σύνολο δεδομένων, το Reality Mining Dataset από το MIT (Eagle & Pentland, 2006), στο οποίο υπάρχουν διαθέσιμα δεδομένα που αφορούν τη σημασιολογική τοποθεσία των χρηστών. Το σύνολο δεδομένων συλλέχθηκε κατά τη διάρκεια ενός πειράματος που αφορούσε 100 χρήστες (κυρίως φοιτητές του MIT) στους οποίους δόθηκαν συσκευές Nokia 6600 και διήρκεσε μια ακαδημαϊκή χρονιά (9 μήνες). Η πληροφορία που συλλέχθηκε αφορά κυρίως αρχεία κλήσεων, κοντινές συσκευές Bluetooth, Cell-IDs από GSM κεραίες, χρήση εφαρμογών, κατάσταση τηλεφώνου (π.χ. σε αδράνεια ή σε φόρτιση κ.λπ.) και προέρχεται κυρίως από την εφαρμογή Context (Raento, Oulasvirta, Petit & Toivonen, 2005). Το σύνολο των δεδομένων που συλλέχθηκαν αντιπροσωπεύει 450.000 ώρες πληροφορίας σχετικά με την επικοινωνία των χρηστών, την τοποθεσία τους και τη χρήση των συσκευών. Το σύνολο δεδομένων

είναι αρκετά πλήρες, καθώς περιλαμβάνει δεδομένα για περίπου 85% του χρόνου που διήρκεσε το πείραμα, με την κυριότερη αιτία μη καταγραφής να είναι το γεγονός πως οι χρήστες απενεργοποιούσαν τις συσκευές κατά τη διάρκεια της νύχτας. Οι ερευνητές βασίστηκαν στα δεδομένα από τις κεραίες κινητής τηλεφωνίας, από τις κοντινές συσκευές Bluetooth και από στατικές Bluetooth κεραίες που βρίσκονταν σε γνωστά σημεία, καθώς και σε χρονικές πληροφορίες που συλλέχθηκαν από τις συσκευές ώστε να προτείνουν ένα μοντέλο ταξινόμησης σε τρεις καταστάσεις: «σπίτι», «δουλειά» και «αλλού». Συνεπώς, η σημασιολογική αυτή πληροφορία για την τοποθεσία των χρηστών είναι διαθέσιμη (με τη μορφή πιθανοτήτων ανά ώρα για κάθε χρήστη, οπότε και επιλέγεται η ετικέτα με τη μεγαλύτερη πιθανότητα όπως φαίνεται στην εικόνα 35), έστω και αν πρόκειται μόνο για τρεις σημασιολογικές ετικέτες, από τις οποίες η μια («αλλού») δεν ορίζει με σαφήνεια την κατάσταση του χρήστη.

user_id	id	day	event_id	contact	description	type	home	elsewhere	work	nosig	cluster	time
4	18	2004-08-05 16:40:39	317	95	Voice call	2	0	0,0025	0,893056	0,0486111	3	1091713239
4	19	2004-08-05 21:02:30	318	106	Voice call	2	0	0,493333	0,473056	0	2	1091728950
4	20	2004-08-06 21:39:05	319	106	Voice call	2	0,0666667	0,767222	0	0,148056	2	1091817545
4	21	2004-08-06 22:35:35	320	106	Voice call	3	0	0,942222	0	0,0322222	2	1091820935
4	23	2004-08-06 23:06:21	322	9	Voice call	2	0	0,985	0	0,00388889	2	1091822781

Εικόνα 35. Παράδειγμα εγγραφών από τον πίνακα κλήσεων, όπου έχουν ενσωματωθεί οι πιθανότητες του ταξινομητή για τις σημασιολογικές τοποθεσίες και έχει αποδοθεί η μεγαλύτερη



Γράφημα 52. Ποσοστό γεω-χαρακτηρισμένων κλήσεων ανά χρήστη σε φθίνουσα σειρά

Πριν από την εκτέλεση των πειραμάτων για τη μελέτη της επίπτωσης της σημασιολογικής τοποθεσίας, εξετάστηκαν ποιοι από τους χρήστες που περιλαμβάνονταν στο σύνολο δεδομένων ήταν κατάλληλοι για επεξεργασία. Και σε

αυτή την περίπτωση υπάρχουν κλήσεις που για το διάστημα κατά το οποίο πραγματοποιήθηκαν δεν υπάρχουν δεδομένα ή η συσκευή δεν είχε σήμα ή για κάποιον άγνωστο λόγο δεν είναι διαθέσιμη η σημασιολογική ετικέτα της τοποθεσίας. Επομένως, όπως έγινε και με το σύνολο δεδομένων της Nokia, αποκλείστηκαν από την ανάλυση όσοι χρήστες είχαν σε ποσοστό μικρότερο από 70% γεω-χαρακτηρισμένες κλήσεις. Ακόμα, εφαρμόστηκαν και πάλι τα ίδια κριτήρια αποκλεισμού που σχετίζονται με την επικοινωνία. Τελικά, απέμειναν 24 χρήστες με ποσοστό γεω-χαρακτηρισμένων κλήσεων κατά μέσο όρο 86,78% (sd=8,12%), όπως φαίνεται και στο γράφημα 52.

Στον πίνακα 28 παρουσιάζεται η κατανομή των γεω-χαρακτηρισμένων κλήσεων ανά χρήστη και ανά σημασιολογική τοποθεσία. Μπορεί να παρατηρήσει κανείς ότι για κάποιους χρήστες δεν υπάρχουν καθόλου κλήσεις από την τοποθεσία «σπίτι». Αυτό μπορεί να οφείλεται σε αδυναμία ή λάθη του ταξινομητή λόγω ελλιπών δεδομένων (σε αυτή την εσωτερική τοποθεσία είναι πιο δύσκολος ο εντοπισμός της τοποθεσίας λόγω της έλλειψης σταθερών κεραιών Bluetooth σε αντίθεση με τα εργαστήρια του MIT όπου φοιτούσαν οι συμμετέχοντες) ή στο ότι πράγματι δε χρησιμοποιούσαν ιδιαίτερα για επικοινωνία τις συσκευές σε αυτή την τοποθεσία. Οι κλήσεις που έχουν χαρακτηριστεί ως θόρυβος είναι κλήσεις που έγιναν σε χρονικές περιόδους που η συσκευή δεν είχε σήμα (κατάσταση “nosig”) για αρκετά μεγάλο διάστημα και επομένως ο ταξινομητής δεν τις κατέταξε σε κάποια από τις τρεις τοποθεσίες.

Πίνακας 28. Κατανομή κλήσεων στις σημασιολογικές τοποθεσίες

ID Χρήστη	Σπίτι%	Δουλειά%	Αλλού%	Θόρυβος %
10		49,90	48,97	1,13
15	75,99	20,24	1,98	1,79
17	48,97	43,79	3,79	3,45
26	36,55	53,80	9,24	0,41
27	63,08	21,27	13,45	2,20
33		15,40	69,98	14,62
37	30,68	18,53	46,61	4,18
38	53,81	31,23	12,83	2,14
40		27,64	70,57	1,79
42	44,76	32,97	20,54	1,73
50		10,08	89,35	0,57
52		39,88	56,67	3,44
55	31,41	40,25	26,01	2,33
61	49,32	38,96	11,17	0,54
62		81,89	18,11	
63		27,56	69,60	2,84
67	53,77	42,45	1,97	1,81

71		39,13	53,14	7,73
73	19,55	59,34	19,72	1,38
76		91,59	5,61	2,80
77	41,46	35,32	19,58	3,65
94	76,99	22,19	0,66	0,17
96	49,76	40,38	8,89	0,96
97	18,84	23,95	55,96	1,25

Επιπλέον, εφαρμόστηκε η ίδια ανάλυση για την κατηγοριοποίηση των χρηστών ανάλογα με την απόκλιση που παρουσιάζει η επικοινωνία τους ανάμεσα στις διαφορετικές σημασιολογικές τοποθεσίες, όπως είχε γίνει και για την περίπτωση του συνόλου δεδομένων που χρησιμοποιήθηκε για την μελέτη της φυσικής τοποθεσίας. Οι χρήστες κατατάχθηκαν στις τρεις τύπους, όπως φαίνεται στον πίνακα 29.

Πίνακας 29. Πλήθος χρηστών ανά τύπο χρήστη με βάση την απόκλιση της επικοινωνίας

Τύπος Χρήστη	Πλήθος Χρηστών
Χαμηλή απόκλιση επικοινωνίας	15
Μέτρια απόκλιση επικοινωνίας	7
Υψηλή απόκλιση επικοινωνίας	2
Σύνολο	24

5.4.2 Αποτελέσματα πρόβλεψης

Σε αυτή την περίπτωση δε χρειάστηκε να πραγματοποιηθεί κάποια συσταδοποίηση, εφόσον οι τοποθεσίες ήταν δεδομένες (σπίτι, δουλειά, αλλού). Επομένως, οι ίδιες πειραματικές προσομοιώσεις που σχεδιάστηκαν για τη μελέτη της επίπτωσης της φυσικής τοποθεσίας εκτελέστηκαν και για την περίπτωση της σημασιολογικής τοποθεσίας. Επομένως, αρχικά εκτελέστηκε ένα πείραμα ως βάση σύγκρισης για 1,3 και 5 προτάσεις (στήλες B1, B3 και B5 στον πίνακα 30) με γραμμικό παράθυρο εκπαίδευσης μήκους 15 ημερών, όπου χρησιμοποιήθηκε η συνολική συχνότητα επικοινωνίας ως διάσταση πλαισίου στον αλγόριθμο πρόβλεψης. Στη συνέχεια, εκτελέστηκε ξανά η ίδια προσομοίωση, αυτή τη φορά χρησιμοποιώντας ως διάσταση πλαισίου την εξαρτημένη από τη σημασιολογική τοποθεσία συχνότητα, δηλαδή για κάθε πρόβλεψη λήφθηκαν υπόψη μόνο οι κλήσεις από τη σημασιολογική τοποθεσία στην οποία βρισκόταν ο χρήστης τη στιγμή που πραγματοποιήθηκε η κλήση που ο αλγόριθμος προσπαθούσε να προβλέψει και τα αποτελέσματα συγκρίθηκαν με αυτά του πειράματος βάσης (στήλες ΔL1, ΔL3 και ΔL5 του πίνακα 30).

