ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

Ασφάλεια Πληροφοριακών και Επικοινωνιακών Συστημάτων

Προστασία της Ιδιωτικότητας σε χωρικά δεδομένα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Κατσίδη Π. Ηλία

Επιβλέπων : Βλάχου Ακριβή

Μέλη εξεταστικής επιτροπής: **Κωστούλας Θεόδωρος & Μαλιάτσος Κωνσταντίνος**

Σάμος, Αύγουστος 2022

Πρόλογος και ευχαριστίες

Θα ήθελα να αφιερώσω λίγο χρόνο για να εκφράσω την ευγνωμοσύνη μου στους ανθρώπους που έκαναν την συγγραφή της παρούσας εργασίας δυνατή.

Πρώτον, θέλω να ευχαριστήσω την όμορφη σύζυγό μου, που υπήρξε ο βράχος μου, η κολλητή μου και η μεγαλύτερη υποστηρίκτρια μου σε όλη αυτή τη διαδικασία. Η αταλάντευτη υποστήριξη, η αγάπη και η πίστη σου σε έχουν σημάνει τον κόσμο για μένα και είμαι πραγματικά ευλογημένος που σε έχω δίπλα μου. Στάθηκες δίπλα μου και ήσουν πάντα εκεί για να με σηκώνεις όταν έπεφτα. Ήσουν σύντροφός μου σε κάθε πτυχή της ζωής μου και είμαι θα για πάντα ευγνώμων για όλα όσα έκανες για μένα.

Στη συνέχεια, θέλω να ευχαριστήσω την οικογένειά μου, που πάντα με ενθάρρυνε να προσπαθώ για το καλύτερο και να μην εγκαταλείπω ποτέ τα όνειρά μου. Η αταλάντευτη υποστήριξη και η αγάπη σας ήταν η κινητήρια δύναμη πίσω από την επιτυχία μου και είμαι περήφανος που είμαι μέλος μιας τόσο στοργικής και υποστηρικτικής οικογένειας. Σας ευχαριστώ που ήσασταν εκεί για μένα και που με βοηθήσατε να παραμείνω συγκεντρωμένος και με κίνητρο, ακόμα και όταν τα πράγματα ήταν δύσκολα.

Τέλος, θέλω να ευχαριστήσω την καθηγήτρια μου, η οποία υπήρξε μέντορας, οδηγός και έμπνευση σε όλη αυτή τη διαδικασία. Οι γνώσεις, η τεχνογνωσία και η αφοσίωσή σας στη διδασκαλία με βοήθησαν να εξελιχθώ τόσο επαγγελματικά όσο και προσωπικά. Η καθοδήγηση και η υποστήριξή σας ήταν ανεκτίμητη και είμαι πραγματικά ευγνώμων για τον χρόνο και την ενέργεια που έχετε επενδύσει σε μένα.

Κλείνοντας, θα ήθελα να πω ότι είμαι βαθιά ευγνώμων για την υποστήριξη και την ενθάρρυνση που έχω λάβει από τη σύζυγό μου, την οικογένειά μου και την καθηγήτρια μου. Χωρίς την αγάπη και την καθοδήγησή σας, δεν θα μπορούσα να πετύχω αυτή την επιτυχία και είμαι πραγματικά ευγνώμων για τον καθένα από εσάς. Ήταν τιμή μου που είχα ένα τόσο υπέροχο σύστημα υποστήριξης και θα αγαπώ πάντα τις αναμνήσεις που έχουμε δημιουργήσει μαζί.

Ευχαριστώ!

ΚΑΤΣΙΔΗ Π. ΗΛΙΑ

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Πίνακας περιεχομένων

Λίστα Σχημάτων

Λίστα Πινάκων

Ακρωνύμια

NN-Cloak	Non-Neighbourhood-Cloak
ARX	Automated Recoding eXchange
LBS	Location-Based Services
GPS	Global Positioning System
KML	Keyhole Markup Language
GDPR	General Data Protection Regulation
SSL	Secure Sockets Layer
TLS	Transport Layer Security
CASPER	Clustering-based Anonymity using SPatially-bounded Errors
CryptDB	Encrypted and Controlled Database System
SQL	Structured Query Language

Περίληψη

Σε αυτή τη διατριβή, διερευνάται η έννοια της ανωνυμίας των Σημειακών χωρικών δεδομένων, στις υπηρεσίες που βασίζονται στην τοποθεσία και τις διάφορες τεχνικές που χρησιμοποιούνται για τη διατήρηση του απορρήτου των ατόμων σε αυτές τις υπηρεσίες. Αυτή η έρευνα παρέχει μια ολοκληρωμένη επισκόπηση των διαφόρων τεχνικών απόκρυψης που χρησιμοποιούνται σε υπηρεσίες που βασίζονται σε τοποθεσία, συμπεριλαμβανομένου του cloaking data base, του cloaking που εξαρτάται από το διάστημα, του Clique cloak, του Center Cloak, του NN-Cloak, της τεχνικής Casper, του interval cloak και του Hilbert cloak. Επίσης συζητιούνται οι διάφορες μεθόδους προστασίας του απορρήτου, όπως η καταστολή δεδομένων, η διαταραχή δεδομένων και η γενίκευση δεδομένων.

Η μελέτη εμβαθύνει περαιτέρω στην έννοια των ιχνών δεδομένων και των ανησυχιών περί απορρήτου που σχετίζονται με αυτά, καθώς και στα μέτρα ασφαλείας που εφαρμόζονται για την προστασία από κακόβουλες επιθέσεις. Επίσης αναλύονται δύο από τα πιο δημοφιλή εργαλεία προστασίας δεδομένων, τα CryptDB και ARX, και συγκρίνονται τα δυνατά και τα αδύνατα σημεία τους, επισημαίνοντας τα οφέλη του ARX έναντι του CryptDB.

Επιπλέον, έχει γίνει διερεύνηση του θέματος της δημοσίευσης δεδομένων για τη διατήρηση του απορρήτου, παρουσιάζοντας ένα γενικό πλαίσιο για το σκοπό αυτό. Αυτό το πλαίσιο παρέχει μια δομημένη προσέγγιση για την ανάλυση των ανησυχιών περί απορρήτου που σχετίζονται με τη δημοσίευση δεδομένων και παρέχει καθοδήγηση σχετικά με τις μεθόδους που χρησιμοποιούνται για τον μετριασμό αυτών των κινδύνων.

Τέλος, γίνεται αναθεώρηση μερικών από τις πιο σημαντικές ερευνητικές εργασίες σε αυτόν τον τομέα, συμπεριλαμβανομένων των "Clique Cloak", "A General Framework for Privacy-Preserving Data Publishing", "NN-Cloak: An Effective and Efficient K-Anonymization Algorithm for Trajectory Data ", "Anonymizing Spatial Trajectory Data" και "Cloaking Location-Based Services: A Survey of Techniques and Attacks".

Συνολικά, αυτή η διατριβή συμβάλλει σημαντικά στον τομέα της προστασίας της ιδιωτικής ζωής στις υπηρεσίες που βασίζονται στην τοποθεσία και θα έχει μεγάλη αξία τόσο για ερευνητές όσο και για επαγγελματίες που ενδιαφέρονται να εξερευνήσουν αυτόν τον τομέα σε μεγαλύτερο βάθος.

Abstract

In this thesis, the concept of anonymity in location-based services and the various techniques used to maintain the privacy of individuals in these services are explored. This research provides a comprehensive overview of various cloaking techniques used in location-based services, including data base cloaking, space-dependent cloaking, Clique cloak, central cloak, NN-Cloak, Casper technique, the interval cloak and the Hilbert cloak. Also discussed are the various methods of protecting privacy, such as data suppression, data perturbation, and data generalization.

The study further delves into the concept of data traces and the privacy concerns associated with them, as well as the security measures implemented to protect against malicious attacks. You've also analyzed two of the most popular data protection tools, CryptDB and ARX, and compared their strengths and weaknesses, highlighting the benefits of ARX over CryptDB.

In addition, we have explored the issue of publishing data to preserve privacy, presenting a general framework for this purpose. This framework provides a structured approach to analyzing privacy concerns related to data publication and provides guidance on methods used to mitigate these risks.

Finally, we have reviewed some of the most important research papers in this field, including "Clique Cloak", "A General Framework for Privacy-Preserving Data Publishing", "NN-Cloak: An Effective and Efficient K-Anonymization Algorithm for Trajectory Data", "Anonymizing Spatial Trajectory Data" and "Cloaking Location-Based Services: A Survey of Techniques and Attacks".

Overall, this thesis makes a significant contribution to the field of privacy protection in location-based services and will be of great value to both researchers and practitioners interested in exploring this area in greater depth.

1

Εισαγωγή

Η πρόοδος της τεχνολογίας έχει οδηγήσει σε σημαντική αύξηση στη συλλογή, αποθήκευση και επεξεργασία προσωπικών δεδομένων. Η ευρεία χρήση ψηφιακών συσκευών, όπως τα smartphone και η τεχνολογία φορητών συσκευών, έχει διευκολύνει τη συλλογή και αποθήκευση τεράστιων ποσοτήτων προσωπικών δεδομένων, συμπεριλαμβανομένων των δεδομένων τοποθεσίας. Οι υπηρεσίες που βασίζονται στην τοποθεσία (LBS) βασίζονται στη συλλογή αυτών των δεδομένων για την παροχή εξατομικευμένων υπηρεσιών στους χρήστες. Ωστόσο, η συλλογή και αποθήκευση προσωπικών δεδομένων εγείρει σοβαρές ανησυχίες για το απόρρητο, καθώς μπορεί να χρησιμοποιηθεί για κακόβουλους σκοπούς, όπως κλοπή ταυτότητας, απάτη ή παρακολούθηση.

Για να αντιμετωπιστούν αυτές οι ανησυχίες, το πεδίο της ανωνυμοποίησης δεδομένων έχει προκύψει για να παρέχει προστασία απορρήτου για ευαίσθητα δεδομένα αφαιρώντας ή μετασχηματίζοντας στοιχεία προσωπικής ταυτοποίησης. Αυτή η διαδικασία στοχεύει να εξισορροπήσει την ανάγκη για απόρρητο δεδομένων με την ανάγκη για χρησιμότητα δεδομένων, επιτρέποντας στους οργανισμούς να χρησιμοποιούν προσωπικά δεδομένα για αναλυτικούς σκοπούς προστατεύοντας παράλληλα το απόρρητο των ατόμων.

Ένας τύπος ευαίσθητων δεδομένων που είναι ιδιαίτερα δύσκολο να ανωνυμοποιηθούν είναι τα χωρικά δεδομένα. Τα χωρικά δεδομένα είναι δεδομένα που περιγράφουν τη θέση και το σχήμα των αντικειμένων στο χώρο, συμπεριλαμβανομένων σημείων, γραμμών και πολυγώνων. Τα σημειακά χωρικά δεδομένα είναι ένα υποσύνολο χωρικών δεδομένων που περιγράφει τη θέση ενός μόνο σημείου στο χώρο. Τα σημειακά χωρικά δεδομένα χρησιμοποιούνται συχνά για να περιγράψουν τη θέση των ατόμων και είναι ο πιο συχνά χρησιμοποιούμενος τύπος χωρικών δεδομένων στο LBS.

Το πρόβλημα της ανωνυμίας στα σημειακά χωρικά δεδομένα αναφέρεται στη δυσκολία προστασίας του απορρήτου των ατόμων όταν συλλέγονται και κοινοποιούνται τα δεδομένα τοποθεσίας τους. Αυτό μπορεί να είναι ιδιαίτερα προβληματικό όταν εμπλέκονται ευαίσθητες πληροφορίες, όπως οι διευθύνσεις κατοικίας. Επιπλέον, η ακρίβεια των δεδομένων τοποθεσίας, όπως οι συντεταγμένες GPS, έχει αυξηθεί σημαντικά, καθιστώντας ευκολότερο τον εντοπισμό ατόμων με βάση το ιστορικό τοποθεσίας τους.

Το ARX είναι μια λύση λογισμικού που στοχεύει στην αντιμετώπιση του προβλήματος της ανωνυμίας στα σημειακά χωρικά δεδομένα. Λειτουργεί εκτελώντας ανωνυμοποίηση δεδομένων σε δεδομένα τοποθεσίας προτού κοινοποιηθούν, καθιστώντας πολύ πιο δύσκολο τον εντοπισμό ατόμων με βάση το ιστορικό τοποθεσίας τους. Αυτό επιτυγχάνεται μετατρέποντας τα δεδομένα σε λιγότερο ακριβή μορφή, διατηρώντας παράλληλα τη χρησιμότητα των δεδομένων για τον επιδιωκόμενο σκοπό. Για παράδειγμα, το λογισμικό μπορεί να στρογγυλοποιεί τις συντεταγμένες GPS σε χαμηλότερο επίπεδο ακρίβειας ή να συγκεντρώνει τα δεδομένα σε μεγαλύτερες περιοχές, καθιστώντας πιο δύσκολο τον εντοπισμό μεμονωμένων χρηστών.