Πίνακας 30. Επίπτωση της σημασιολογικής τοποθεσίας στην ακρίβεια πρόβλεψης (ΔLx) στα αποτελέσματα της βασικής απόδοσης (Bx) για 1, 3 και 5 θέσεις προτάσεων

User ID	B1%	$\Delta L1\%$	B3%	$\Delta L3\%$	B5%	$\Delta L5\%$
Χαμηλή απόκλιση επικοινωνίας						
10	22	0	52	0	67	-2
26	20	-7	48	-4	66	-8
27	58	4	77	-2	85	0
33	41	-4	60	-2	73	-5
37	67	1	84	0	91	-2
38	8	-1	20	-1	29	0
40	40	1	55	4	66	2
42	26	-1	53	0	68	4
52	15	2	30	2	41	-2
55	30	-1	59	10	75	12
67	34	2	57	2	69	2
73	59	5	93	2	96	0
77	24	-1	53	-1	69	5
96	23	-12	36	-5	51	-5
97	37	0	53	-3	65	-2
Μέτρια απόκλιση επικοινωνίας						
15	39	1	63	2	76	1
17	15	-2	40	3	56	8
50	66	2	82	1	93	0
61	62	-2	82	1	90	0
63	40	-1	70	-3	83	0
71	37	3	60	1	63	6
76	10	0	26	-3	36	-2
Υψηλή απόκλιση επικοινωνίας						
62	28	3	48	-5	67	-6
94	32	-1	61	-2	74	-2

Στην περίπτωση εκτέλεσης των προσομοιώσεων για μια θέση πρόβλεψης η χρησιμοποίηση της σημασιολογικής τοποθεσίας είχε μια μέση επίπτωση για τον συνολικό πληθυσμό $-0,37\%$ ($sd=3,57\%$). Αντίστοιχα, για την περίπτωση των τριών θέσεων πρόβλεψης, η επίπτωση της τοποθεσίας κατά μέσο όρο ήταν $-0,12\%$ ($sd=3,30\%$), ενώ για τις πέντε θέσεις πρόβλεψης υπήρξε μια μικρή μέση βελτίωση $0,17\%$ ($sd=4,50\%$). Η στατιστική ανάλυση έδειξε ότι σε καμία από τις παραπάνω περιπτώσεις οι διαφορές δεν ήταν στατιστικά σημαντικές (Paired Samples T-Test με $t(23)=0,514$ $p=0,612$ για 1 θέση πρόβλεψης, $t(23)=0,186$ $p=0,854$ για 3 θέσεις και $t(23)=-0,181$ $p=0,858$ για 5 θέσεις).

Στη συνέχεια παρουσιάζεται η ανάλυση των αποτελεσμάτων ανά τύπο χρήστη. Για τους χρήστες που παρουσιάζουν χαμηλή απόκλιση επικοινωνίας, βρέθηκε μια μέση

επίπτωση -0,80% (sd=4,26%) από τη χρήση της σημασιολογικής τοποθεσίας για την εξαγωγή των προβλέψεων για τη λίστα μιας θέσης, ενώ για τις τρεις θέσεις η μέση διαφορά ήταν 0,13% (sd=3,66%) και για τις 5 θέσεις -0,07% (sd=4,80%). Καμία από αυτές τις διαφορές δεν είναι στατιστικά σημαντικές (Paired Samples T-Test με $t(14)=0,727$ $p=0,479$ για 1 θέση πρόβλεψης, $t(14)=-0,141$ $p=0,890$ για 3 θέσεις πρόβλεψης και $t(14)=0,054$ $p=0,958$ για 5 θέσεις πρόβλεψης). Για τους χρήστες που παρουσιάζουν μέτρια απόκλιση επικοινωνίας, η χρήση της σημασιολογικής τοποθεσίας οδήγησε σε μια μέση βελτίωση 0,14% (sd=1,95%) στην περίπτωση της μιας θέσης πρόβλεψης, ενώ είχε μια μέση επίπτωση 0,28% (sd=2,36%) για τις τρεις θέσεις και 1,85% (sd=3,67%) για τις 5 θέσεις πρόβλεψης. Και πάλι οι διαφορές που παρατηρήθηκαν δεν είναι στατιστικά σημαντικές (Paired Samples T-Test με $t(6)=-0,194$ $p=0,853$ για τη 1 θέση πρόβλεψης, $t(6)=-0,320$ $p=0,760$ για τις 3 θέσεις πρόβλεψης και $t(6)=-1,338$ $p=0,229$ για τις 5 θέσεις πρόβλεψης). Τέλος, δεν έχει νόημα η εξέταση των χρηστών υψηλής απόκλισης επικοινωνίας ξεχωριστά, λόγω του ότι αυτός ο τύπος χρήστη περιελάμβανε μόνο δύο χρήστες.

5.5. Συμπεράσματα

Στις προηγούμενες παραγράφους εξετάστηκε η επίπτωση της τοποθεσίας (φυσικής και σημασιολογικής) ή παραγόντων που σχετίζονται έμμεσα με αυτή (όπως το χρονικό πλαίσιο) ως παράγοντα πρόβλεψης της επόμενης κλήσης, μια διαδικασία χρήσιμη για το σχεδιασμό προσαρμοστικών διεπαφών για την αξιοποίηση των επαφών. Η τοποθεσία θεωρείται στη βιβλιογραφία ένας σημαντικός παράγοντας πλαισίου για τους χρήστες συσκευών κινητού υπολογισμού και προτείνεται συχνά ως μια υποψήφια διάσταση σε αλγορίθμους πρόβλεψης εξερχόμενων κλήσεων. Εντούτοις, αυτό είναι το πρώτο εγχείρημα για την αξιολόγηση της επίπτωσης της τοποθεσίας στο συγκεκριμένο πρόβλημα και γι' αυτό τον λόγο πρόκειται για μια από τις κυριότερες συνεισφορές της παρούσας διδακτορικής έρευνας.

Όσον αφορά τη φυσική τοποθεσία, από τα πειραματικά αποτελέσματα προέκυψε ότι η υιοθέτησή της για την πρόβλεψη εξερχόμενων κλήσεων δεν οδηγεί απαραίτητα σε βελτίωση της ακρίβειας πρόβλεψης, ενώ για κάποιες περιπτώσεις (5 θέσεις πρόβλεψης, χαμηλής και μέτριας απόκλισης επικοινωνίας τύποι χρηστών) η στατιστική ανάλυση έδειξε μια ελαφριά, αλλά στατιστικά σημαντική μείωση. Αν και τα αποτελέσματα ήταν αναμενόμενα για χρήστες που χαρακτηρίζονται από παρόμοια πρότυπα επικοινωνίας σε όλες τις τοποθεσίες που επισκέπτονται, αποτέλεσε έκπληξη η μη βελτίωση της ακρίβειας (ή η ελάχιστη βελτίωση χωρίς όμως αυτή η διαφορά να είναι στατιστικά σημαντική) για χρήστες που φαίνεται να αλλάζουν τη συμπεριφορά τους καθώς μετακινούνται μεταξύ διαφορετικών περιοχών. Αναλύθηκαν τα πρότυπα

επικοινωνίας για μερικούς τυπικούς χρήστες που ανήκουν σε αυτή την κατηγορία. Βρέθηκε ότι υπήρχαν χρήστες για τους οποίους οι επικοινωνίες που πραγματοποιήθηκαν στις γεωγραφικές τοποθεσίες όπου εμφανίζεται αποκλίνουσα συμπεριφορά αντιπροσώπευαν ένα μικρό μόνο ποσοστό της συνολικής επικοινωνίας και συνεπώς πιθανές βελτιώσεις στην ακρίβεια πρόβλεψης για κλήσεις από αυτές τις τοποθεσίες δεν επηρεάζουν σημαντικά τη συνολική ακρίβεια. Άλλοι χρήστες βρέθηκαν να αλληλεπιδρούν κυρίως με ένα μικρό πλήθος από επαφές, που είναι εύκολο να προβλεφθούν ανεξάρτητα από την προσθήκη της τοποθεσίας. Σε αυτή την περίπτωση, οι αλληλεπιδράσεις με άλλες επαφές είναι σπάνιες, επομένως ο αλγόριθμος αποτυγχάνει να προβλέψει αυτές τις κλήσεις. Το εύρημα αυτό συμφωνεί με τα αποτελέσματα που παρουσιάστηκαν στο (Bentley & Chen, 2015), όπου βρέθηκε ότι το 84% των κλήσεων ανταλλάσσεται με τις 5 κορυφαίες επαφές από άποψη συχνότητας, ενώ παρόμοια συμπεράσματα προέκυψαν και στο (Stefanis, Plessas, Komninos & Garofalakis, 2012). Επιπροσθέτως, για την περίπτωση της λίστας πρόβλεψης με μήκος μιας θέσης, οι Wiese, Min, Hong και Zimmerman (2015) έχουν εισηγηθεί πως οι άνθρωποι με τους οποίους επικοινωνεί κανείς πιο συχνά, είναι αυτοί με τον υψηλότερο βαθμό διασύνδεσης και επομένως είναι λογικό για τις αντίστοιχες επαφές να χρησιμοποιούνται ανεξάρτητα από την τοποθεσία.