Για την προστασία του απορρήτου των ατόμων, έχουν αναπτυχθεί διάφορες τεχνικές ανωνυμοποίησης, συμπεριλαμβανομένων των k-anonymity, l-diversity και t-closeness. Αυτές οι τεχνικές στοχεύουν να διασφαλίσουν ότι τα προσωπικά στοιχεία ενός χρήστη δεν μπορούν να ανιχνευθούν σε αυτά ακόμη και αν τα δεδομένα αποκαλυφθούν ή διαρρεύσουν. Εκτός από αυτές τις μεθόδους, έχουν αναπτυχθεί διάφορες τεχνικές απόκρυψης για την προστασία του απορρήτου των χρηστών στο LBS, συμπεριλαμβανομένης της καταστολής δεδομένων, της διαταραχής δεδομένων, της γενίκευσης δεδομένων και της απόκρυψης που εξαρτάται από το διάστημα.

Αυτή η διπλωματική εργασία παρέχει μια επισκόπηση του απορρήτου των χωρικών δεδομένων και των διαφόρων τεχνικών και εργαλείων που είναι διαθέσιμα για την προστασία του απορρήτου των χρήστες στο LBS, συμπεριλαμβανομένων των τεχνικών k-anonymity, l-diversity, t-closeness και cloaking. Θα συζητήσουμε επίσης τα κομμάτια δεδομένων και το CryptDB, επισημαίνοντας τις κύριες διαφορές μεταξύ ARX και CryptDB και τα πλεονεκτήματα και τα μειονεκτήματα του καθενός.

Συμπερασματικά, η προστασία της ιδιωτικής ζωής είναι μια κρίσιμη πτυχή της συλλογής και αποθήκευσης δεδομένων, ιδιαίτερα όταν πρόκειται για ευαίσθητα δεδομένα, όπως τα χωρικά δεδομένα. Οι τεχνικές ανωνυμοποίησης και οι μέθοδοι απόκρυψης παρέχουν σημαντικά εργαλεία για τους οργανισμούς για την προστασία του απορρήτου των χρηστών, ενώ παράλληλα επιτρέπουν στα δεδομένα να είναι χρήσιμα για αναλυτικούς σκοπούς. Ωστόσο, η επιλογή της μεθόδου ανωνυμοποίησης εξαρτάται από τις συγκεκριμένες απαιτήσεις και περιορισμούς κάθε μεμονωμένης περίπτωσης και είναι σημαντικό να ληφθούν υπόψη προσεκτικά οι συμβιβασμούς μεταξύ της ιδιωτικής ζωής και της χρησιμότητας δεδομένων.

2

Βιβλιογραφική Ανασκόπηση

2.1 Χωρικά δεδομένα και ιδιωτικότητα

Τα χωρικά δεδομένα αναφέρονται σε δεδομένα που περιέχουν πληροφορίες για μια τοποθεσία. Με την αυξανόμενη χρήση του GPS και άλλων τεχνολογιών που βασίζονται στην τοποθεσία, μεγάλος αριθμός χωρικών δεδομένων συλλέγεται, αποθηκεύεται και αναλύεται. Ωστόσο, η συλλογή και η χρήση αυτών των δεδομένων εγείρει σημαντικές ανησυχίες σχετικά με το απόρρητο.

Ένα από τα κύρια προβλήματα με τα χωρικά δεδομένα είναι η δυνατότητα εντοπισμού και επιτήρησης τοποθεσίας. Όταν η τοποθεσία ενός ατόμου παρακολουθείται με την πάροδο του χρόνου, μπορεί να αποκαλύψει ευαίσθητες πληροφορίες, όπως η καθημερινή του ρουτίνα, τα μέρη που συχνάζουν, ακόμη και οι προσωπικές του σχέσεις. Αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν για να συναχθούν ευαίσθητες πληροφορίες σχετικά με τη ζωή ενός ατόμου, όπως οι πολιτικές του απόψεις ή ο σεξουαλικός προσανατολισμός. Επιπλέον, η συλλογή χωρικών δεδομένων μπορεί επίσης να αποκαλύψει ευαίσθητες πληροφορίες σχετικά με την περιουσία και τα περιουσιακά στοιχεία ενός ατόμου, τα οποία μπορούν να χρησιμοποιηθούν για στοχευμένη διαφήμιση ή άλλες μορφές εκμετάλλευσης.

Μια άλλη ανησυχία με τα χωρικά δεδομένα είναι η πιθανότητα παραβιάσεων δεδομένων. Τα χωρικά δεδομένα αποθηκεύονται συχνά σε μεγάλες βάσεις δεδομένων και εάν αυτές οι βάσεις δεδομένων δεν είναι σωστά ασφαλισμένες, τα δεδομένα που περιέχουν μπορούν να έχουν πρόσβαση από μη εξουσιοδοτημένα μέρη. Αυτό μπορεί να οδηγήσει σε απώλεια του απορρήτου για τα άτομα των οποίων τα δεδομένα περιέχονται στη βάση δεδομένων.

Για την αντιμετώπιση αυτών των ανησυχιών, είναι σημαντικό τα άτομα να ενημερώνονται για τη συλλογή και τη χρήση των χωρικών τους δεδομένων και να έχουν τη δυνατότητα να ελέγχουν ποιος έχει πρόσβαση σε αυτά τα δεδομένα. Αυτό μπορεί να γίνει με τη χρήση των πολιτικών απορρήτου και των συμφωνιών παροχής υπηρεσιών, καθώς και με τη χρήση τεχνολογιών που ενισχύουν το απόρρητο, όπως ασφαλή πρωτόκολλα επικοινωνίας και τεχνικές ανωνυμοποίησης δεδομένων.

Επιπλέον, υπάρχει ανάγκη για νόμους και κανονισμούς για την προστασία της ιδιωτικής ζωής των ατόμων σε σχέση με τα χωρικά δεδομένα. Αυτό θα μπορούσε να περιλαμβάνει νόμους που απαιτούν από τις εταιρείες να λαμβάνουν ρητή συναίνεση πριν συλλέγουν, χρησιμοποιούν ή κοινοποιούν χωρικά δεδομένα ατόμων, καθώς και νόμους που απαιτούν από τις εταιρείες να ειδοποιούν τα άτομα όταν τα δεδομένα τους έχουν προσπελαστεί ή χρησιμοποιηθεί με μη εξουσιοδοτημένο τρόπο.

Συμπερασματικά, καθώς η χρήση χωρικών δεδομένων συνεχίζει να αυξάνεται, είναι σημαντικό να ληφθούν υπόψη οι επιπτώσεις αυτών των δεδομένων στο απόρρητο και να ληφθούν μέτρα για την προστασία της ιδιωτικής ζωής των ατόμων. Αυτό μπορεί να γίνει μέσω ενός συνδυασμού διαφάνειας, ελέγχου χρήστη και νομικής προστασίας.

2.2 Τι είναι χωρικά δεδομένα, με έμφαση στα σημειακά χωρικά δεδομένα

Τα χωρικά δεδομένα αναφέρονται σε δεδομένα που περιέχουν πληροφορίες για μια τοποθεσία. Αυτό μπορεί να περιλαμβάνει δεδομένα σχετικά με τη φυσική θέση ενός χαρακτηριστικού, όπως οι συντεταγμένες ενός κτιρίου ή το σχήμα μιας λίμνης, καθώς και δεδομένα σχετικά με τα χαρακτηριστικά ενός χαρακτηριστικού, όπως ο πληθυσμός μιας πόλης ή το όνομα ενός ποταμού . Τα χωρικά δεδομένα μπορούν να αναπαρασταθούν σε διάφορες μορφές, συμπεριλαμβανομένων χαρτών, εικόνων και πινάκων, και μπορούν να χρησιμοποιηθούν για ένα ευρύ φάσμα εφαρμογών, όπως η πλοήγηση, ο πολεοδομικός σχεδιασμός και η διαχείριση φυσικών πόρων.

Ένας τύπος χωρικών δεδομένων είναι τα σημειακά χωρικά δεδομένα, τα οποία αποτελούνται από ένα μόνο σημείο ή σύνολο σημείων που αντιπροσωπεύουν μια θέση στην επιφάνεια της γης. Τα σημειακά δεδομένα μπορούν να χρησιμοποιηθούν για να αναπαραστήσουν ένα ευρύ φάσμα χαρακτηριστικών, συμπεριλαμβανομένων κτιρίων, ορόσημων και φυσικών χαρακτηριστικών, όπως δέντρα και βράχοι. Τα σημειακά δεδομένα μπορούν να συλλεχθούν χρησιμοποιώντας μια ποικιλία μεθόδων, όπως GPS, αεροφωτογράφιση και επιτόπιες έρευνες, και μπορούν να αποθηκευτούν σε διάφορες μορφές, όπως shapefiles και KML.

Τα σημειακά χωρικά δεδομένα έχουν πολλές χρήσεις, συμπεριλαμβανομένης της πλοήγησης, του σχεδιασμού διαδρομής και των υπηρεσιών που βασίζονται στην τοποθεσία. Για παράδειγμα, τα σημειακά δεδομένα μπορούν να χρησιμοποιηθούν για τη χαρτογράφηση της τοποθεσίας των επιχειρήσεων, των ορόσημων και άλλων σημείων ενδιαφέροντος, για την παροχή αναλυτικών οδηγιών προς μια τοποθεσία ή για τη βοήθεια των ατόμων να βρουν κοντινές επιχειρήσεις ή

υπηρεσίες. Τα σημειακά δεδομένα μπορούν επίσης να χρησιμοποιηθούν στο πλαίσιο του GIS (Geographic Information System) για χωρική ανάλυση και χαρτογράφηση.

Τα σημειακά δεδομένα μπορούν επίσης να χρησιμοποιηθούν στον τομέα της τρισδιάστατης χαρτογράφησης και μοντελοποίησης, γεγονός που επιτρέπει την ακριβέστερη αναπαράσταση της επιφάνειας της γης και των χαρακτηριστικών σε αυτήν. Τα σημειακά δεδομένα μπορούν να χρησιμοποιηθούν για τη δημιουργία λεπτομερών τρισδιάστατων μοντέλων κτιρίων, τοπίων και άλλων χαρακτηριστικών, τα οποία μπορούν να χρησιμοποιηθούν για ένα ευρύ φάσμα εφαρμογών όπως ο πολεοδομικός σχεδιασμός, ο αρχιτεκτονικός σχεδιασμός και η εικονική πραγματικότητα.

Επιπλέον, τα σημειακά δεδομένα μπορούν να χρησιμοποιηθούν για τη βελτίωση της αποτελεσματικότητας και της ασφάλειας των συστημάτων μεταφοράς παρέχοντας πληροφορίες για τα κυκλοφοριακά μοτίβα, τον σχεδιασμό διαδρομής και τον εντοπισμό περιοχών όπου μπορούν να γίνουν βελτιώσεις.

Συμπερασματικά, σημειακά χωρικά δεδομένα είναι ένας συγκεκριμένος τύπος χωρικών δεδομένων που χρησιμοποιείται για να αναπαραστήσει ένα μόνο σημείο ή σύνολο σημείων στην επιφάνεια της γης. Τα σημειακά δεδομένα έχουν ένα ευρύ φάσμα χρήσεων, συμπεριλαμβανομένης της πλοήγησης, του σχεδιασμού διαδρομής, των υπηρεσιών που βασίζονται στη θέση, του GIS, της τρισδιάστατης χαρτογράφησης και μοντελοποίησης και των συστημάτων μεταφοράς.

2.3 Τι είναι το k-anonimity και ποιες είναι οι βασικές τεχνικές εφαρμογής του

Η κ-ανωνυμία είναι μια συγκεκριμένη τεχνική που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων στα δεδομένα. Εισήχθη για πρώτη φορά από τον Sweeney το 2002 ως ένας τρόπος για την προστασία του απορρήτου των ατόμων σε μεγάλα σύνολα δεδομένων, γενικεύοντας τα δεδομένα με τέτοιο τρόπο ώστε να μην είναι δυνατός ο εντοπισμός οποιουδήποτε ατόμου με έναν μοναδικό συνδυασμό τιμών για ένα σύνολο ευαίσθητων χαρακτηριστικών.

Η βασική ιδέα πίσω από την κ-ανωνυμία είναι ότι ένα σύνολο δεδομένων θεωρείται κ-ανώνυμο εάν για κάθε εγγραφή στο σύνολο δεδομένων, υπάρχουν τουλάχιστον κ-1 άλλες εγγραφές στο σύνολο δεδομένων που έχουν την ίδια τιμή για τα ευαίσθητα χαρακτηριστικά. Αυτό σημαίνει ότι για κάθε εγγραφή στο σύνολο δεδομένων, υπάρχουν τουλάχιστον κ-1 άλλες εγγραφές που έχουν την ίδια τιμή για τα ευαίσθητα χαρακτηριστικά, καθιστώντας αδύνατη την αναγνώριση οποιασδήποτε μεμονωμένης εγγραφής με έναν μοναδικό συνδυασμό τιμών.

Μία από τις βασικές τεχνικές που χρησιμοποιούνται για την επίτευξη της κ-ανωνυμίας είναι η γενίκευση, η οποία περιλαμβάνει τη μείωση του επιπέδου λεπτομέρειας στα δεδομένα με τη συγκέντρωση δεδομένων σε υψηλότερο επίπεδο ή με την αφαίρεση ορισμένων χαρακτηριστικών. Αυτό μπορεί να είναι χρήσιμο για την προστασία ευαίσθητων πληροφοριών, ενώ εξακολουθεί να επιτρέπει τη χρήση των δεδομένων για ανάλυση.

Μια άλλη τεχνική είναι η καταστολή, η οποία περιλαμβάνει την αφαίρεση ορισμένων τιμών ή εγγραφών από το σύνολο δεδομένων, εάν καθιστούν δυνατή την αναγνώριση ενός ατόμου με έναν μοναδικό συνδυασμό τιμών.

Μια πιο περίπλοκη τεχνική είναι η χρήση ομαδοποίησης, η οποία είναι μια τεχνική που χρησιμοποιείται για την ομαδοποίηση παρόμοιων εγγραφών. Αυτό μπορεί να είναι χρήσιμο για την προστασία ευαίσθητων πληροφοριών, ενώ εξακολουθεί να επιτρέπει τη χρήση των δεδομένων για ανάλυση.

Μια άλλη τεχνική είναι η χρήση της προσθήκης θορύβου, η οποία περιλαμβάνει την προσθήκη τυχαίου θορύβου στα δεδομένα για να κρύψει τις πραγματικές τιμές των χαρακτηριστικών, καθιστώντας πιο δύσκολο τον εντοπισμό ατόμων με έναν μοναδικό συνδυασμό τιμών.

Συμπερασματικά, η κ-ανωνυμία είναι μια τεχνική που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων στα δεδομένα με τη γενίκευση των δεδομένων με τέτοιο τρόπο ώστε να μην είναι δυνατός ο εντοπισμός οποιουδήποτε ατόμου με έναν μοναδικό συνδυασμό τιμών για ένα σύνολο ευαίσθητων χαρακτηριστικών. Οι βασικές τεχνικές που χρησιμοποιούνται για την επίτευξη της κ-ανωνυμίας περιλαμβάνουν τη γενίκευση, την καταστολή, τη ομαδοποίηση και την προσθήκη θορύβου. Κάθε μία από αυτές τις τεχνικές έχει τα δικά της πλεονεκτήματα και μειονεκτήματα και η καλύτερη προσέγγιση θα εξαρτηθεί από τη συγκεκριμένη περίπτωση χρήσης και το επίπεδο ανωνυμίας που απαιτείται.

2.4 Τι είναι το l-diversity και ποιες είναι οι βασικές τεχνικές εφαρμογής του

Το L-diversity είναι μια συγκεκριμένη τεχνική που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων στα δεδομένα διασφαλίζοντας ότι υπάρχει επαρκής ποικιλομορφία σε μια ομάδα εγγραφών που μοιράζονται τις ίδιες τιμές για ένα σύνολο ευαίσθητων χαρακτηριστικών. Εισήχθη για πρώτη φορά από τους Machanavajjhala et al. το 2006 ως ένας τρόπος αντιμετώπισης των περιορισμών της k-ανωνυμίας, που διασφαλίζει μόνο ότι υπάρχει επαρκής αριθμός εγγραφών που μοιράζονται τις ίδιες τιμές για ευαίσθητα χαρακτηριστικά, αλλά όχι ότι οι τιμές είναι διαφορετικές.

Η βασική ιδέα πίσω από την l-diversity είναι ότι ένα σύνολο δεδομένων θεωρείται ότι είναι l-diverse εάν για κάθε ευαίσθητη τιμή υπάρχουν τουλάχιστον l καλά αντιπροσωπευόμενες, διακριτές, ευαίσθητες τιμές στην κλάση ισοδυναμίας (η ομάδα εγγραφών που μοιράζονται το ίδιο τιμή για το ευαίσθητο χαρακτηριστικό). Αυτό σημαίνει ότι για κάθε ευαίσθητη τιμή, υπάρχουν τουλάχιστον l διαφορετικές τιμές στην ομάδα εγγραφών, γεγονός που καθιστά δύσκολο τον προσδιορισμό οποιασδήποτε μεμονωμένης εγγραφής με έναν μοναδικό συνδυασμό τιμών.

Μία από τις βασικές τεχνικές που χρησιμοποιούνται για την επίτευξη l-diversity είναι η γενίκευση, η οποία περιλαμβάνει τη μείωση του επιπέδου λεπτομέρειας στα δεδομένα με τη συγκέντρωση δεδομένων σε υψηλότερο επίπεδο ή με την αφαίρεση ορισμένων χαρακτηριστικών.

Αυτό μπορεί να είναι χρήσιμο για την προστασία ευαίσθητων πληροφοριών, ενώ εξακολουθεί να επιτρέπει τη χρήση των δεδομένων για ανάλυση.

Μια άλλη τεχνική είναι η καταστολή, η οποία περιλαμβάνει την αφαίρεση ορισμένων τιμών ή εγγραφών από το σύνολο δεδομένων, εάν καθιστούν δυνατή την αναγνώριση ενός ατόμου με έναν μοναδικό συνδυασμό τιμών.

Μια πιο περίπλοκη τεχνική είναι η χρήση ομαδοποίησης, η οποία είναι μια τεχνική που χρησιμοποιείται για την ομαδοποίηση παρόμοιων εγγραφών. Αυτό μπορεί να είναι χρήσιμο για την προστασία ευαίσθητων πληροφοριών, ενώ εξακολουθεί να επιτρέπει τη χρήση των δεδομένων για ανάλυση.

Μια άλλη τεχνική είναι η χρήση της προσθήκης θορύβου, η οποία περιλαμβάνει την προσθήκη τυχαίου θορύβου στα δεδομένα για να κρύψει τις πραγματικές τιμές των χαρακτηριστικών, καθιστώντας πιο δύσκολο τον εντοπισμό ατόμων με έναν μοναδικό συνδυασμό τιμών.

Συμπερασματικά, η l-diversity είναι μια τεχνική που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων στα δεδομένα διασφαλίζοντας ότι υπάρχει επαρκής ποικιλομορφία σε μια ομάδα εγγραφών που μοιράζονται τις ίδιες τιμές για ένα σύνολο ευαίσθητων χαρακτηριστικών. Οι βασικές τεχνικές που χρησιμοποιούνται για την επίτευξη της ποικιλομορφίας l περιλαμβάνουν τη γενίκευση, την καταστολή, τη ομαδοποίηση και την προσθήκη θορύβου. Κάθε μία από αυτές τις τεχνικές έχει τα δικά της πλεονεκτήματα και μειονεκτήματα και η καλύτερη προσέγγιση θα εξαρτηθεί από τη συγκεκριμένη περίπτωση χρήσης και το επίπεδο ποικιλομορφίας που απαιτείται.

2.5 Τι είναι το t-Closeness και ποιες είναι οι βασικές τεχνικές εφαρμογής του

Το T-closeness είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που χρησιμοποιείται για την προστασία ευαίσθητων δεδομένων διασφαλίζοντας ότι οι τιμές των ευαίσθητων χαρακτηριστικών σε ένα σύνολο δεδομένων δεν είναι πολύ παρόμοιες με τις τιμές αυτών των χαρακτηριστικών στον πληθυσμό. Αυτή η τεχνική χρησιμοποιείται για την προστασία από επιθέσεις επαναπροσδιορισμού, στις οποίες ένας εισβολέας χρησιμοποιεί πρόσθετες πληροφορίες για να επαναπροσδιορίσει τα άτομα σε ένα ανώνυμο σύνολο δεδομένων.

Το T-closeness λειτουργεί εισάγοντας ένα όριο, γνωστό ως τιμή "ανοχής", το οποίο χρησιμοποιείται για τον προσδιορισμό της μέγιστης επιτρεπόμενης διαφοράς μεταξύ των τιμών των ευαίσθητων χαρακτηριστικών στο σύνολο δεδομένων και εκείνων στον πληθυσμό. Για παράδειγμα, εάν η τιμή ανοχής οριστεί σε 0,1, τότε οι τιμές των ευαίσθητων χαρακτηριστικών στο σύνολο δεδομένων δεν πρέπει να διαφέρουν περισσότερο από 0,1 από τις αντίστοιχες τιμές στον πληθυσμό.

Για να εφαρμοστεί το t-closeness, τα δεδομένα πρέπει πρώτα να γενικευθούν ή να διαταραχθούν για να διασφαλιστεί ότι οι τιμές των ευαίσθητων χαρακτηριστικών δεν είναι πολύ παρόμοιες με τις τιμές στον πληθυσμό. Αυτό μπορεί να γίνει χρησιμοποιώντας έναν αριθμό διαφορετικών τεχνικών, όπως η καταστολή δεδομένων, ο μετασχηματισμός δεδομένων και η διαταραχή δεδομένων.

Η καταστολή δεδομένων είναι μια τεχνική που περιλαμβάνει την αφαίρεση ευαίσθητων τιμών από το σύνολο δεδομένων, είτε αφαιρώντας ολόκληρη τη γραμμή ή τη στήλη, είτε αντικαθιστώντας την τιμή με μια μάσκα. Ο μετασχηματισμός δεδομένων είναι μια τεχνική που περιλαμβάνει την αλλαγή των τιμών των ευαίσθητων χαρακτηριστικών στο σύνολο δεδομένων, όπως με στρογγυλοποίηση ή τοποθέτηση σε κάδο των τιμών. Η διαταραχή δεδομένων είναι μια τεχνική που περιλαμβάνει την προσθήκη τυχαίου θορύβου στις τιμές των ευαίσθητων χαρακτηριστικών στο σύνολο δεδομένων.

Μια άλλη τεχνική που μπορεί να εφαρμοστεί στο t-closeness, ονομάζεται δημοσίευση δεδομένων με βάση το T-Closeness, είναι μια μέθοδος που ανωνυμοποιεί τα δεδομένα αφαιρώντας τα ευαίσθητα χαρακτηριστικά και ομαδοποιώντας τα δεδομένα σε συμπλέγματα με βάση τα μη ευαίσθητα χαρακτηριστικά και στη συνέχεια προσθέτοντας τυχαίο θόρυβο στα ευαίσθητα χαρακτηριστικά κάθε συστάδας, με αυτόν τον τρόπο τα ευαίσθητα χαρακτηριστικά κάθε συστάδας θα είναι παρόμοια με τον πληθυσμό αλλά όχι πανομοιότυπα.

Συμπερασματικά, το T-closeness είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που χρησιμοποιείται για την προστασία ευαίσθητων δεδομένων διασφαλίζοντας ότι οι τιμές των ευαίσθητων χαρακτηριστικών σε ένα σύνολο δεδομένων δεν είναι πολύ παρόμοιες με τις τιμές αυτών των χαρακτηριστικών στον πληθυσμό. Αυτή η τεχνική υλοποιείται χρησιμοποιώντας τεχνικές γενίκευσης δεδομένων ή διαταραχών όπως η καταστολή δεδομένων, ο μετασχηματισμός δεδομένων και η διαταραχή δεδομένων. Η δημοσίευση δεδομένων με βάση το T-Closeness είναι μια μέθοδος που μπορεί να χρησιμοποιηθεί για την ανωνυμοποίηση των δεδομένων αφαιρώντας τα ευαίσθητα χαρακτηριστικά και ομαδοποιώντας τα δεδομένα σε συμπλέγματα με βάση τα μη ευαίσθητα χαρακτηριστικά και στη συνέχεια προσθέτοντας τυχαίο θόρυβο στα ευαίσθητα χαρακτηριστικά κάθε συμπλέγματος.