Μια προσεχτική παρατήρηση των αρχείων κλήσεων των χρηστών και μια βηματική εκτέλεση του αλγορίθμου πρόβλεψης, αποκάλυψε ότι υπήρχαν συχνά αποτυχίες όταν οι χρήστες κινούνταν από τη συστάδα που είχαν το μεγαλύτερο ποσοστό επικοινωνίας σε άλλες τοποθεσίες, καθώς τα σύνολα εκπαίδευσης που χρησιμοποιούνταν σε αυτές τις «δευτερεύουσες» συστάδες αποτύγχαναν να αναπαραστήσουν τις πρόσφατες τάσεις της επικοινωνίας.

Φυσικά, τα συμπεράσματα από την παραπάνω ανάλυση δεν μπορούν να γενικευθούν ανεπιφύλακτα. Παρότι αυτή τη στιγμή δεν υπάρχει δημόσια διαθέσιμο κάποιο πιο λεπτομερές σύνολο δεδομένων στο οποίο να δοκιμαστούν οι αλγόριθμοι πρόβλεψης, είναι πιθανό το σύνολο δεδομένων που χρησιμοποιήθηκε να μην περιλαμβάνει χρήστες που αντιστοιχούν στα σενάρια που περιγράφηκαν στην ενότητα 5.2. Είναι πιθανό ότι η φυσική τοποθεσία παίζει σημαντικό ρόλο στην επικοινωνιακή συμπεριφορά ανθρώπων που ο τρόπος ζωής τους περιλαμβάνει συχνά ταξίδια και επικοινωνία με περισσότερες επαφές σε σχέση με τον μέσο χρήστη. Πρέπει να τονιστεί σε αυτό το σημείο ότι τα αποτελέσματα είναι ενθαρρυντικά για κάποιους από τους χρήστες που ανήκουν στην κατηγορία υψηλής απόκλισης της επικοινωνίας, η οποία αριθμεί μόλις 6 μέλη, ένας αριθμός ενδεχομένως πολύ μικρός για να οδηγήσει σε στατιστικά ασφαλή συμπεράσματα. Εντούτοις, είναι σημαντικό ότι αυτή η

ανάλυση υποδεικνύει ότι η φυσική τοποθεσία ίσως δεν είναι τόσο χρήσιμη διάσταση πλαισίου για την πρόβλεψη εξερχόμενων κλήσεων και κατ' επέκταση για την αξιοποίηση των επαφών για την πλειοψηφία των χρηστών συσκευών κινητού υπολογισμού, όσο οι περισσότεροι ερευνητές στην επιστημονική βιβλιογραφία υποστηρίζουν.

Παρόμοια ήταν τα ευρήματα και στην περίπτωση της μελέτης της επίπτωσης της σημασιολογικής τοποθεσίας. Οι διαφορές που προέκυψαν από την εισαγωγή στον αλγόριθμο πρόβλεψης αυτής της διάστασης ήταν συνολικά, αλλά και για την πλειοψηφία των χρηστών του συνόλου δεδομένων οριακές (είτε θετικά είτε αρνητικά). Επιπλέον, οι διαφορές αυτές δεν ήταν στατιστικά σημαντικές. Και πάλι θα πρέπει να είναι κανείς επιφυλακτικός με τα αποτελέσματα, καθώς το δείγμα των υπό ανάλυση χρηστών ήταν μικρό (μόλις 24 χρήστες), ενώ και το πλήθος των σημασιολογικών τοποθεσιών περιελάμβανε μόνο τις ετικέτες «σπίτι», «δουλειά» και «αλλού». Προφανώς, στη ζωή των ανθρώπων υπάρχει ένα πολύ μεγαλύτερο πλήθος από σημασιολογικές ετικέτες που επηρεάζουν τη συμπεριφορά τους και πολλές φορές σχετίζονται και με τη δραστηριότητα που πραγματοποιούν (π.χ. με φίλους, για διασκέδαση, στον κινηματογράφο, σε ταξίδι κ.α.).

Τέλος, όσον αφορά το χρονικό πλαίσιο (temporal context) που υιοθετήθηκε με τη μορφή της διάστασης της ώρας της ημέρας (time of day) και όπως αναφέρθηκε σχετίζεται έμμεσα με την τοποθεσία, αφού μπορεί να χρησιμοποιηθεί για τον εντοπισμό προτύπων κίνησης και σημασιολογικής πληροφορίας για την τοποθεσία, τα αποτελέσματα δεν ήταν και πάλι ενθαρρυντικά, καθώς η ακρίβεια πρόβλεψης μειώθηκε ελαφρώς. Συνδυάζοντας αυτό το συμπέρασμα με όσα ειπώθηκαν νωρίτερα για τη φυσική και τη σημασιολογική τοποθεσία, γίνεται εμφανές ότι η προσθήκη διαστάσεων πλαισίου σε αλγορίθμους πρόβλεψης δεν είναι μια απλή διαδικασία. Η δουλειά που παρουσιάστηκε υπονοεί ότι τα διανυσματικά μοντέλα αναπαράστασης του πλαισίου απαιτούν μια λεπτομερή εξέταση της επίπτωσης των διανυσματικών διαστάσεων, αντίθετα με τη συνήθη προσέγγιση που ακολουθείται στην επιστημονική βιβλιογραφία, σύμφωνα με την οποία ενσωματώνονται στα μοντέλα πρόβλεψης διαστάσεις που φαίνεται λογικό να επηρεάζουν την επικοινωνία. Αυτός είναι και ο λόγος που στο προηγούμενο κεφάλαιο προτάθηκαν ως διαστάσεις του αλγόριθμου πρόβλεψης η συχνότητα και η προσφατότητα της επικοινωνίας, οι οποίες είχαν μελετηθεί εκτενώς στο κεφάλαιο 3. Ένα παρόμοιο συμπέρασμα που επιβεβαιώνει αυτό το επιχείρημα μπορεί να εξαχθεί από το (Lee, Seo & Lee, 2010), όπου αν και δεν υπογραμμίζεται ευθέως από τους συγγραφείς, αναφέρεται πως δεν είχαν όλες οι

διαστάσεις του πλαισίου που χρησιμοποιήθηκαν για την πρόβλεψη την ίδια συνεισφορά στα αποτελέσματα σε κάθε περίπτωση.

5.6 Σύνοψη

Η τοποθεσία είναι μια διάσταση πλαισίου που πολύ συχνά προτείνεται ως υποψήφια ή και ενσωματώνεται στα μοντέλα πρόβλεψης χρησιμοποίησης της προσωπικής πληροφορίας και ιδιαίτερα για την αξιοποίηση των επαφών μέσω της πρόβλεψης κλήσεων. Σε αυτό το κεφάλαιο, μελετήθηκε η επίπτωση της φυσικής και της σημασιολογικής τοποθεσίας, καθώς και του χρονικού πλαισίου που σχετίζεται έμμεσα με αυτή. Πρόκειται για την πρώτη αντίστοιχη προσπάθεια στην επιστημονική βιβλιογραφία, όσον αφορά το πρόβλημα των επαφών και αυτή είναι μια σημαντική συνεισφορά της διδακτορικής έρευνας.

Η ανάλυση έδειξε ότι η τοποθεσία, σε όλες τις μορφές της (φυσική, σημασιολογική και έμμεσα με τη μορφή του χρονικού πλαισίου) ίσως να μην είναι τόσο σημαντική διάσταση όσο προτείνεται στη βιβλιογραφία για την πλειοψηφία των χρηστών, όμως η έλλειψη δημόσια διαθέσιμων συνόλων δεδομένων για την περαιτέρω μελέτη του προβλήματος δεν επιτρέπει τη γενίκευση των αποτελεσμάτων.

6. Επίλογος

Όπως εξηγήθηκε στην εισαγωγή, η διδακτορική έρευνα που παρουσιάζεται σε αυτή τη διατριβή έδωσε έμφαση από τα στάδια του κύκλου ζωής της διαχείρισης προσωπικής πληροφορίας, ειδικά για τις συσκευές κινητού υπολογισμού, στην αξιοποίηση. Οι λόγοι γι' αυτή την επιλογή έχουν ήδη αναλυθεί:

- Οι χρήστες προτιμούν να αποθηκεύουν όλη την προσωπική πληροφορία και είναι διστακτικοί όσον αφορά τη διαγραφή της, μια διαδικασία που απαιτεί χρόνο και κόπο.
- Οι συσκευές έχουν περιορισμούς (π.χ. μέγεθος οθόνης) που δυσκολεύουν την ανάκτηση της πληροφορίας.
- Οι ανάγκες των κινητών χρηστών απαιτούν την παροχή της σωστής πληροφορίας την κατάλληλη στιγμή.
- Το κινητό πλαίσιο είναι ένας σημαντικός παράγοντας που διαφοροποιεί τις απαιτήσεις για αξιοποίηση της προσωπικής πληροφορίας σε σύγκριση με τη χρήση σταθερών συσκευών, ενώ ταυτόχρονα δεν αφήνει χρόνο για την οργάνωση της.