2.6 Τι είναι το ARX, πώς λειτουργεί και ποιος είναι ο κύριος σκοπός του

Το ARX (Automated Recoding eXchange) είναι ένα εργαλείο λογισμικού ανωνυμοποίησης δεδομένων που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων στα δεδομένα. Έχει σχεδιαστεί για να αυτοματοποιεί τη διαδικασία ανωνυμοποίησης δεδομένων, διευκολύνοντας τους οργανισμούς να συμμορφώνονται με τους κανονισμούς προστασίας δεδομένων, όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) και ο Νόμος Φορητότητας και Υπευθυνότητας Ασφάλισης Υγείας (HIPAA).

Το ARX λειτουργεί εντοπίζοντας ευαίσθητα δεδομένα μέσα σε ένα σύνολο δεδομένων και στη συνέχεια εφαρμόζοντας διαφορετικές τεχνικές ανωνυμοποίησης για να κρύψει τις τιμές αυτών των χαρακτηριστικών, ενώ εξακολουθεί να επιτρέπει στα δεδομένα να χρησιμοποιηθούν για ανάλυση. Αυτές οι τεχνικές περιλαμβάνουν τη γενίκευση, την καταστολή και την προσθήκη θορύβου, οι οποίες είναι παρόμοιες με αυτές που χρησιμοποιούνται στην l-diversity.

Ένας από τους κύριους σκοπούς του ARX είναι η υποστήριξη k-ανωνυμίας, l-diversity, t-closeness και διαφορικής ιδιωτικότητας. Το λογισμικό μπορεί να χρησιμοποιηθεί για την

ανωνυμοποίηση ενός ευρέος φάσματος τύπων δεδομένων, συμπεριλαμβανομένων δομημένων δεδομένων, ημιδομημένων δεδομένων και μη δομημένων δεδομένων. Μπορεί επίσης να χρησιμοποιηθεί για την ανωνυμοποίηση δεδομένων από διαφορετικές πηγές, όπως βάσεις δεδομένων, υπολογιστικά φύλλα και αρχεία κειμένου.

Το ARX περιλαμβάνει μια φιλική προς το χρήστη διεπαφή που επιτρέπει στους χρήστες να ορίζουν τις παραμέτρους ανωνυμοποίησης και να τις εφαρμόζουν στα δεδομένα σε ένα μόνο βήμα. Περιλαμβάνει επίσης μια σειρά από προηγμένα χαρακτηριστικά, όπως τη δυνατότητα χειρισμού διανομών δεδομένων, ιεραρχίες δεδομένων και σχέσεις δεδομένων.

Συμπερασματικά, το ARX είναι ένα ισχυρό εργαλείο λογισμικού ανωνυμοποίησης δεδομένων που μπορεί να χρησιμοποιηθεί για την προστασία του απορρήτου των ατόμων στα δεδομένα, αποκρύπτοντας τις τιμές των ευαίσθητων χαρακτηριστικών. Υποστηρίζει k-ανωνυμία, l-diversity, t-closeness και διαφορικό απόρρητο. Το λογισμικό μπορεί να χρησιμοποιηθεί για την ανωνυμοποίηση ενός ευρέος φάσματος τύπων δεδομένων και μπορεί να εφαρμοστεί σε δεδομένα από διαφορετικές πηγές. Η φιλική προς τον χρήστη διεπαφή και οι προηγμένες λειτουργίες διευκολύνουν τους οργανισμούς να συμμορφωθούν με τους κανονισμούς προστασίας δεδομένων.

2.7 Πώς η k-anonimity και η l-diversity υλοποιούνται στο ARX

Η K-ανωνυμία και η l-diversity είναι δύο σημαντικές τεχνικές που μπορούν να χρησιμοποιηθούν για την προστασία του απορρήτου των ατόμων στα δεδομένα. Και οι δύο αυτές τεχνικές μπορούν να εφαρμοστούν χρησιμοποιώντας το εργαλείο λογισμικού ARX.

Η K-ανωνυμία υλοποιείται στο ARX με τον εντοπισμό ευαίσθητων δεδομένων μέσα σε ένα σύνολο δεδομένων και στη συνέχεια με την εφαρμογή διαφορετικών τεχνικών ανωνυμοποίησης για την απόκρυψη των τιμών αυτών των χαρακτηριστικών, επιτρέποντας παράλληλα τη χρήση των δεδομένων για ανάλυση. Αυτές οι τεχνικές περιλαμβάνουν τη γενίκευση, την καταστολή και την προσθήκη θορύβου. Η γενίκευση περιλαμβάνει τη μείωση του επιπέδου λεπτομέρειας στα δεδομένα με τη συγκέντρωση δεδομένων σε υψηλότερο επίπεδο ή με την αφαίρεση ορισμένων χαρακτηριστικών. Η καταστολή περιλαμβάνει την αφαίρεση ορισμένων τιμών ή εγγραφών από το σύνολο δεδομένων, εάν καθιστούν δυνατή την αναγνώριση ενός ατόμου με έναν μοναδικό συνδυασμό τιμών. Η προσθήκη θορύβου περιλαμβάνει την προσθήκη τυχαίου θορύβου στα δεδομένα για να κρύψει τις πραγματικές τιμές των χαρακτηριστικών, καθιστώντας πιο δύσκολο τον εντοπισμό ατόμων με έναν μοναδικό συνδυασμό τιμών. Η k-ανωνυμία εφαρμόζεται ορίζοντας ένα όριο "k" για τον αριθμό των εγγραφών που θα πρέπει να μοιράζονται τις ίδιες τιμές για τα ευαίσθητα χαρακτηριστικά.

Το L-diversity υλοποιείται στο ARX διασφαλίζοντας ότι για κάθε ευαίσθητη τιμή, υπάρχουν τουλάχιστον l καλά αντιπροσωπευόμενες, διακριτές, ευαίσθητες τιμές στην κλάση ισοδυναμίας (η ομάδα εγγραφών που μοιράζονται την ίδια τιμή για το ευαίσθητο χαρακτηριστικό). Αυτό σημαίνει ότι για κάθε ευαίσθητη τιμή, υπάρχουν τουλάχιστον l διαφορετικές τιμές στην ομάδα εγγραφών, γεγονός που καθιστά δύσκολο τον προσδιορισμό οποιασδήποτε μεμονωμένης εγγραφής με έναν

μοναδικό συνδυασμό τιμών. Η l-diversity εφαρμόζεται ορίζοντας ένα όριο "l" για τον αριθμό των διακριτών τιμών που πρέπει να υπάρχουν στην ομάδα εγγραφών που μοιράζονται την ίδια τιμή για το ευαίσθητο χαρακτηριστικό.

Το ARX επιτρέπει στον χρήστη να καθορίσει τις παραμέτρους για την k-ανωνυμία και την l-diversity σε μια φιλική προς το χρήστη διεπαφή και εφαρμόζει αυτές τις τεχνικές στα δεδομένα σε ένα μόνο βήμα. Περιλαμβάνει επίσης μια σειρά από προηγμένα χαρακτηριστικά, όπως τη δυνατότητα χειρισμού διανομών δεδομένων, ιεραρχίες δεδομένων και σχέσεις δεδομένων. Το λογισμικό μπορεί να χρησιμοποιηθεί για την ανωνυμοποίηση ενός ευρέος φάσματος τύπων δεδομένων, συμπεριλαμβανομένων δομημένων δεδομένων, ημιδομημένων δεδομένων και μη δομημένων δεδομένων. Μπορεί επίσης να χρησιμοποιηθεί για την ανωνυμοποίηση δεδομένων από διαφορετικές πηγές, όπως βάσεις δεδομένων, υπολογιστικά φύλλα και αρχεία κειμένου.

2.8 Προστασία χωρικών δεδομένων και κίνδυνοι

Η προστασία των χωρικών δεδομένων είναι μια σημαντική πτυχή του απορρήτου των δεδομένων, καθώς αφορά τις πληροφορίες που βασίζονται στην τοποθεσία που συλλέγονται και αποθηκεύονται από οργανισμούς. Αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν για την παρακολούθηση των κινήσεων και των δραστηριοτήτων των ατόμων, αποκαλύπτοντας ευαίσθητες πληροφορίες για την προσωπική τους ζωή. Ως εκ τούτου, είναι σημαντικό να προστατεύονται τα χωρικά δεδομένα προκειμένου να διατηρηθεί το απόρρητο των ατόμων.

Ένας από τους κύριους κινδύνους που συνδέονται με τα χωρικά δεδομένα είναι η δυνατότητα επαναπροσδιορισμού. Αυτό συμβαίνει όταν ένας εισβολέας μπορεί να συνδέσει ανώνυμα δεδομένα με ένα άτομο, χρησιμοποιώντας πρόσθετες πληροφορίες. Για παράδειγμα, εάν ένας εισβολέας έχει πρόσβαση στο προφίλ ενός ατόμου στα μέσα κοινωνικής δικτύωσης, μπορεί να μπορεί να χρησιμοποιήσει τα δεδομένα τοποθεσίας από το προφίλ για να αναγνωρίσει εκ νέου το άτομο σε ένα ανώνυμο σύνολο δεδομένων. Αυτό μπορεί να οδηγήσει σε σοβαρή παραβίαση του απορρήτου, καθώς ο εισβολέας μπορεί να αποκτήσει πρόσβαση σε ευαίσθητες πληροφορίες σχετικά με το άτομο, όπως τα ταξιδιωτικά του μοτίβα, προσωπικές σχέσεις και πληροφορίες υγείας.

Ένας άλλος κίνδυνος που σχετίζεται με τα χωρικά δεδομένα είναι η δυνατότητα εντοπισμού τοποθεσίας. Αυτό συμβαίνει όταν ένας εισβολέας χρησιμοποιεί δεδομένα τοποθεσίας για να παρακολουθεί τις κινήσεις ενός ατόμου με την πάροδο του χρόνου. Αυτό μπορεί να γίνει παρακολουθώντας την τοποθεσία ενός ατόμου μέσω της κινητής συσκευής του ή παρακολουθώντας τις κινήσεις του μέσω της χρήσης καμερών ή άλλων τεχνολογιών παρακολούθησης. Αυτό μπορεί να οδηγήσει σε σοβαρή παραβίαση του απορρήτου, καθώς ο εισβολέας μπορεί να αποκτήσει πρόσβαση σε ευαίσθητες πληροφορίες σχετικά με το άτομο, όπως τα ταξιδιωτικά του μοτίβα, προσωπικές σχέσεις και πληροφορίες υγείας.

Για την προστασία των χωρικών δεδομένων, οι οργανισμοί πρέπει να κάνουν μια σειρά από βήματα για την ανωνυμοποίηση και την κρυπτογράφηση των δεδομένων. Μια προσέγγιση είναι η χρήση τεχνικών όπως η k-ανωνυμία και η l-ποικιλομορφία, οι οποίες στοχεύουν να αποκρύψουν τις τιμές των ευαίσθητων χαρακτηριστικών ενώ παράλληλα επιτρέπουν στα δεδομένα να

χρησιμοποιηθούν για ανάλυση. Μια άλλη προσέγγιση είναι η χρήση τεχνικών χωρικής κάλυψης, οι οποίες περιλαμβάνουν την προσθήκη τυχαίου θορύβου στα δεδομένα για να κρύψουν τις πραγματικές τιμές των χαρακτηριστικών.

Ένας άλλος τρόπος για την προστασία των χωρικών δεδομένων είναι η χρήση ασφαλών πρωτοκόλλων επικοινωνίας, όπως το ασφαλές στρώμα υποδοχής (SSL) και η ασφάλεια επιπέδου μεταφοράς (TLS) για την κρυπτογράφηση της μετάδοσης δεδομένων. Αυτό είναι σημαντικό γιατί θα αποτρέψει μη εξουσιοδοτημένα μέρη από την υποκλοπή και την πρόσβαση στα δεδομένα κατά τη μετάδοση.

Επιπλέον, οι οργανισμοί θα πρέπει να εφαρμόζουν αυστηρούς ελέγχους πρόσβασης για να διασφαλίζουν ότι μόνο εξουσιοδοτημένο προσωπικό έχει πρόσβαση σε ευαίσθητα χωρικά δεδομένα. Αυτό μπορεί να περιλαμβάνει την εφαρμογή μηχανισμών ελέγχου ταυτότητας και εξουσιοδότησης, καθώς και την εφαρμογή ελέγχων πρόσβασης βάσει ρόλων για να διασφαλιστεί ότι μόνο όσοι χρειάζονται πρόσβαση στα δεδομένα μπορούν να έχουν πρόσβαση σε αυτά.

Συμπερασματικά, η προστασία των χωρικών δεδομένων είναι σημαντική για τη διατήρηση της ιδιωτικής ζωής των ατόμων. Οι κίνδυνοι που σχετίζονται με τα χωρικά δεδομένα περιλαμβάνουν τη δυνατότητα επαναπροσδιορισμού και εντοπισμού θέσης. Για την προστασία των χωρικών δεδομένων, οι οργανισμοί θα πρέπει να λαμβάνουν ορισμένα βήματα για την ανωνυμοποίηση και την κρυπτογράφηση των δεδομένων, καθώς και να εφαρμόζουν αυστηρούς ελέγχους πρόσβασης για να διασφαλίζουν ότι μόνο εξουσιοδοτημένο προσωπικό έχει πρόσβαση σε ευαίσθητα δεδομένα.

3

Ανωνυμία και τεχνικές απόκρυψης

3.1 Ανωνυμία σε Υπηρεσίες Εντοπισμού Θέσης (Anonymity in Location-Based Services)

Η ανωνυμία στις υπηρεσίες που βασίζονται στην τοποθεσία (LBS) είναι ένα σημαντικό ζήτημα, καθώς το LBS βασίζεται συχνά στη συλλογή και ανάλυση δεδομένων τοποθεσίας για την παροχή υπηρεσιών όπως η πλοήγηση, η διαφήμιση βάσει τοποθεσίας και η κοινωνική δικτύωση βάσει τοποθεσίας. Ωστόσο, αυτή η συλλογή και ανάλυση δεδομένων τοποθεσίας μπορεί επίσης να θέσει σε σημαντικό κίνδυνο το απόρρητο, καθώς τα δεδομένα τοποθεσίας των ατόμων μπορούν να χρησιμοποιηθούν για την εκ νέου ταυτοποίησή τους και την εκμάθηση ευαίσθητων πληροφοριών σχετικά με τις δραστηριότητες και τις κινήσεις τους.

Ένας τρόπος προστασίας της ανωνυμίας στο LBS είναι μέσω της χρήσης της ψευδωνυμίας. Αυτή είναι η πρακτική της αντικατάστασης της πραγματικής ταυτότητας ενός ατόμου με ένα ψευδώνυμο ή ένα ψεύτικο όνομα, προκειμένου να προστατευθεί το απόρρητό του. Για παράδειγμα, ένας χρήστης μιας εφαρμογής κοινωνικής δικτύωσης που βασίζεται σε τοποθεσία θα μπορούσε να χρησιμοποιήσει ένα ψευδώνυμο αντί για το πραγματικό του όνομα όταν δημιουργεί ένα προφίλ και μοιράζεται δεδομένα τοποθεσίας.

Ένας άλλος τρόπος προστασίας της ανωνυμίας στο LBS είναι μέσω της χρήσης της συγκέντρωσης και γενίκευσης δεδομένων. Αυτό περιλαμβάνει τη συλλογή και ανάλυση δεδομένων τοποθεσίας σε υψηλό επίπεδο αφαίρεσης, όπως σε επίπεδο πόλης ή γειτονιάς, παρά στο επίπεδο της

ακριβούς τοποθεσίας ενός ατόμου. Για παράδειγμα, ένα LBS θα μπορούσε να παρέχει πληροφορίες για τον καιρό για μια συγκεκριμένη πόλη, αντί για μια συγκεκριμένη διεύθυνση.

Μια άλλη τεχνική είναι η χρήση του διαφορικού απορρήτου, είναι ένα μαθηματικό πλαίσιο για την προστασία του απορρήτου στην ανάλυση δεδομένων προσθέτοντας προσεκτικά επιλεγμένο θόρυβο στα δεδομένα, με αυτόν τον τρόπο οι ιδιωτικές πληροφορίες προστατεύονται ενώ παράλληλα επιτρέπει την εκτέλεση χρήσιμης ανάλυσης.

Επιπλέον, οι πάροχοι LBS μπορούν επίσης να εφαρμόσουν τεχνικές όπως k-anonymity, l-diversity και t-closeness για να προστατεύσουν την ανωνυμία των χρηστών τους. Αυτές οι τεχνικές περιλαμβάνουν την ομαδοποίηση των χρηστών σε σύνολα ανωνυμίας, τη διασφάλιση ότι οι ευαίσθητες πληροφορίες δεν είναι πολύ παρόμοιες εντός μιας ομάδας και την εισαγωγή ενός ορίου για τη μέγιστη επιτρεπόμενη διαφορά μεταξύ των τιμών των ευαίσθητων χαρακτηριστικών στο σύνολο δεδομένων και εκείνων του πληθυσμού.

Συμπερασματικά, η προστασία της ανωνυμίας στις υπηρεσίες που βασίζονται στην τοποθεσία είναι σημαντική για να διασφαλιστεί ότι τα δεδομένα τοποθεσίας των ατόμων δεν χρησιμοποιούνται για την εκ νέου ταυτοποίησή τους και την εκμάθηση ευαίσθητων πληροφοριών σχετικά με τις δραστηριότητες και τις κινήσεις τους. Τεχνικές όπως ψευδωνυμία, συγκέντρωση και γενίκευση δεδομένων, διαφορικό απόρρητο και k-ανωνυμία, l-diversity και t-closeness μπορούν να χρησιμοποιηθούν για την προστασία της ανωνυμίας των χρηστών σε υπηρεσίες που βασίζονται στην τοποθεσία.

3.2 Οι τεχνικές της απόκρυψης (Cloaking Techniques)

Οι τεχνικές απόκρυψης είναι μέθοδοι που χρησιμοποιούνται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία (LBS) αφαιρώντας ή αποκρύπτοντας ευαίσθητα δεδομένα τοποθεσίας. Μία από τις πιο συχνά χρησιμοποιούμενες τεχνικές cloaking είναι η k-anonymity.

Η K-anonymity είναι μια τεχνική διατήρησης του απορρήτου που χρησιμοποιείται για την προστασία ευαίσθητων δεδομένων διασφαλίζοντας ότι κάθε άτομο σε ένα σύνολο δεδομένων δεν μπορεί να επαναπροσδιοριστεί από οποιοδήποτε συνδυασμό ευαίσθητων χαρακτηριστικών. Λειτουργεί ομαδοποιώντας τα άτομα σε ένα σύνολο δεδομένων σε "σύνολα ανωνυμίας" με τουλάχιστον k άτομα, όπου το k είναι μια παράμετρος που καθορίζεται από τον χρήστη. Για παράδειγμα, εάν k=3, αυτό θα σήμαινε ότι κάθε άτομο στο σύνολο δεδομένων είναι μέρος μιας ομάδας τουλάχιστον 3 άλλων ατόμων με παρόμοια ευαίσθητα χαρακτηριστικά, γεγονός που καθιστά δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου οποιοδήποτε άτομο με βάση τα ευαίσθητα χαρακτηριστικά του μόνος.

Στο πλαίσιο του LBS, η k-ανωνυμία μπορεί να εφαρμοστεί σε δεδομένα τοποθεσίας γενικεύοντας ή διαταράσσοντας τα δεδομένα τοποθεσίας σε ένα ορισμένο επίπεδο ευαισθησίας. Για παράδειγμα, αντί να παρέχεται η ακριβής τοποθεσία ενός χρήστη, η τοποθεσία θα μπορούσε να γενικευτεί στο επίπεδο μιας πόλης ή μιας γειτονιάς. Αυτό θα καθιστούσε πιο δύσκολο για έναν εισβολέα να αναγνωρίσει ξανά τον χρήστη με βάση μόνο τα δεδομένα τοποθεσίας του.

Μια άλλη τεχνική που χρησιμοποιείται στο cloaking είναι η χρήση της τυχαιοποιημένης απόκρισης, είναι μια τεχνική που επιτρέπει στον χρήστη να απαντήσει σε μια ευαίσθητη ερώτηση με ειλικρίνεια ή όχι ειλικρινή με ορισμένες πιθανότητες, με αυτόν τον τρόπο η αληθινή απάντηση προστατεύεται αλλά εξακολουθεί να επιτρέπει στα δεδομένα να είναι χρήσιμα για την ανάλυση .

Επιπλέον, οι πάροχοι LBS μπορούν επίσης να χρησιμοποιήσουν τεχνικές όπως l-diversity και t-closeness για την περαιτέρω προστασία της ανωνυμίας των χρηστών τους. Αυτές οι τεχνικές περιλαμβάνουν τη διασφάλιση ότι οι ευαίσθητες πληροφορίες δεν είναι πολύ παρόμοιες μέσα σε μια ομάδα και την εισαγωγή ενός ορίου για τη μέγιστη επιτρεπόμενη διαφορά μεταξύ των τιμών των ευαίσθητων χαρακτηριστικών στο σύνολο δεδομένων και εκείνων του πληθυσμού.

3.2.1 Απόκρυψη βασιζόμενη στα δεδομένα (Data-based cloaking)

Η απόκρυψη που βασίζεται σε δεδομένα αναφέρεται σε ένα σύνολο τεχνικών που χρησιμοποιούνται για την ανωνυμοποίηση ευαίσθητων δεδομένων αφαιρώντας ή αποκρύπτοντας τα δεδομένα με τρόπο που διατηρεί τη χρησιμότητά τους για ανάλυση. Αυτό έρχεται σε αντίθεση με το "απόκρυψη απόκρυψης βάσει μοντέλου" που βασίζεται σε ένα μοντέλο διανομής δεδομένων για τη δημιουργία συνθετικών δεδομένων που διατηρούν ορισμένες ιδιότητες των αρχικών δεδομένων.

Μία από τις πιο κοινές τεχνικές απόκρυψης που βασίζονται σε δεδομένα είναι η k-anonymity, η οποία λειτουργεί ομαδοποιώντας άτομα σε ένα σύνολο δεδομένων σε "σύνολα ανωνυμίας" τουλάχιστον k ατόμων, όπου το k είναι μια παράμετρος που καθορίζεται από τον χρήστη. Για παράδειγμα, εάν k=3, αυτό θα σήμαινε ότι κάθε άτομο στο σύνολο δεδομένων είναι μέρος μιας ομάδας τουλάχιστον 3 άλλων ατόμων με παρόμοια ευαίσθητα χαρακτηριστικά, γεγονός που καθιστά δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου οποιοδήποτε άτομο με βάση τα ευαίσθητα χαρακτηριστικά του μόνος.

Μια άλλη τεχνική απόκρυψης που βασίζεται σε δεδομένα είναι η l-diversity, η οποία λειτουργεί διασφαλίζοντας ότι οι ευαίσθητες πληροφορίες δεν είναι πολύ παρόμοιες μέσα σε μια ομάδα και εισάγοντας ένα όριο για τη μέγιστη επιτρεπόμενη διαφορά μεταξύ των τιμών των ευαίσθητων χαρακτηριστικών στο σύνολο δεδομένων και εκείνων του πληθυσμού. Αυτή η προσέγγιση είναι χρήσιμη όταν υπάρχουν ευαίσθητα χαρακτηριστικά που είναι κατηγορηματικά και ο στόχος είναι να διασφαλιστεί ότι τα δεδομένα δεν είναι πολύ προβλέψιμα.

Οι τεχνικές απόκρυψης που βασίζονται σε δεδομένα περιλαμβάνουν επίσης το t-closeness, το οποίο είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που λειτουργεί διασφαλίζοντας ότι η κατανομή πιθανοτήτων των ευαίσθητων χαρακτηριστικών στα δεδομένα είναι παρόμοια με αυτή στον πληθυσμό.

Μια άλλη τεχνική είναι η τυχαιοποιημένη απόκριση, επιτρέπει στον χρήστη να απαντήσει σε μια ευαίσθητη ερώτηση με ειλικρίνεια ή όχι ειλικρινά με συγκεκριμένη πιθανότητα, με αυτόν τον τρόπο η αληθινή απάντηση προστατεύεται αλλά εξακολουθεί να επιτρέπει στα δεδομένα να είναι χρήσιμα για την ανάλυση.

Οι τεχνικές απόκρυψης βάσει δεδομένων μπορούν να εφαρμοστούν σε δεδομένα τοποθεσίας γενικεύοντας ή διαταράσσοντας τα δεδομένα τοποθεσίας σε ένα ορισμένο επίπεδο ευαισθησίας. Για παράδειγμα, αντί να παρέχεται η ακριβής τοποθεσία ενός χρήστη, η τοποθεσία θα μπορούσε να

γενικευτεί στο επίπεδο μιας πόλης ή μιας γειτονιάς. Αυτό θα καθιστούσε πιο δύσκολο για έναν εισβολέα να αναγνωρίσει ξανά τον χρήστη με βάση μόνο τα δεδομένα τοποθεσίας του.