Το είδος της προσωπικής πληροφορίας στο οποίο επικεντρώθηκε η έρευνα ήταν οι επαφές, καθώς οι συσκευές κινητού υπολογισμού και κυρίως τα έξυπνα κινητά τηλέφωνα χρησιμοποιούνται πρωτίστως για την εξυπηρέτηση της επικοινωνίας. Επιπλέον, το αποθετήριο των επαφών περιέχει μερικές εκατοντάδες έως χιλιάδες εγγραφές, οι περισσότερες από τις οποίες δε χρησιμοποιούνται, εντούτοις δυσχεραίνουν την ανάκτηση των υπολοίπων.

Το διανυσματικό μοντέλο που παρουσιάστηκε μπορεί εύκολα να επεκταθεί και να εφαρμοστεί σε οποιοδήποτε είδος προσωπικής πληροφορίας. Μάλιστα, υπάρχουν αναφορές στη βιβλιογραφία όπου αντίστοιχα διανυσματικά μοντέλα μπορούν να εφαρμοστούν εκτός από τις επαφές και σε άλλα στοιχεία προσωπικής πληροφορίας. Για παράδειγμα, οι Barzaiq και Loke (2011) πρότειναν την εφαρμογή του αλγορίθμου πρόβλεψης στα μηνύματα και στα τραγούδια που αποθηκεύονται στις συσκευές κινητού υπολογισμού, ενώ οι Sun, Wang, Zheng & Hsu (2013) εφάρμοσαν την προσέγγισή τους και στην περίπτωση των εφαρμογών που είναι εγκατεστημένες στις συσκευές, για την πρόβλεψη εκείνης που θα χρειαστεί να ανοίξει ο χρήστης. Η μεθοδολογία που παρουσιάστηκε μπορεί να προσαρμοστεί σε οποιοδήποτε τύπο προσωπικής πληροφορίας ως εξής:

1. Εντοπισμός των διαστάσεων του κινητού κοινωνικού πλαισίου που σχετίζονται με την ανάκτηση του συγκεκριμένου τύπου προσωπικής πληροφορίας. Σε αυτή την περίπτωση, είναι χρήσιμη η ανάλυση δεδομένων από τη χρήση της προσωπικής πληροφορίας σε συσκευές κινητού υπολογισμού. Το πλαίσιο για κάθε στοιχείο προσωπικής πληροφορίας x_i ορίζεται ως

$$x_i = (d_1, d_2, \dots, d_n)$$

2. Για κάθε διάσταση που ταυτοποιείται ως σημαντική ορίζεται μια συνάρτηση σκορ Sd_j και αποδίδεται ένα βάρος w_j . Οι επιμέρους συναρτήσεις (κανονικοποιημένες) συνδυάζονται σε ένα συνολικό σκορ ως εξής:

$$S = \sum_{j=1}^n w_j \times Sd_j$$

$$\sum_{j=1}^n w_j = 1$$

3. Τα στοιχεία προσωπικής πληροφορίας ταξινομούνται με βάση το συνολικό σκορ και παρουσιάζονται τα κορυφαία k στον χρήστη, οπότε η ανάκτηση πραγματοποιείται σε ένα πολύ μικρότερο πλήθος αντικειμένων.
4. Η μέθοδος αξιολογείται αρχικά μέσω προσομοίωσης και εν συνεχεία στο πεδίο.

Στην εισαγωγή τέθηκαν τα ερευνητικά ερωτήματα που είχε σκοπό να απαντήσει η διδακτορική έρευνα. Στη συνέχεια, ακολουθούν συνοπτικά τα συμπεράσματα που προέκυψαν από τη μελέτη κάθε ερωτήματος.

E1: Ποιο μέρος του πλαισίου των συσκευών κινητού υπολογισμού σχετίζεται με την ανάκτηση επαφών από τους χρήστες για την έναρξη επικοινωνίας;

Ένα πείραμα μελέτης της σύνθεσης του αποθετηρίου επαφών και των προτύπων επικοινωνίας των χρηστών πραγματοποιήθηκε για την απάντηση αυτού του ερωτήματος. Από το πείραμα προέκυψαν αποτελέσματα που επιβεβαίωσαν το μεγάλο σώμα στο οποίο χρειάζεται να πραγματοποιηθεί η ανάκτηση, όπως και το μεγάλο πλήθος των μη χρησιμοποιούμενων επαφών. Η επικοινωνία των χρηστών φάνηκε να ακολουθεί μια μορφή της αρχής Pareto, καθώς το μεγαλύτερο μέρος της επικοινωνίας βρέθηκε να πραγματοποιείται με ένα μικρό αριθμό επαφών. Ακόμα ένα ενδιαφέρον συμπέρασμα που προέκυψε ήταν πως οι υπάρχουσες μέθοδοι ανάκτησης δεν είναι οι

βέλτιστες, αλλά υπάρχει περιθώριο βελτίωσης. Η ανάλυση οδήγησε σε ευρήματα που συμφωνούν με αντίστοιχα, άλλων ερευνητών.

Τελικά, ταυτοποιήθηκαν ως σημαντικές διαστάσεις του κινητού κοινωνικού πλαισίου για την ανάκτηση των επαφών η συχνότητα και η προσφατότητα της επικοινωνίας. Για κάποιους χρήστες βρέθηκε ότι και η προσωπική προτίμηση, όπως εκφράζεται με το χαρακτηρισμό επαφών ως αγαπημένες, παίζει σημαντικό ρόλο στην εξερχόμενη επικοινωνία με τις επαφές. Επειδή, όμως, ο αριθμός τους ήταν σχετικά μικρός, θεωρήθηκε διφορούμενη διάσταση και δε χρησιμοποιήθηκε στο μοντέλο που προτάθηκε. Δυστυχώς, στα υπόλοιπα σύνολα δεδομένων που αναλύθηκαν κατά καιρούς, δεν υπήρχε διαθέσιμη αυτή η πληροφορία. Σε επόμενη φάση μελετήθηκαν, επίσης, το χρονικό πλαίσιο και η τοποθεσία, χωρίς όμως να προκύψει βελτίωση του μοντέλου από την ενσωμάτωσή τους.

E2: Πώς μπορεί το πλαίσιο αυτό να μοντελοποιηθεί και να αξιοποιηθεί για τη βελτίωση της πρόσβασης στις επαφές;

Για την αξιοποίηση του πλαισίου προτάθηκε ένα διανυσματικό μοντέλο αναπαράστασης του. Το μοντέλο αυτό έχει το πλεονέκτημα της εύκολης επέκτασής του με επιπλέον διαστάσεις πλαισίου, ενώ λόγω της απλότητάς του είναι κατάλληλο για απευθείας εφαρμογή στις συσκευές κινητού υπολογισμού. Σε κάθε διάσταση αντιστοιχεί μια συνάρτηση σκορ. Για κάθε επαφή υπολογίζεται ένα συνολικό σκορ, από το συνδυασμό των επιμέρους συναρτήσεων, ακολουθώντας τη θεωρία της μίξης χαρακτηριστικών. Τελικά, παράγεται μια λίστα από τις top-k προτάσεις, οι οποίες τροφοδοτούν μια προσαρμοζόμενη λίστα, απ' όπου πραγματοποιείται η ανάκτηση. Το μοντέλο περιγράφεται αναλυτικά, για τη γενική περίπτωση που περιλαμβάνει οποιοδήποτε τύπο στοιχείων προσωπικής πληροφορίας, στις προηγούμενες παραγράφους.

E3: Ποια τα αποτελέσματα αξιολόγησης ενός συστήματος που αξιοποιεί το πλαίσιο για να διευκολύνει την πρόσβαση στις επαφές;

Το μοντέλο πρόβλεψης της χρήσης των επαφών αξιολογήθηκε τόσο μέσω της διαδικασίας προσομοίωσης στα υπάρχοντα σύνολα δεδομένων, όσο και στο πεδίο. Συγκεκριμένα, πραγματοποιήθηκαν ένα προκαταρκτικό πείραμα με ένα μικρό σύνολο δεδομένων και ένα κυρίως πείραμα με το πληρέστερο σύνολο που υπάρχει δημόσια διαθέσιμο. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη προσέγγιση μπορεί να εξυπηρετήσει καλύτερα τους χρήστες σε σχέση με τις δημοφιλέστερες μεθόδους ανάκτησης (λίστα πρόσφατων κλήσεων και λίστα συχνότερων επαφών). Επιπλέον,

αυτή η προσέγγιση φαίνεται να παρέχει καλύτερη απόδοση, όσον αφορά τις προβλέψεις, σε σχέση με άλλους αλγορίθμους που έχουν προταθεί στη βιβλιογραφία. Το συμπέρασμα αυτό προέκυψε τόσο από τη σύγκριση των αναφερόμενων ποσοστών επιτυχίας από τις προσομοιώσεις, όσο και με απευθείας σύγκριση με άλλους αλγορίθμους μετά από εφαρμογή στο ίδιο σύνολο δεδομένων. Το πείραμα πεδίου επιβεβαίωσε πλήρως τα συμπεράσματα που εξήχθησαν από τις πειραματικές προσομοιώσεις, ενώ αποκάλυψε και την αποδοχή της προτεινόμενης προσέγγισης από τους χρήστες.