Συνοπτικά, τεχνικές απόκρυψης που βασίζονται σε δεδομένα, όπως k-anonymity, l-diversity, t-closeness, τυχαιοποιημένη απόκριση και γενίκευση ή διαταραχή χρησιμοποιούνται για την ανωνυμοποίηση ευαίσθητων δεδομένων αφαιρώντας ή αποκρύπτοντας τα δεδομένα με τρόπο που διατηρεί τη χρησιμότητά τους για ανάλυση . Αυτές οι τεχνικές βοηθούν στην προστασία ευαίσθητων δεδομένων καθιστώντας πιο δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου τα άτομα με βάση μόνο τα ευαίσθητα χαρακτηριστικά τους.

3.2.2 Απόκρυψη βασιζόμενη στο χώρο (Space-dependant cloaking).

Η απόκρυψη που εξαρτάται από το διάστημα είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που λαμβάνει υπόψη τις χωρικές εξαρτήσεις των δεδομένων τοποθεσίας προκειμένου να προστατεύσει την ανωνυμία των ατόμων. Αυτή η προσέγγιση είναι ιδιαίτερα σημαντική για τις υπηρεσίες που βασίζονται στην τοποθεσία (LBS), όπου ο στόχος είναι να αποτραπεί ο επαναπροσδιορισμός των ατόμων με βάση μόνο τα δεδομένα τοποθεσίας τους.

Μια προσέγγιση για την απόκρυψη που εξαρτάται από το διάστημα είναι η χρήση της χωρικής γενίκευσης, η οποία λειτουργεί με τη γενίκευση των δεδομένων τοποθεσίας σε ένα υψηλότερο επίπεδο ευαισθησίας. Για παράδειγμα, αντί να παρέχεται η ακριβής τοποθεσία ενός χρήστη, η τοποθεσία θα μπορούσε να γενικευτεί στο επίπεδο μιας πόλης ή μιας γειτονιάς. Αυτό θα καθιστούσε πιο δύσκολο για έναν εισβολέα να αναγνωρίσει ξανά τον χρήστη με βάση μόνο τα δεδομένα τοποθεσίας του.

Μια άλλη προσέγγιση είναι η χρήση της χωρικής διαταραχής, η οποία λειτουργεί προσθέτοντας τυχαίο θόρυβο στα δεδομένα τοποθεσίας. Αυτό μπορεί να γίνει προσθέτοντας μια τυχαία μετατόπιση στις συντεταγμένες τοποθεσίας ή επιλέγοντας μια τυχαία τοποθεσία σε μια συγκεκριμένη ακτίνα από την πραγματική θέση. Αυτή η προσέγγιση βοηθά στην προστασία της ανωνυμίας των ατόμων, καθιστώντας πιο δύσκολο για έναν εισβολέα να τα αναγνωρίσει ξανά με βάση τα δεδομένα τοποθεσίας και μόνο.

Μια τρίτη προσέγγιση είναι η χρήση της χωρικής συνάθροισης, η οποία λειτουργεί ομαδοποιώντας τα άτομα με βάση τα δεδομένα τοποθεσίας τους και στη συνέχεια εφαρμόζοντας τεχνικές ανωνυμίας, όπως η k-ανωνυμία ή η l-ποικιλομορφία στις ομάδες. Αυτή η προσέγγιση βοηθά στην προστασία της ανωνυμίας των ατόμων, καθιστώντας πιο δύσκολο για έναν εισβολέα να τα αναγνωρίσει ξανά με βάση τα δεδομένα τοποθεσίας και μόνο.

Συνοπτικά, η απόκρυψη που εξαρτάται από το διάστημα είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που λαμβάνει υπόψη τις χωρικές εξαρτήσεις των δεδομένων τοποθεσίας προκειμένου να προστατεύσει την ανωνυμία των ατόμων. Αυτή η προσέγγιση είναι ιδιαίτερα σημαντική για υπηρεσίες που βασίζονται σε τοποθεσία και μπορεί να επιτευχθεί μέσω της χωρικής γενίκευσης, της χωρικής διαταραχής και της χωρικής συνάθροισης.

3.2.3 Χωρική k-ανωνυμία (Location k-anonymity)

Οι προσεγγίσεις της χωρικής k-ανωνυμίας, προστατεύουν την ιδιωτικότητα των χρηστών χρησιμοποιώντας την τρέχουσα θέση του κάθε χρήστη στο σύστημα. Οι πιο δημοφιλείς προσεγγίσεις είναι οι εξής:

3.2.3.1 Η απόκρυψη κλίκας (clique cloak)

Το Clique cloak είναι μια τεχνική διατήρησης απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία (LBS). Η βασική ιδέα πίσω από τον μανδύα της κλίκας είναι να χωριστούν τα δεδομένα τοποθεσίας σε ομάδες, που ονομάζονται κλίκες, και στη συνέχεια να εφαρμοστούν τεχνικές ανωνυμίας, όπως η k-anonymity ή η l-diversity στις ομάδες.

Η τεχνική clique cloak λειτουργεί προσδιορίζοντας ομάδες τοποθεσιών που είναι χωρικά κοντά η μία στην άλλη και που είναι πιθανό να τις επισκεφτεί το ίδιο σύνολο ατόμων. Αυτές οι ομάδες ονομάζονται κλίκες. Στη συνέχεια, τα δεδομένα τοποθεσίας για κάθε άτομο διαταράσσονται, έτσι ώστε να εμπίπτουν στην ίδια κλίκα με την πραγματική τοποθεσία του ατόμου. Αυτό καθιστά πιο δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου άτομα με βάση μόνο τα δεδομένα τοποθεσίας τους.

Το κύριο πλεονέκτημα του clique cloak είναι ότι μπορεί να διατηρήσει την ανωνυμία των ατόμων, ενώ εξακολουθεί να παρέχει δεδομένα τοποθεσίας που είναι αρκετά ακριβή ώστε να είναι χρήσιμα για το LBS. Ο μανδύας Clique επιτρέπει επίσης τη διατήρηση των χωρικών εξαρτήσεων στα δεδομένα τοποθεσίας, κάτι που είναι σημαντικό για πολλές εφαρμογές LBS.

Θα πρέπει να σημειωθεί ότι, όπως με κάθε τεχνική ανωνυμοποίησης, το επίπεδο προστασίας που παρέχεται από τον μανδύα κλίκ εξαρτάται από τις συγκεκριμένες παραμέτρους που χρησιμοποιούνται, όπως το μέγεθος των κλίκων, το επίπεδο διαταραχής που εφαρμόζεται και την τεχνική ανωνυμίας που χρησιμοποιείται. Επίσης, όπως κάθε τεχνική που βασίζεται στις χωρικές εξαρτήσεις, θα μπορούσε να επηρεαστεί από τη σπανιότητα των δεδομένων σε ορισμένες περιοχές ή τη συγκέντρωση ατόμων σε ορισμένα μέρη.

Συνοπτικά, το Clique cloak είναι μια τεχνική διατήρησης απορρήτου που λειτουργεί διαμερίζοντας τα δεδομένα τοποθεσίας σε ομάδες, που ονομάζονται κλίκες, και στη συνέχεια εφαρμόζοντας τεχνικές ανωνυμίας, όπως k-anonymity ή l-diversity στις ομάδες. Αυτή η τεχνική επιτρέπει τη διατήρηση των χωρικών εξαρτήσεων στα δεδομένα τοποθεσίας, κάτι που είναι σημαντικό για πολλές εφαρμογές LBS, ενώ εξακολουθεί να παρέχει ένα επίπεδο προστασίας για το απόρρητο των ατόμων.

3.2.3.2 Η κεντρική απόκρυψη (center cloak)

Ο κεντρικός μανδύας είναι μια άλλη τεχνική διατήρησης του απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία (LBS). Η βασική ιδέα πίσω από τον κεντρικό μανδύα είναι να προσδιορίσει ένα σύνολο κεντρικών

τοποθεσιών, που ονομάζονται κέντρα, και στη συνέχεια να διαταράξει τα δεδομένα τοποθεσίας κάθε ατόμου στο πλησιέστερο κέντρο.

Η τεχνική του κεντρικού μανδύα λειτουργεί προσδιορίζοντας ένα σύνολο κεντρικών τοποθεσιών, που ονομάζονται κέντρα, που είναι πιθανό να επισκέπτονται μεγάλος αριθμός ατόμων. Στη συνέχεια, τα δεδομένα τοποθεσίας για κάθε άτομο διαταράσσονται, έτσι ώστε να πέφτει σε μια ορισμένη απόσταση από το πλησιέστερο κέντρο. Αυτό καθιστά πιο δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου άτομα με βάση μόνο τα δεδομένα τοποθεσίας τους.

Το κύριο πλεονέκτημα του κεντρικού μανδύα είναι ότι είναι υπολογιστικά αποδοτικός και εύκολος στην εφαρμογή του. Παρέχει επίσης υψηλό επίπεδο προστασίας για το απόρρητο των ατόμων, καθώς μειώνει τη χωρική ανάλυση των δεδομένων τοποθεσίας και καθιστά πιο δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου τα άτομα.

Θα πρέπει να σημειωθεί ότι το επίπεδο προστασίας που παρέχεται από τον κεντρικό μανδύα εξαρτάται από τις συγκεκριμένες παραμέτρους που χρησιμοποιούνται, όπως ο αριθμός των κέντρων, το μέγεθος της ακτίνας διαταραχής και η μετρική απόστασης που χρησιμοποιείται. Επίσης, όπως κάθε τεχνική που βασίζεται στις χωρικές εξαρτήσεις, θα μπορούσε να επηρεαστεί από τη σπανιότητα των δεδομένων σε ορισμένες περιοχές ή τη συγκέντρωση ατόμων σε ορισμένα μέρη.

Συνοπτικά, το Center cloak είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που λειτουργεί προσδιορίζοντας ένα σύνολο κεντρικών τοποθεσιών και διαταράσσοντας τα δεδομένα τοποθεσίας κάθε ατόμου στο πλησιέστερο κέντρο. Αυτή η τεχνική είναι υπολογιστικά αποδοτική και εύκολη στην εφαρμογή, μειώνει τη χωρική ανάλυση των δεδομένων τοποθεσίας, καθιστώντας πιο δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου άτομα και παρέχει υψηλό επίπεδο προστασίας για το απόρρητο των ατόμων.

3.2.3.3 Η τεχνική της απόκρυψης του κοντινότερου γείτονα (NN-Cloak)

Το NN-Cloak είναι μια τεχνική διατήρησης απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία (LBS). Η βασική ιδέα πίσω από το NN-Cloak είναι να αντικατασταθεί η ακριβής τοποθεσία ενός ατόμου με τη θέση του πλησιέστερου γειτονικού ατόμου.

Η τεχνική NN-Cloak λειτουργεί προσδιορίζοντας πρώτα τον πλησιέστερο γείτονα για κάθε άτομο σε ένα σύνολο δεδομένων. Στη συνέχεια, η ακριβής τοποθεσία κάθε ατόμου αντικαθίσταται από την τοποθεσία του πλησιέστερου γείτονά του. Αυτό καθιστά πιο δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου άτομα με βάση μόνο τα δεδομένα τοποθεσίας τους.

Το κύριο πλεονέκτημα του NN-Cloak είναι ότι είναι υπολογιστικά αποδοτικό και εύκολο στην εφαρμογή του. Παρέχει επίσης υψηλό επίπεδο προστασίας για το απόρρητο των ατόμων, καθώς μειώνει τη χωρική ανάλυση των δεδομένων τοποθεσίας και καθιστά πιο δύσκολο για έναν εισβολέα να αναγνωρίσει εκ νέου άτομα.

Θα πρέπει να σημειωθεί ότι το επίπεδο προστασίας που παρέχεται από το NN-Cloak εξαρτάται από τις συγκεκριμένες παραμέτρους που χρησιμοποιούνται, όπως η μετρική απόστασης που χρησιμοποιείται και η αραιότητα των δεδομένων σε ορισμένες περιοχές ή η συγκέντρωση ατόμων σε ορισμένα μέρη.

Ένα άλλο μειονέκτημα του NN-Cloak είναι ότι μπορεί να μην είναι αποτελεσματικό σε περιπτώσεις όπου η πυκνότητα των ατόμων είναι χαμηλή, καθώς μπορεί να μην υπάρχει αρκετά μεγάλος αριθμός πλησιέστερων γειτόνων για να αντικαταστήσει την ακριβή τοποθεσία ενός ατόμου.