E4: Ποια είναι η επίπτωση της τοποθεσίας ως παράγοντα πλαισίου για την πρόσβαση στις επαφές;

Η τοποθεσία είναι ένας παράγοντας πλαισίου που αναφέρεται συχνά στη βιβλιογραφία ως σημαντικός για τους χρήστες κινητών συσκευών. Όσον αφορά την πρόβλεψη εξερχόμενων κλήσεων, προτείνεται σε πολλές ερευνητικές δημοσιεύσεις ως υποψήφια διάσταση του πλαισίου, ενώ σε κάποιες ενσωματώνεται κιόλας στο προτεινόμενο μοντέλο. Στα πλαίσια της διδακτορικής έρευνας μελετήθηκε η επίπτωση που είχε στην πρόβλεψη κλήσεων η εισαγωγή του χωρικού πλαισίου με τρεις διαφορετικές μορφές: φυσική τοποθεσία, σημασιολογική τοποθεσία και χρονικό πλαίσιο (σχετίζεται έμμεσα με την τοποθεσία). Η ανάλυση έδειξε πως, τουλάχιστον για τους υπό εξέταση χρήστες, η χρησιμοποίηση της τοποθεσίας δε βελτίωσε απαραίτητα την ακρίβεια των προβλέψεων. Αν και είναι πιθανό να μην υπήρχαν στο σύνολο δεδομένων που αναλύθηκε χρήστες με πρότυπα επικοινωνίας που να εξαρτώνται από την τοποθεσία, αυτό είναι μια ένδειξη πως η τοποθεσία ίσως να μην είναι για τους περισσότερους χρήστες τόσο σημαντικός παράγοντας πλαισίου όσο αναφέρεται στη βιβλιογραφία, όσον αφορά την εξερχόμενη επικοινωνία.

Ο συγγραφέας θεωρεί πως η εκπόνηση της διδακτορικής έρευνας οδήγησε σε μια πολύπλευρη συνεισφορά στην ερευνητική κοινότητα. Συγκεκριμένα:

- Η προσέγγιση που παρουσιάστηκε ήταν η πρώτη όπου οι διαστάσεις του πλαισίου αναλύθηκαν ενδελεχώς πριν την ενσωμάτωσή τους στον αλγόριθμο πρόβλεψης για την αξιοποίηση των επαφών. Σε όλες τις υπόλοιπες περιπτώσεις που αναφέρονται στη βιβλιογραφία, οι διαστάσεις επιλέγονταν χωρίς κάποια αιτιολόγηση ή μόνο με βάση λογικές παρατηρήσεις. Η μοναδική εξαίρεση είναι η εργασία των Sun, Wang, Zheng και Hsu (2013), όπου οι ερευνητές βασίστηκαν στα αποτελέσματα του πειράματος μελέτης της σύνθεσης του αποθετηρίου και των προτύπων επικοινωνίας που παρουσιάστηκε σε αυτή τη διατριβή. Επιπλέον, το συγκεκριμένο πείραμα

ήταν ίσως το πιο αναλυτικό σε αυτόν τον τομέα, μέχρι το αντίστοιχο των Bentley και Chen (2015).

- Ο αλγόριθμος πρόβλεψης εξερχόμενων κλήσεων που προτάθηκε παρουσιάζει την υψηλότερη ακρίβεια σε σχέση με τους αντίστοιχους αλγορίθμους που περιγράφονται στη βιβλιογραφία, ενώ ταυτόχρονα είναι ιδιαίτερα απλός και εύκολα επεκτάσιμος.
- Εξετάστηκε αναλυτικά η επίπτωση που έχουν οι διαφοροποιήσεις στον τρόπο επικοινωνίας των χρηστών, ανάλογα με την «κοινωνικότητά» τους. Μάλιστα, για πρώτη φορά ποσοτικοποιήθηκε η έννοια αυτή επιτυχώς με την εισαγωγή της K-Μετρικής, ενώ τα πειραματικά αποτελέσματα επιβεβαίωσαν την εγκυρότητα αυτής της προσέγγισης.
- Πραγματοποιήθηκε πείραμα πεδίου για την αξιολόγηση του αλγορίθμου πρόβλεψης. Στη βιβλιογραφία υπάρχει μόνο άλλη μια περίπτωση αξιολόγησης αλγορίθμου πρόβλεψης στο πεδίο, με λιγότερους χρήστες και με σημαντικά μικρότερο πλήθος από κλήσεις.
- Για πρώτη φορά μελετήθηκε διεξοδικά η επίπτωση των διαστάσεων που σχετίζονται με το χωρικό πλαίσιο (χρονικό πλαίσιο, φυσική τοποθεσία, σημασιολογική τοποθεσία) στην ακρίβεια των προβλέψεων. Καθώς τα αποτελέσματα δεν ήταν τα αναμενόμενα (δεν παρατηρήθηκε βελτίωση για την πλειοψηφία των χρηστών), τονίστηκε η ανάγκη εξέτασης των υποψηφίων διαστάσεων πριν από την ενσωμάτωσή τους σε μοντέλα πρόβλεψης για τη βελτίωση της ανάκτησης.

Κατά την εκπόνηση της διδακτορικής έρευνας προέκυψαν διάφορα ανοιχτά ζητήματα που θα μπορούσαν να αποτελέσουν μελλοντικές επεκτάσεις. Αρχικά, όπως έχει ήδη αναφερθεί, υπάρχουν τόσο ενδείξεις όσο και η πεποίθηση του συγγραφέα πως το πλαίσιο εκείνο που επηρεάζει τη συμπεριφορά των χρηστών είναι μεταβαλλόμενο. Αυτό σημαίνει πως τα βάρη των διαστάσεων του πλαισίου ενδεχομένως πρέπει να είναι δυναμικά όχι μόνο για διαφορετικούς χρήστες, αλλά ακόμα και για τον ίδιο χρήστη κάτω από διαφορετικές συνθήκες. Πρόκειται για μια παρατήρηση που αξίζει να διερευνηθεί περαιτέρω. Επιπλέον, από τα αποτελέσματα φάνηκε πως για την ομάδα με τους περισσότερους «κοινωνικούς» χρήστες υπάρχει σημαντικό περιθώριο βελτίωσης για την ακρίβεια των προβλέψεων. Αυτό θα μπορούσε να επιτευχθεί με τη διερεύνηση και άλλων διαστάσεων που δεν υπήρχαν διαθέσιμες, όπως οι προγραμματισμένες εργασίες, η δραστηριότητα του χρήστη ή ακόμα και με την ενσωμάτωση και άλλων καναλιών επικοινωνίας (π.χ. μηνύματα, emails, αλληλεπιδράσεις σε κοινωνικά δίκτυα κ.λπ.) στις υπάρχουσες διαστάσεις της

συχνότητας και της προσφατότητας. Μάλιστα, οι Wiese, Min, Hong και Zimmerman (2015) έχουν τονίσει το γεγονός πως οι άνθρωποι χρησιμοποιούν πολλά διαφορετικά κανάλια επικοινωνίας και επομένως τα αρχεία κλήσεων και μηνυμάτων δεν είναι αντιπροσωπευτικά των προτύπων επικοινωνίας. Ακόμα, είναι σημαντικό να μελετηθεί περισσότερο η επίπτωση της τοποθεσίας, ειδικά για χρήστες που παρουσιάζουν συχνή κινητικότητα και ταυτόχρονα ένα «κοινωνικό» πρότυπο επικοινωνίας. Δυστυχώς, φαίνεται ότι τέτοιοι χρήστες δεν περιλαμβάνονταν στα διαθέσιμα σύνολα δεδομένων που αναλύθηκαν. Για όλα τα παραπάνω θα μπορούσε να αξιοποιηθεί η δεξαμενή χρηστών που έχουν εγκαταστήσει και χρησιμοποιούν την εφαρμογή Calchas. Τέλος, το μοντέλο που παρουσιάστηκε θα μπορούσε να εφαρμοστεί και σε άλλους τύπους στοιχείων προσωπικής πληροφορίας ή και σε συνδυασμούς αυτών, οδηγώντας στο σχεδιασμό νέων διεπαφών ανάκτησης για τις συσκευές κινητού υπολογισμού.

Βιβλιογραφικές Αναφορές

- Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. (1999, January). Towards a better understanding of context and context-awareness. In *Handheld and ubiquitous computing* (pp. 304-307). Springer Berlin Heidelberg.
- Absar, R., O'Brien, H., & Webster, E. T. (2014). Exploring social context in mobile information behavior. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-10.
- Allan, J., Croft, B., Moffat, A., & Sanderson, M. (2012, May). Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. In *ACM SIGIR Forum* (Vol. 46, No. 1, pp. 2-32). ACM.
- Ancona, M., Doderer, G., & Gianuzzi, V. (1999, February). Mobile computing in outdoor environments. In *Proceedings of the 1999 ACM symposium on Applied computing* (pp. 413-415). ACM.
- Anguera, X., Xu, J., & Oliver, N. (2008, October). Multimodal photo annotation and retrieval on a mobile phone. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval* (pp. 188-194). ACM.
- Ankolekar, A., Szabo, G., Luon, Y., Huberman, B. A., Wilkinson, D., & Wu, F. (2009, September). Friendlee: a mobile application for your social life. In *Proceedings of the 11th international Conference on Human-Computer interaction with Mobile Devices and Services* (p. 27). ACM.
- Arminen, I. (2006). Social functions of location in mobile telephony. *Personal and Ubiquitous Computing*, 10(5), 319-323.
- Aula, A., Jhaveri, N., & Käki, M. (2005, May). Information search and re-access strategies of experienced web users. In *Proceedings of the 14th international conference on World Wide Web* (pp. 583-592). ACM.
- Baltrunas, L., Ludwig, B., Peer, S., & Ricci, F. (2011). Context-aware places of interest recommendations for mobile users. In *Design, User Experience, and Usability. Theory, Methods, Tools and Practice* (pp. 531-540). Springer Berlin Heidelberg.
- Barkhuus, L., Brown, B., Bell, M., Sherwood, S., Hall, M., & Chalmers, M. (2008, April). From awareness to repartee: sharing location within social groups. In *proceedings of the SIGCHI conference on human factors in computing systems* (pp. 497-506). ACM.