Συνοπτικά, το NN-Cloak είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που λειτουργεί αντικαθιστώντας την ακριβή τοποθεσία ενός ατόμου με τη θέση του πλησιέστερου γειτονικού ατόμου. Αυτή η τεχνική είναι υπολογιστικά αποτελεσματική και εύκολη στην εφαρμογή, μειώνει τη χωρική ανάλυση των δεδομένων τοποθεσίας, καθιστώντας πιο δύσκολο για έναν εισβολέα να αναγνωρίσει ξανά άτομα και παρέχει υψηλό επίπεδο προστασίας για το απόρρητο των ατόμων.

3.2.3.4 Η τεχνική Casper (Clustering-based Anonymity using SPatially-bounded Errors)

Το CASPER (Ανωνυμία βασισμένη σε ομαδοποίηση με χρήση σφαλμάτων χωρικά περιορισμένων) είναι μια τεχνική διατήρησης του απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία (LBS). Η βασική ιδέα πίσω από το CASPER είναι να χρησιμοποιήσει αλγόριθμους ομαδοποίησης για να ομαδοποιήσει άτομα με βάση τα δεδομένα τοποθεσίας τους και στη συνέχεια να προσθέσει θόρυβο στα δεδομένα τοποθεσίας κάθε ομάδας προκειμένου να διατηρηθεί το απόρρητο των ατόμων.

Το CASPER λειτουργεί ομαδοποιώντας πρώτα τα άτομα σε ένα σύνολο δεδομένων με βάση τα δεδομένα τοποθεσίας τους. Ο αλγόριθμος ομαδοποίησης που χρησιμοποιείται βασίζεται συνήθως στην πυκνότητα των ατόμων στην περιοχή και μπορεί να προσαρμοστεί ώστε να ταιριάζει στις συγκεκριμένες απαιτήσεις του συνόλου δεδομένων. Μόλις τα άτομα ομαδοποιηθούν σε συμπλέγματα, προστίθεται θόρυβος στα δεδομένα τοποθεσίας κάθε ομάδας προκειμένου να διατηρηθεί το απόρρητο των ατόμων.

Ο θόρυβος που προστίθεται στα δεδομένα τοποθεσίας μπορεί να είναι είτε χωρικός είτε χρονικός. Ο χωρικός θόρυβος αναφέρεται στην προσθήκη τυχαίου θορύβου στα δεδομένα τοποθεσίας με τη μορφή μικρών μετατοπίσεων στην πραγματική θέση. Ο προσωρινός θόρυβος αναφέρεται στην προσθήκη τυχαίου θορύβου στα δεδομένα τοποθεσίας μετατοπίζοντας τις χρονικές σημάνσεις των δεδομένων τοποθεσίας κατά μικρές ποσότητες.

Το κύριο πλεονέκτημα του CASPER είναι ότι είναι σε θέση να διατηρήσει το απόρρητο των ατόμων, ενώ εξακολουθεί να επιτρέπει τη διεξαγωγή χρήσιμης ανάλυσης βάσει τοποθεσίας. Επιτρέπει επίσης την προστασία του απορρήτου σε σύνολα δεδομένων με ποικίλες χωρικές πυκνότητες.

Ωστόσο, θα πρέπει να σημειωθεί ότι το CASPER δεν εγγυάται ότι η τοποθεσία των ατόμων δεν μπορεί να επαναπροσδιοριστεί και το επίπεδο απορρήτου εξαρτάται από τις παραμέτρους και το όριο που επιλέγονται για ομαδοποίηση και την ποσότητα του προστιθέμενου θορύβου.

Συνοπτικά, το CASPER (Clustering-based Anonymity using SPatially-bounded Errors) είναι μια τεχνική διατήρησης απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία, χρησιμοποιώντας αλγόριθμους ομαδοποίησης για την ομαδοποίηση ατόμων με βάση τα δεδομένα τοποθεσίας τους και στη συνέχεια προσθέτοντας θόρυβο σε τα δεδομένα τοποθεσίας κάθε ομάδας προκειμένου να διατηρηθεί το απόρρητο των ατόμων. Επιτρέπει την προστασία του απορρήτου σε σύνολα δεδομένων με ποικίλες χωρικές πυκνότητες,

αλλά δεν εγγυάται πλήρη προστασία και η απόδοσή του εξαρτάται από τις παραμέτρους και το όριο που επιλέγονται.

3.2.3.5 Η τεχνική της απόκρυψης διαστήματος (interval cloak)

Το Interval cloak είναι μια τεχνική διατήρησης απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία (LBS) καλύπτοντας τα δεδομένα τοποθεσίας τους χρησιμοποιώντας διαστήματα. Η βασική ιδέα πίσω από τον μανδύα διαστήματος είναι να αναπαραστήσει την πραγματική θέση ενός ατόμου ως ένα διάστημα και όχι ως ένα συγκεκριμένο σημείο. Αυτό επιτρέπει την εισαγωγή ενός βαθμού αβεβαιότητας στα δεδομένα τοποθεσίας, τα οποία μπορούν να χρησιμοποιηθούν για τη διατήρηση του απορρήτου των ατόμων.

Η τεχνική του διαστήματος μανδύα λειτουργεί προσδιορίζοντας πρώτα το σύνολο των τοποθεσιών που είναι πιθανό να έχει επισκεφτεί ένα άτομο, με βάση το ιστορικό τοποθεσίας του. Στη συνέχεια, αυτές οι τοποθεσίες ομαδοποιούνται σε ένα σύνολο μη επικαλυπτόμενων διαστημάτων, με το μέγεθος των διαστημάτων να καθορίζεται από το απαιτούμενο επίπεδο απορρήτου. Στη συνέχεια, η πραγματική θέση του ατόμου αναπαρίσταται ως το κέντρο ενός από τα διαστήματα.

Το κύριο πλεονέκτημα του ενδιάμεσου μανδύα είναι ότι παρέχει έναν απλό και αποτελεσματικό τρόπο διατήρησης του απορρήτου των ατόμων στο LBS. Επιτρέπει επίσης τη διατήρηση της ιδιωτικής ζωής σε σύνολα δεδομένων με ποικίλες χωρικές πυκνότητες.

Ωστόσο, θα πρέπει να σημειωθεί ότι ο μανδύας διαστήματος δεν εγγυάται ότι η τοποθεσία των ατόμων δεν μπορεί να επαναπροσδιοριστεί και το επίπεδο απορρήτου εξαρτάται από το μέγεθος των διαστημάτων που επιλέγονται. Επιπλέον, η μέθοδος μπορεί να μην είναι κατάλληλη για όλους τους τύπους δεδομένων και θα μπορούσε να οδηγήσει σε απώλεια της χρησιμότητας δεδομένων εάν τα διαστήματα είναι πολύ μεγάλα.

Συνοπτικά, το Interval cloak είναι μια τεχνική διατήρησης απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία, καλύπτοντας τα δεδομένα τοποθεσίας τους χρησιμοποιώντας διαστήματα. Επιτρέπει τη διατήρηση του απορρήτου εισάγοντας αβεβαιότητα στα δεδομένα τοποθεσίας και είναι απλό και αποτελεσματικό, αλλά δεν εγγυάται πλήρη προστασία και η απόδοσή του εξαρτάται από το μέγεθος των επιλεγμένων διαστημάτων και θα μπορούσε να οδηγήσει σε απώλεια της χρησιμότητας δεδομένων εάν τα διαστήματα είναι πολύ μεγάλα.

3.2.3.6 Η τεχνική της απόκρυψης του Hilbert (Hilbert cloak)

Ο μανδύας Hilbert είναι μια τεχνική διατήρησης του απορρήτου που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία (LBS) καλύπτοντας τα δεδομένα τοποθεσίας τους χρησιμοποιώντας μια παραλλαγή της καμπύλης Hilbert. Η βασική ιδέα πίσω από τον μανδύα Hilbert είναι να χρησιμοποιηθούν οι ιδιότητες της καμπύλης Hilbert για να χαρτογραφηθούν οι γεωγραφικές τοποθεσίες των ατόμων σε έναν μονοδιάστατο χώρο,

όπου η πραγματική θέση ενός ατόμου μπορεί να αντικατασταθεί από ένα διάστημα χωρίς να χαθεί η χωρική σχέση μεταξύ των δεδομένων. σημεία.

Η τεχνική του μανδύα Hilbert λειτουργεί διαιρώντας πρώτα τη γεωγραφική περιοχή ενδιαφέροντος σε ένα πλέγμα κελιών χρησιμοποιώντας μια μέθοδο χωρικής κατανομής, όπως ένα τετράδεντρο. Στη συνέχεια, σε κάθε κελί εκχωρείται ένας μοναδικός κωδικός με βάση τη θέση του στην καμπύλη Hilbert. Στη συνέχεια, η πραγματική θέση ενός ατόμου αντικαθίσταται από τον κωδικό του κελιού μέσα στο οποίο εμπίπτει, καλύπτοντας αποτελεσματικά τη θέση του ατόμου.

Το κύριο πλεονέκτημα του μανδύα Hilbert είναι ότι παρέχει έναν τρόπο διατήρησης της ιδιωτικής ζωής των ατόμων στο LBS, διατηρώντας παράλληλα τη χωρική σχέση μεταξύ των σημείων δεδομένων. Επιτρέπει επίσης τη διατήρηση της ιδιωτικής ζωής σε σύνολα δεδομένων με ποικίλες χωρικές πυκνότητες.

Ωστόσο, θα πρέπει να σημειωθεί ότι όπως και άλλες τεχνικές, ο μανδύας Hilbert δεν εγγυάται ότι η τοποθεσία των ατόμων δεν μπορεί να επαναπροσδιοριστεί και το επίπεδο απορρήτου εξαρτάται από το μέγεθος των επιλεγμένων κελιών. Επιπλέον, η μέθοδος μπορεί να μην είναι κατάλληλη για όλους τους τύπους δεδομένων και θα μπορούσε να οδηγήσει σε απώλεια της χρησιμότητας δεδομένων εάν τα κελιά είναι πολύ μεγάλα.

Συνοπτικά, ο μανδύας Hilbert είναι μια τεχνική διατήρησης της ιδιωτικής ζωής που χρησιμοποιείται για την προστασία της ανωνυμίας των ατόμων σε υπηρεσίες που βασίζονται σε τοποθεσία, καλύπτοντας τα δεδομένα τοποθεσίας τους χρησιμοποιώντας μια παραλλαγή της καμπύλης Hilbert. Διατηρεί τη χωρική σχέση μεταξύ των σημείων δεδομένων και επιτρέπει τη διατήρηση του απορρήτου σε σύνολα δεδομένων με ποικίλες χωρικές πυκνότητες. Ωστόσο, δεν εγγυάται πλήρη προστασία και η απόδοσή του εξαρτάται από το μέγεθος των επιλεγμένων κελιών και μπορεί να οδηγήσει σε απώλεια της χρησιμότητας δεδομένων εάν τα κελιά είναι πολύ μεγάλα.

3.3 Ανωνυμία σε σημασιολογικά ευαίσθητες τοποθεσίες

Η ανωνυμία σε σημασιολογικά ευαίσθητους ιστότοπους αναφέρεται στην προστασία του απορρήτου των ατόμων όταν τα δεδομένα τοποθεσίας τους συλλέγονται σε περιοχές που θεωρούνται ευαίσθητες, όπως νοσοκομεία, θρησκευτικοί χώροι και πολιτικά κτίρια. Σε αυτές τις περιπτώσεις, τα δεδομένα τοποθεσίας που συλλέγονται θα μπορούσαν να αποκαλύψουν ευαίσθητες πληροφορίες σχετικά με την υγεία, τις θρησκευτικές πεποιθήσεις ή τις πολιτικές πεποιθήσεις ενός ατόμου, κάτι που θα μπορούσε να οδηγήσει σε διακρίσεις ή άλλες αρνητικές συνέπειες.

Υπάρχουν διάφορες τεχνικές που μπορούν να χρησιμοποιηθούν για την προστασία της ανωνυμίας σε σημασιολογικά ευαίσθητους ιστότοπους, όπως:

3.3.1 Καταστολή δεδομένων

Αυτή η τεχνική περιλαμβάνει την αφαίρεση δεδομένων τοποθεσίας που εμπίπτουν στα όρια σημασιολογικά ευαίσθητων τοποθεσιών από το σύνολο δεδομένων. Αυτό μπορεί να προστατεύσει

αποτελεσματικά την ανωνυμία των ατόμων, αλλά έχει επίσης ως αποτέλεσμα την απώλεια της χρησιμότητας των δεδομένων.