Barkhuus, L. (2003). Context information in mobile telephony. In *Human-Computer Interaction with Mobile Devices and Services* (pp. 451-455). Springer Berlin Heidelberg.

Barreau, D. K. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, 46(5), 327-339.

Barzaiq, O. O., & Loke, S. W. (2011). Adapting the mobile phone for task efficiency: the case of predicting outgoing calls using frequency and regularity of historical calls. *Personal and Ubiquitous Computing*, 15(8), 857-870.

Beach, A., Gartrell, M., Xing, X., Han, R., Lv, Q., Mishra, S., & Seada, K. (2010, February). Fusing mobile, sensor, and social data to fully enable context-aware computing. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications* (pp. 60-65). ACM.

Bellman, R., 1957. *Dynamic Programming*, Princeton University Press. Princeton, NJ, USA.

Bentley, F., & Chen, Y. Y. (2015, April). The Composition and Use of Modern Mobile Phonebooks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2749-2758). ACM.

Bentley, F. R., Kames, J., Ahmed, R., Zivin, R. S., & Schwendimann, L. (2010, April). Contacts 3.0: Bringing together research and design teams to reinvent the phonebook. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 4677-4690). ACM.

Bergman, O., Beyth Marom, R., & Nachmias, R. (2003). The user subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology*, 54(9), 872-878.

Bergman, O., Beyth Marom, R., & Nachmias, R. (2008). The user subjective approach to personal information management systems design: evidence and implementations. *Journal of the American Society for Information Science and Technology*, 59(2), 235-246.

Bergman, O., Komninos, A., Liarokapis, D., & Clarke, J. (2012). You never call: Demoting unused contacts on mobile phones using DMTR. *Personal and Ubiquitous Computing*, 16(6), 757-766.

Bergman, O., Tucker, S., Beyth-Marom, R., Cutrell, E., & Whittaker, S. (2009, April). It's not that important: demoting personal information of low subjective importance using GrayArea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 269-278). ACM.

Bergman, O., Whittaker, S., Sanderson, M., Nachmias, R., & Ramamoorthy, A. (2010). The effect of folder structure on personal file navigation. *Journal of the American Society for Information Science and Technology*, 61(12), 2426-2441.

Bernstein, M., Van Kleek, M., Schraefel, M., & Karger, D. R. (2008). Evolution and evaluation of an information scrap manager.

Bettini, C., Brdiczka, O., Henriksen, K., Indulska, J., Nicklas, D., Ranganathan, A., & Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, 6(2), 161-180.

Bhamidipaty, A., & Deepak, P. (2007). SymAB: symbol-based address book for the semi-literate mobile user. In *Human-Computer Interaction-INTERACT 2007* (pp. 389-392). Springer Berlin Heidelberg.

Bicocchi, N., Castelli, G., Mamei, M., Rosi, A., & Zambonelli, F. (2008, February). Supporting location-aware services for mobile users with the whereabouts diary. In *Proceedings of the 1st international conference on MOBILE Wireless MiddleWARE, Operating Systems, and Applications* (p. 6). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Boardman, R. (2004). *Improving tool support for personal information management* (Doctoral dissertation, Imperial College).

Boardman, R., & Sasse, M. A. (2004, April). Stuff goes into the computer and doesn't come out: a cross-tool study of personal information management. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 583-590). ACM.

Böcker, M., & Suwita, A. (1999). Evaluating the Siemens C10 mobile phone: going beyond "quick and dirty" usability testing. *Proceedings of HFT'99*.

Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011, August). Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services* (pp. 47-56). ACM.

Bridle, R., & McCreath, E. (2006, January). Inducing shortcuts on a mobile phone interface. In *Proceedings of the 11th international conference on Intelligent user interfaces* (pp. 327-329). ACM.

Brown, B., & Randell, R. (2004). Building a context sensitive telephone: some hopes and pitfalls for context sensitive computing. *Computer Supported Cooperative Work (CSCW)*, 13(3-4), 329-345.

Butler, J., Morrice, D. J., & Mullarkey, P. W. (2001). A multiple attribute utility theory approach to ranking and selection. *Management Science*, 47(6), 800-816.

- Calabrese, F., Smoreda, Z., Blondel, V. D., & Ratti, C. (2011). Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PloS one*, 6(7).
- Cao, X., Cong, G., & Jensen, C. S. (2010). Mining significant semantic locations from GPS data. *Proceedings of the VLDB Endowment*, 3(1-2), 1009-1020.
- Chamasemani, F. F., & Affendey, L. S. (2014). Impact of mobile context-aware applications on human computer interaction. *Journal of Theoretical & Applied Information Technology*, 62(1), 281-287.
- Chande, S., Vartiainen, P., & Rämö, K. (2007, September). Active notes: context-sensitive notes for mobile devices. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology* (pp. 716-723). ACM.
- Chen, G., & Kotz, D. (2000). *A survey of context-aware mobile computing research* (Vol. 1, No. 2.1, pp. 2-1). Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College.
- Chihani, B., Bertin, E., Jeanne, F., & Crespi, N. (2011). Context-aware systems: a case study. In *Digital Information and Communication Technology and Its Applications* (pp. 718-732). Springer Berlin Heidelberg.
- Chihani, B., Bertin, E., Jeanne, F., & Crespi, N. (2011). HEP: context-aware communication system. *International Journal of New Computer Architectures and their Applications (IJNCAA)*, 1(1), 15-24.
- Church, K., & Smyth, B. (2009, February). Understanding the intent behind mobile information needs. In *Proceedings of the 14th international conference on Intelligent user interfaces* (pp. 247-256). ACM.
- Cui, Y., & Honkala, M. (2013). A novel mobile device user interface with integrated social networking services. *International Journal of Human-Computer Studies*, 71(9), 919-932.
- Dai, L., Lutters, W. G., & Bower, C. (2005, April). Why use memo for all?: restructuring mobile applications to support informal note taking. In *CHI'05 extended abstracts on Human factors in computing systems* (pp. 1320-1323). ACM.
- Dai, P., & Ho, S. S. (2014, July). A smartphone user activity prediction framework utilizing partial repetitive and landmark behaviors. In *Mobile Data Management (MDM), 2014 IEEE 15th International Conference on* (Vol. 1, pp. 205-210). IEEE.

- Dai, P., Yang, W., & Ho, S. S. (2013). Predicting mobile call behavior via subspace methods. In *Social Computing, Behavioral-Cultural Modeling and Prediction* (pp. 466-475). Springer Berlin Heidelberg.
- Davis, M., Van House, N., Towle, J., King, S., Ahern, S., Burgener, C., ... & Rothenberg, M. (2005, April). MMM2: mobile media metadata for media sharing. In *CHI'05 extended abstracts on Human factors in computing systems* (pp. 1335-1338). ACM.
- Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, 5(1), 4-7.
- Dey, A. K., & Häkkinen, J. (2008). Context-awareness and mobile devices. *User interface design and evaluation for mobile technology*, 1, 205-217.
- Du, W., & Wang, L. (2008, May). Context-aware application programming for mobile devices. In *Proceedings of the 2008 C³S²E conference* (pp. 215-227). ACM.
- Eagle, N., Clauset, A., & Quinn, J. A. (2009). Location Segmentation, Inference and Prediction for Anticipatory Computing. In *AAAI Spring Symposium: Technosocial Predictive Analytics* (pp. 20-25).
- Eagle, N., & Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4), 255-268.
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD)* (Vol. 96, No. 34, pp. 226-231).
- Falke, E. (2008). *The Associative PDA 2.0: A Prototype of an Innovative Mobile System for Personal Information Management*. VDM Verlag.
- Faloutsos, C., & Lin, K. I. (1995). FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *Proceedings of ACM SIGMOD '95* (pp. 163-174). ACM.
- Ferscha, A., Hechinger, M., Riener, A., Schmitzberger, H., Franz, M., Rocha, M., & Zeidler, A. (2006, July). Context-aware profiles. In *Autonomic and Autonomous Systems, 2006. ICAS'06. 2006 International Conference on* (pp. 48-48). IEEE.
- Fisher, D., Brush, A. J., Gleave, E., & Smith, M. A. (2006, November). Revisiting Whittaker & Sidner's email overload ten years later. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 309-312). ACM.
- Fodor, I. K., 2002. A survey of dimension reduction techniques. *LLNL Technical Report*, UCRL-ID-148494.

- Gilbert, P., Cuervo, E., & Cox, L. P. (2009, February). Experimenting in mobile social contexts using JellyNets. In *Proceedings of the 10th workshop on Mobile Computing Systems and Applications* (p. 16). ACM.
- Gong, J., & Tarasewich, P. (2004, November). Guidelines for handheld mobile device interface design. In *Proceedings of DSI 2004 Annual Meeting* (pp. 3751-3756).
- Gray, P., & Salber, D. (2001). Modelling and using sensed context information in the design of interactive applications. In *Engineering for Human-Computer Interaction* (pp. 317-335). Springer Berlin Heidelberg.
- Gurrin, C., Jones, G. J., Lee, H., O'Hare, N., Smeaton, A. F., & Murphy, N. (2005, September). Mobile access to personal digital photograph archives. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services* (pp. 311-314). ACM.
- Hang, A., Von Zezschwitz, E., De Luca, A., & Hussmann, H. (2012, October). Too much information!: user attitudes towards smartphone sharing. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (pp. 284-287). ACM.
- Heimonen, T. (2009, September). Information needs and practices of active mobile internet users. In *Proceedings of the 6th International Conference on Mobile Technology, Application & Systems* (p. 50). ACM.
- Hofer, T., Schwinger, W., Pichler, M., Leonhartsberger, G., Altmann, J., & Retschitzegger, W. (2003, January). Context-awareness on mobile devices-the hydrogen approach. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on* (pp. 10-pp). IEEE.
- Hsu, D. F., Chung, Y. S., & Kristal, B. S. (2006). Combinatorial fusion analysis: methods and practice of combining multiple scoring systems. *Advanced Data Mining Technologies in Bioinformatics*, 32-62.
- Hsu, D. F., & Taksa, I. (2005). Comparing rank and score combination methods for data fusion in information retrieval. *Information retrieval*, 8(3), 449-480.
- Huvila, I., Eriksen, J., Häusner, E. M., & Jansson, I. M. (2014). Continuum thinking and the contexts of personal information management. *INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL*, 19(1).
- Jan, S., Li, M., & Al-Raweshidy, H. (2011). FARM: file annotation and retrieval on mobile devices. *Personal and Ubiquitous Computing*, 15(7), 771-779.

- Jones, W. (2004). Finders, keepers? The present and future perfect in support of personal information management. *First monday*, 9(3).
- Jones, W. (2007). Personal information management. *Annual review of information science and technology*, 41(1), 453-504.
- Jung, Y., Anttila, A., & Blom, J. (2008, September). Designing for the evolution of mobile contacts application. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 449-452). ACM.
- Kang, S., Won, H., Jeon, G., & Lee, Y. S. (2012). Call prediction model based on smartphone users behavior. In *Convergence and Hybrid Information Technology* (pp. 180-185). Springer Berlin Heidelberg.
- Karikoski, J., & Soikkeli, T. (2013). Contextual usage patterns in smartphone communication services. *Personal and ubiquitous computing*, 17(3), 491-502.
- Karypidis, A., & Lalis, S. (2007). Automated context aggregation and file annotation for PAN-based computing. *Personal and Ubiquitous Computing*, 11(1), 33-44.
- Kim, J., Lee, S., Won, J. S., & Moon, Y. S. (2011, December). Photo cube: An automatic management and search for photos using mobile smartphones. In *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on* (pp. 1228-1234). IEEE.
- Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D., & Laurila, J. (2010). Towards rich mobile phone datasets: Lausanne data collection campaign. In *Proceedings of the 7th Int. Conf. on Pervasive Services (ICPS), Berlin*.
- Klockar, T., Carr, D. A., Hedman, A., Johansson, T., & Bengtsson, F. (2003, December). Usability of mobile phones. In *Proceedings of the 19th international symposium on human factors in telecommunication* (pp. 197-204).
- Komninos, A., Baillie, L., & Barrie, P. (2008). Holistic PIM: Managing personal information for a nomadic generation. In *Conference of Human Factors in Computing Systems, CHI* (pp. 5-10).
- Komninos, A., Baillie, L., & Dunlop, M. (2007). Augmenting entry: the possibilities for utilizing geo-referenced information to improve mobile calendar applications. In *ACM CHI 2007 Workshop on Mobile Spatial Interaction*.
- Komninos, A., & Liarokapis, D. (2009, September). The use of mobile contact list applications and a context-oriented framework to support their design. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services* (p. 79). ACM.

- Komninos, A., Plessas, A., Stefanis, V., & Garofalakis, J. (2011a, September). Application of dimensionality reduction techniques for mobile social context. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 583-584). ACM.
- Komninos, A., Plessas, A., Stefanis, V., & Garofalakis, J. D. (2011b). Context Dimensionality Reduction for Mobile Personal Information Access. In *KDIR* (pp. 493-498).
- Korpipää, P., Häkkinen, J., Kela, J., Ronkainen, S., & Käsälä, I. (2004, October). Utilising context ontology in mobile device application personalisation. In *Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia* (pp. 133-140). ACM.
- Korpipää, P., Jani, M., Kela, J., & Malm, E. J. (2003). Managing context information in mobile devices. *IEEE pervasive computing*, (3), 42-51.
- Korpipää, P., & Mäntylä, J. (2003). An ontology for mobile device sensor-based context awareness. In *Modeling and Using Context* (pp. 451-458). Springer Berlin Heidelberg.
- Koshman, S. (2011, February). iSchool agenda: mobile context research and teaching. In *Proceedings of the 2011 iConference* (pp. 211-216). ACM.
- Kristoffersen, S., & Ljungberg, F. (1999, November). "Making place" to make IT work: empirical explorations of HCI for mobile CSCW. In Proceedings of the international ACM SIGGROUP conference on Supporting group work (pp. 276-285). ACM.
- Krumm, J., & Rouhana, D. (2013, September). Placer: semantic place labels from diary data. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (pp. 163-172). ACM.
- Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9), 140-150.
- Langley, P. (1999). User Modeling in Adaptive Interface. In *Proceedings of the seventh international conference on user modeling* (pp. 357-370). Springer Vienna.
- Lansdale, M. W. (1988). The psychology of personal information management. *Applied ergonomics*, 19(1), 55-66.
- LaRue, E. M., Mitchell, A. M., Terhorst, L., & Karimi, H. A. (2010). Assessing mobile phone communication utility preferences in a social support network. *Telematics and Informatics*, 27(4), 363-369.
- Laurila, J. K., Gatica-Perez, D., Aad, I., Bornet, O., Do, T. M. T., Dousse, O., ... & Miettinen, M. (2012). The mobile data challenge: Big data for mobile computing research. In *Pervasive Computing* (No. EPFL-CONF-192489).

- Le, A. S. S. H. V., Henze, N., & Schmidt, A. (2013, August). Automatic Classification of Mobile Phone Contacts. In *Mensch & Computer 2013–Tagungsband: 13. fachübergreifende Konferenz für interaktive und kooperative Medien* (p. 27). Walter de Gruyter.
- Lee, S., Seo, J., & Lee, G. (2010, April). An adaptive speed-call list algorithm and its evaluation with ESM. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2019-2022). ACM.
- Lehtiniemi, A. (2008, December). Evaluating SuperMusic: streaming context-aware mobile music service. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology* (pp. 314-321). ACM.
- Leino, J., Finnberg, S., & Rähkä, K. J. (2010, September). The times they are a-changin': mobile PIM is leaving the paper trail behind. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 259-268). British Computer Society.
- Liao, Z. X., Li, S. C., Peng, W. C., Yu, P. S., & Liu, T. C. (2013, December). On the feature discovery for app usage prediction in smartphones. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on* (pp. 1127-1132). IEEE.
- Lugano, G. (2008). Mobile social networking in theory and practice. *First Monday*, 13(11).
- Marshall, C. C. (2008a). Rethinking personal digital archiving, Part 1: Four challenges from the field. *D-Lib Magazine*, 14(3), 2.
- Marshall, C. C. (2008). Rethinking personal digital archiving, part 2: implications for services, applications, and institutions. *D-Lib Magazine*, 14(3), 3.
- Mikalsen, M., & Kofod-Petersen, A. (2004, September). Representing and reasoning about context in a mobile environment. In *Proceedings of the First International Workshop on Modeling and Retrieval of Context. CEUR Workshop Proceedings* (Vol. 114, pp. 25-35).
- Min, J. K., Kim, H. T., & Cho, S. B. (2008, December). Social and Personal Context Modeling for Contact List Recommendation on Mobile Device. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on* (Vol. 3, pp. 381-384). IEEE.
- Montoliu, R., Martínez-Uso, A., Martínez-Sotoca, J., & McInerney, J. (2012). Semantic place prediction by combining smart binary classifiers. In *Nokia Mobile Data Challenge 2012 Workshop. p. Dedicated task* (Vol. 1, No. 1).
- Moore, P., Hu, B., & Wan, J. (2010). Smart-context: A context ontology for pervasive mobile computing. *The Computer Journal*, 53(2), 191-207.

Myers, R., Zapata, E., & Singh, G. (2007, September). Linking information for mobile use. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology* (pp. 607-613). ACM.

Nguyen, T. V., & Oh, A. H. Y. (2010, September). Users' needs for social tagging and sharing on mobile contacts. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 387-388). ACM.

Oulasvirta, A., Raento, M., & Tiitta, S. (2005, September). ContextContacts: re-designing SmartPhone's contact book to support mobile awareness and collaboration. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services* (pp. 167-174). ACM.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005, April). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 919-928). ACM.

Page, T. (2013). Usability of text input interfaces in smartphones. *Journal of Design Research*, 11(1), 39-56.

Pareto, V. (1935). *The mind and society: Trattato di sociologia generale*. AMS Press.

Park, M. H., Hong, J. H., & Cho, S. B. (2007). Location-based recommendation system using bayesian user's preference model in mobile devices. In *Ubiquitous Intelligence and Computing* (pp. 1130-1139). Springer Berlin Heidelberg.

Pascoe, J., Ryan, N., & Morse, D. (2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(3), 417-437.

Patel, S. N., & Abowd, G. D. (2004). The contextcam: Automated point of capture video annotation. In *UbiComp 2004: Ubiquitous Computing* (pp. 301-318). Springer Berlin Heidelberg.

Phithakkitnukoon, S., & Dantu, R. (2008). Adequacy of data for characterizing caller behavior. In *Proceedings of KDD Inter. Workshop on Social Network Mining and Analysis (SNAKDD 2008)*.

Phithakkitnukoon, S., & Dantu, R. (2011). Towards ubiquitous computing with call prediction. *ACM SIGMOBILE Mobile Computing and Communications Review*, 15(1), 52-64.

Phithakkitnukoon, S., Dantu, R., Claxton, R., & Eagle, N. (2011). Behavior-based adaptive call predictor. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 6(3), 21.

- Plessas, A., Georgiadou, O., Stefanis, V., Komninos, A., & Garofalakis, J. (2015). Assessing Physical Location as a Potential Contextual Cue For Adaptive Mobile Contact Lists. In *Proceedings of 14th IEEE International Conference on Ubiquitous Computing and Communications* (IUCC 2015). IEEE.
- Plessas, A., Stefanis, V., Komninos, A., & Garofalakis, J. (2013). Using communication frequency and recency context to facilitate mobile contact list retrieval. *International Journal of Handheld Computing Research (IJHCR)*, 4(4), 52-71.
- Plessas, A., Stefanis, V., Komninos, A., & Garofalakis, J. (under revision). Field Evaluation of Context Aware Adaptive Interfaces for Efficient Mobile Contact Retrieval. *Pervasive and Mobile Computing*.
- Qin, C., Bao, X., Roy Choudhury, R., & Nelakuditi, S. (2011, June). Tagsense: a smartphone-based approach to automatic image tagging. In *Proceedings of the 9th international conference on Mobile systems, applications, and services* (pp. 1-14). ACM.
- Raento, M., Oulasvirta, A., Petit, R., & Toivonen, H. (2005). ContextPhone: A prototyping platform for context-aware mobile applications. *Pervasive Computing, IEEE*, 4(2), 51-59.
- Robbins, D. C. (2008). TapGlance: designing a unified smartphone interface for personal information management. In *CHI 2009 Conference on Human Factors and Computing Systems Florence, Italy*.
- Ryan, N. (1999). ConteXtML: Exchanging contextual information between a Mobile Client and the FieldNote Server. *Computing Laboratory, University of Kent at Canterbury*.
- Salehan, M., & Negahban, A. (2013). Social networking on smartphones: When mobile phones become addictive. *Computers in Human Behavior*, 29(6), 2632-2639.
- Sarvas, R., Herrarte, E., Wilhelm, A., & Davis, M. (2004, June). Metadata creation system for mobile images. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services* (pp. 36-48). ACM.
- Schmidt, A., & Gellersen, H. W. (2001). Context-aware mobile telephony. *ACM SIGGROUP Bulletin*, 22(1), 19-21.
- Schmidt, A., Gellersen, H., & Stuhr, T. (2001). Context-phonebook extending mobile phone applications with context.
- Seipp, K., & Devlin, K. (2013, August). Landscape vs portrait mode: which is faster to use on your smart phone?. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services* (pp. 534-539). ACM.

- Shepard, C., Rahmati, A., Tossell, C., Zhong, L., & Kortum, P. (2011). LiveLab: measuring wireless networks and smartphone users in the field. *ACM SIGMETRICS Performance Evaluation Review*, 38(3), 15-20.
- Shin, C., Hong, J. H., & Dey, A. K. (2012, September). Understanding and prediction of mobile application usage for smart phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 173-182). ACM.
- Singh, G. (2006). PIM for mobility. *Personal Information Management – Special Interest Group for Information Retrieval Workshop* (pp. 98-101).
- Smit, L., Stander, A., & Ophoff, J. (2012). An Analysis of Base Station Location Accuracy within Mobile-Cellular Networks. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, 1(4), 272-279.
- Sohn, T., Li, K. A., Griswold, W. G., & Hollan, J. D. (2008, April). A diary study of mobile information needs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 433-442). ACM.
- Srinivasan, V., Moghaddam, S., Mukherji, A., Rachuri, K. K., Xu, C., & Tapia, E. M. (2014, September). Mobileminer: Mining your frequent patterns on your phone. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 389-400). ACM.
- Stefanis, V., Komninos, A., Plessas, A., & Garofalakis, J. (2013, August). An interface for context-aware retrieval of mobile contacts. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services* (pp. 492-497). ACM.
- Stefanis, V., Plessas, A., Komninos, A., & Garofalakis, J. (2012a, October). Efficient Support of Context-Driven Interfaces for Mobile Devices. In *Informatics (PCI), 2012 16th Panhellenic Conference on* (pp. 363-368). IEEE.
- Stefanis, V., Plessas, A., Komninos, A., & Garofalakis, J. (2012b, September). Patterns of usage and context in interaction with communication support applications in mobile devices. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services* (pp. 25-34). ACM.
- Stefanis, V., Plessas, A., Komninos, A., & Garofalakis, J. (2014). Frequency and recency context for the management and retrieval of personal information on mobile devices. *Pervasive and Mobile Computing*, 15, 100-112.
- Strang, T., & Linnhoff-Popien, C. (2004, September). A context modeling survey. In *First International Workshop on Advanced Context Modelling, Reasoning and Management*.

- Subramanya, S. R. (2012, April). Analysis of Mobile Phone Call Data to Determine User Characteristics and to Enhance User Experience. *IJCA Proc. of Nat. Conf. on Recent Trends in Computing* (pp. 9-13).
- Sun, C., Wang, Y., Zheng, J., & Hsu, D. F. (2013, July). Feature fusion for mobile usage prediction using rank-score characteristics. In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2013 12th IEEE International Conference on* (pp. 212-217). IEEE.
- Teevan, J., & Jones, W. (2008, April). The disappearing desktop: pim 2008. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3917-3920). ACM.
- Teevan, J., Jones, W., & Bederson, B. B. (2006). Personal information management. *Communications of the ACM*, 49(1), 40-43.
- Toninelli, A., Khushraj, D., Lassila, O., & Montanari, R. (2008, October). Towards socially aware mobile phones. In *7th International Semantic Web Conference* (pp. 68-75).
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97-136.
- Trung, N. V., & Oh, A. H. User Evaluation of a Prototype for Tagging and Sharing Mobile Contacts. *MobileHCI 2010*.
- Tungare, M., & Pérez-Quinones, M. (2008a). It's not what you have, but how you use it: Compromises in mobile device use. *arXiv preprint arXiv:0801.4423*.
- Tungare, M., & Perez-Quinones, M. (2008b). Thinking outside the (beige) box: Personal information management beyond the desktop. In *Proceedings of the 3rd Invitational Workshop on Personal Information Management* (p. 8).
- Tungare, M., & Pérez-Quinones, M. A. (2009, April). Mental workload in multi-device personal information management. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 3431-3436). ACM.
- Yoon, Y., Ahn, Y., Lee, G., Hong, S., & Kim, M. (2008, September). Context-aware photo selection for promoting photo consumption on a mobile phone. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 33-42). ACM.
- Verkasalo, H. (2009). Contextual patterns in mobile service usage. *Personal and Ubiquitous Computing*, 13(5), 331-342.
- Viana, W., Miron, A. D., Moisuc, B., Gensel, J., Villanova-Oliver, M., & Martin, H. (2011). Towards the semantic and context-aware management of mobile multimedia. *Multimedia Tools and Applications*, 53(2), 391-429.

- Wang, X. H., Zhang, D. Q., Gu, T., & Pung, H. K. (2004, March). Ontology based context modeling and reasoning using OWL. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on* (pp. 18-22). IEEE.
- Wang, X., Rosenblum, D., & Wang, Y. (2012, October). Context-aware mobile music recommendation for daily activities. In *Proceedings of the 20th ACM international conference on Multimedia* (pp. 99-108). ACM.
- Whittaker, S. (2011). Personal information management: from information consumption to curation. *Annual review of information science and technology*, 45(1), 1-62.
- Whittaker, S., Bergman, O., & Clough, P. (2010). Easy on that trigger dad: a study of long term family photo retrieval. *Personal and Ubiquitous Computing*, 14(1), 31-43.
- Whittaker, S., & Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(2), 150-170.
- Whittaker, S., & Sidner, C. (1996, April). Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 276-283). ACM.
- Whittaker, S., Jones, Q., & Terveen, L. (2002a, January). Managing long term communications: Conversation and contact management. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on* (pp. 1070-1079). IEEE.
- Whittaker, S., Jones, Q., & Terveen, L. (2002b, November). Contact management: identifying contacts to support long-term communication. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work* (pp. 216-225). ACM.
- Wiese, J., Min, J. K., Hong, J. I., & Zimmerman, J. (2015, February). You Never Call, You Never Write: Call and SMS Logs Do Not Always Indicate Tie Strength. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 765-774). ACM.
- Woerndl, W., & Hristov, A. (2009, February). Recommending resources in mobile personal information management. In *Digital Society, 2009. ICDS'09. Third International Conference on* (pp. 149-154). IEEE.
- Woerndl, W., & Woehrl, M. (2008, April). SeMoDesk: Towards a Mobile Semantic Desktop. In *Proceedings of Personal Information Management (PIM) Workshop*.
- Zhou, L., Mohammed, A. S., & Zhang, D. (2012). Mobile personal information management agent: Supporting natural language interface and application integration. *Information Processing & Management*, 48(1), 23-31.