Η καταστολή δεδομένων είναι μια τεχνική που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων σε ένα σύνολο δεδομένων, αφαιρώντας ή αντικαθιστώντας ορισμένες ευαίσθητες πληροφορίες. Αυτή η τεχνική μπορεί να εφαρμοστεί σε διάφορους τύπους δεδομένων, συμπεριλαμβανομένων των χωρικών δεδομένων. Στο πλαίσιο των χωρικών δεδομένων, η καταστολή δεδομένων μπορεί να χρησιμοποιηθεί για την αφαίρεση ή την αντικατάσταση της ακριβούς τοποθεσίας ενός ατόμου σε ένα σύνολο δεδομένων, όπως η διεύθυνσή του ή οι συντεταγμένες GPS.

Μια κοινή μέθοδος καταστολής δεδομένων για χωρικά δεδομένα είναι η χρήση της γενίκευσης δεδομένων. Αυτή η τεχνική περιλαμβάνει τη μείωση του επιπέδου λεπτομέρειας στα χωρικά δεδομένα, όπως με τη συγκέντρωση δεδομένων σε μια μεγαλύτερη γεωγραφική περιοχή ή με την κατάργηση συγκεκριμένων λεπτομερειών για μια τοποθεσία. Για παράδειγμα, αντί να παρέχουν την ακριβή διεύθυνση ενός ατόμου, τα δεδομένα ενδέχεται να παρέχουν μόνο τη γειτονιά ή την πόλη στην οποία ζουν.

Μια άλλη μέθοδος καταστολής δεδομένων για χωρικά δεδομένα είναι η διαταραχή δεδομένων, η οποία περιλαμβάνει την προσθήκη μικρών τυχαίων σφαλμάτων στα δεδομένα για να είναι λιγότερο ακριβή. Αυτή η τεχνική μπορεί να χρησιμοποιηθεί για την προστασία του απορρήτου των ατόμων, ενώ εξακολουθεί να επιτρέπει στα δεδομένα να είναι χρήσιμα για ορισμένους τύπους ανάλυσης.

Η καταστολή δεδομένων μπορεί επίσης να συνδυαστεί με άλλες τεχνικές ανωνυμοποίησης, όπως η k-anonymity και η l-diversity, για να παρέχει πρόσθετη προστασία για ευαίσθητα δεδομένα. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η απόκρυψη δεδομένων μπορεί επίσης να μειώσει τη χρησιμότητα των δεδομένων και μπορεί να καταστήσει δύσκολη ή αδύνατη τη διεξαγωγή ορισμένων τύπων ανάλυσης. Ως εκ τούτου, είναι σημαντικό να λαμβάνετε υπόψη προσεκτικά τους συμβιβασμούς μεταξύ του απορρήτου και της χρησιμότητας δεδομένων όταν αποφασίζετε να χρησιμοποιήσετε την καταστολή δεδομένων.

3.3.2 Διαταραχή δεδομένων:

Αυτή η τεχνική περιλαμβάνει την προσθήκη τυχαίου θορύβου στα δεδομένα τοποθεσίας για να κρύψει την πραγματική θέση ενός ατόμου. Αυτό μπορεί να διατηρήσει την ανωνυμία των ατόμων, ενώ εξακολουθεί να επιτρέπει τη χρήση του συνόλου δεδομένων για ανάλυση.

Η διαταραχή δεδομένων είναι μια τεχνική που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων σε ένα σύνολο δεδομένων προσθέτοντας μικρά τυχαία σφάλματα στα δεδομένα. Αυτή η τεχνική μπορεί να εφαρμοστεί σε διάφορους τύπους δεδομένων, συμπεριλαμβανομένων των χωρικών δεδομένων. Στο πλαίσιο των χωρικών δεδομένων, η διαταραχή δεδομένων μπορεί να χρησιμοποιηθεί για την προσθήκη τυχαίων σφαλμάτων στα δεδομένα τοποθεσίας των ατόμων, όπως η διεύθυνσή τους ή οι συντεταγμένες GPS.

Μια κοινή μέθοδος διαταραχής δεδομένων για χωρικά δεδομένα είναι η χρήση διαφορικού απορρήτου. Αυτή η τεχνική περιλαμβάνει την προσθήκη τυχαίου θορύβου στα δεδομένα με ελεγχόμενο τρόπο ώστε να είναι λιγότερο ακριβή. Αυτός ο θόρυβος μπορεί να προστεθεί με διάφορους τρόπους, όπως με την προσθήκη τυχαίων σφαλμάτων στις συντεταγμένες μιας τοποθεσίας

ή με την προσθήκη τυχαίων σφαλμάτων στα χαρακτηριστικά μιας τοποθεσίας. Το διαφορικό απόρρητο μπορεί επίσης να χρησιμοποιηθεί για τον περιορισμό του όγκου των πληροφοριών που μπορούν να συναχθούν από τα δεδομένα.

Μια άλλη μέθοδος διατάραξης δεδομένων για χωρικά δεδομένα είναι η χρήση θορύβου που βασίζεται στη θέση. Αυτή η τεχνική περιλαμβάνει την προσθήκη τυχαίων σφαλμάτων στα δεδομένα τοποθεσίας των ατόμων σε ένα σύνολο δεδομένων. Τα σφάλματα μπορούν να προστεθούν με διάφορους τρόπους, όπως με την προσθήκη τυχαίων σφαλμάτων στις συντεταγμένες γεωγραφικού πλάτους και μήκους ή με την προσθήκη τυχαίων σφαλμάτων στη διεύθυνση ενός ατόμου.

Η διαταραχή δεδομένων μπορεί επίσης να συνδυαστεί με άλλες τεχνικές ανωνυμοποίησης, όπως η k-anonymity και η l-diversity, για να παρέχει πρόσθετη προστασία για ευαίσθητα δεδομένα. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η διαταραχή δεδομένων μπορεί επίσης να μειώσει τη χρησιμότητα των δεδομένων και μπορεί να καταστήσει δύσκολη ή αδύνατη τη διεξαγωγή ορισμένων τύπων ανάλυσης. Ως εκ τούτου, είναι σημαντικό να λαμβάνετε υπόψη προσεκτικά τους συμβιβασμούς μεταξύ του απορρήτου και της χρησιμότητας δεδομένων όταν αποφασίζετε να χρησιμοποιήσετε τη διαταραχή δεδομένων.

3.3.3 Γενίκευση δεδομένων

Αυτή η τεχνική περιλαμβάνει την ομαδοποίηση δεδομένων τοποθεσίας σε μεγαλύτερες περιοχές, όπως γειτονιές ή συνοικίες πόλεων, για να κρύψει την πραγματική τοποθεσία ενός ατόμου. Αυτό μπορεί να διατηρήσει την ανωνυμία των ατόμων, ενώ εξακολουθεί να επιτρέπει τη χρήση του συνόλου δεδομένων για ανάλυση.

Η γενίκευση δεδομένων είναι μια τεχνική που χρησιμοποιείται για την προστασία του απορρήτου των ατόμων σε ένα σύνολο δεδομένων αντικαθιστώντας τις λεπτομερείς πληροφορίες με πιο γενικές πληροφορίες. Αυτή η τεχνική μπορεί να εφαρμοστεί σε διάφορους τύπους δεδομένων, συμπεριλαμβανομένων των χωρικών δεδομένων. Στο πλαίσιο των χωρικών δεδομένων, η γενίκευση δεδομένων μπορεί να χρησιμοποιηθεί για να αντικαταστήσει συγκεκριμένες πληροφορίες τοποθεσίας, όπως μια διεύθυνση δρόμου, με πιο γενικές πληροφορίες, όπως μια γειτονιά ή μια πόλη.

Μια κοινή μέθοδος γενίκευσης δεδομένων για χωρικά δεδομένα είναι η χρήση χωρικής συνάθροισης. Αυτή η τεχνική περιλαμβάνει την ομαδοποίηση δεδομένων τοποθεσίας σε μεγαλύτερες γεωγραφικές περιοχές, όπως γειτονιές ή πόλεις. Στη συνέχεια, τα δεδομένα συνοψίζονται για κάθε γεωγραφική περιοχή και όχι για μεμονωμένες τοποθεσίες. Αυτό μπορεί να γίνει μετρώντας τον αριθμό των ατόμων σε κάθε περιοχή ή υπολογίζοντας στατιστικά στοιχεία όπως ο μέσος όρος ή ο διάμεσος για κάθε περιοχή.

Μια άλλη μέθοδος γενίκευσης δεδομένων για χωρικά δεδομένα είναι η χρήση χωρικής κάλυψης. Αυτή η τεχνική περιλαμβάνει την αφαίρεση ή την απόκρυψη ορισμένων τμημάτων των χωρικών δεδομένων, όπως τα τελευταία ψηφία μιας διεύθυνσης ή τα λιγότερο σημαντικά κομμάτια μιας συντεταγμένης γεωγραφικού πλάτους ή μήκους. Αυτό μπορεί να γίνει για την προστασία ευαίσθητων πληροφοριών, όπως η συγκεκριμένη τοποθεσία του σπιτιού ή του χώρου εργασίας ενός ατόμου.

Μια τρίτη μέθοδος γενίκευσης δεδομένων για χωρικά δεδομένα είναι η ιεραρχική γενίκευση, η οποία είναι η ομαδοποίηση δεδομένων σε επίπεδα ευαισθησίας, όπως χώρες, πολιτείες, πόλεις, γειτονιές και διευθύνσεις. Με αυτόν τον τρόπο, οι λεπτομερείς πληροφορίες μπορούν να αντικατασταθούν από λιγότερο λεπτομερείς πληροφορίες, όπως από διεύθυνση σε επίπεδο πόλης.

Η γενίκευση δεδομένων μπορεί επίσης να συνδυαστεί με άλλες τεχνικές ανωνυμοποίησης, όπως η k-anonymity και η l-diversity, για να παρέχει πρόσθετη προστασία για ευαίσθητα δεδομένα. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η γενίκευση δεδομένων μπορεί επίσης να μειώσει τη χρησιμότητα των δεδομένων και μπορεί να καταστήσει δύσκολη ή αδύνατη τη διεξαγωγή ορισμένων τύπων ανάλυσης. Ως εκ τούτου, είναι σημαντικό να λαμβάνετε υπόψη προσεκτικά τους συμβιβασμούς μεταξύ του απορρήτου και της χρησιμότητας δεδομένων όταν αποφασίζετε να χρησιμοποιήσετε τη γενίκευση δεδομένων.

3.3.4 Άλλοι Μέθοδοι ανωνυμοποίησης

Το k-anonymity, l-diversity, t-closeness κ.λπ. μπορούν να χρησιμοποιηθούν για την προστασία του απορρήτου των ατόμων σε σημασιολογικά ευαίσθητους ιστότοπους.

Είναι σημαντικό να σημειωθεί ότι ενώ αυτές οι τεχνικές μπορούν να προστατεύσουν αποτελεσματικά την ανωνυμία των ατόμων σε σημασιολογικά ευαίσθητους ιστότοπους, μπορεί επίσης να οδηγήσουν σε απώλεια της χρησιμότητας των δεδομένων και είναι απαραίτητο να επιτευχθεί μια ισορροπία μεταξύ της προστασίας του απορρήτου και της διατήρησης της χρησιμότητας δεδομένων. Επιπλέον, το επίπεδο προστασίας μπορεί να διαφέρει ανάλογα με την τεχνική που χρησιμοποιείται και το συγκεκριμένο πλαίσιο των δεδομένων τοποθεσίας.

3.4 Τροχιές δεδομένων

Τα ίχνη δεδομένων αναφέρονται στο ψηφιακό ίχνος που αφήνουν πίσω τους τα άτομα ενώ χρησιμοποιούν διάφορες ψηφιακές συσκευές και υπηρεσίες, όπως κινητά τηλέφωνα, φορητούς υπολογιστές και το διαδίκτυο. Αυτά τα ίχνη δεδομένων συχνά περιλαμβάνουν ευαίσθητες πληροφορίες, όπως προσωπικά δεδομένα και δεδομένα τοποθεσίας, τα οποία μπορούν να χρησιμοποιηθούν για την αναγνώριση και την παρακολούθηση ατόμων. Αυτό μπορεί να αποτελέσει σημαντικό κίνδυνο για το απόρρητο και την ανωνυμία.

Για να διατηρηθεί η ανωνυμία, τα ίχνη δεδομένων πρέπει να διαχειρίζονται, να καλύπτονται ή να εξαλειφθούν. Υπάρχουν διάφορες τεχνικές που χρησιμοποιούνται για να επιτευχθεί αυτό, όπως η καταστολή δεδομένων, η διαταραχή δεδομένων και η γενίκευση δεδομένων. Η καταστολή δεδομένων περιλαμβάνει την πλήρη αφαίρεση δεδομένων από την εγγραφή, ενώ η διαταραχή δεδομένων περιλαμβάνει την αλλαγή των δεδομένων με τρόπο που δεν επηρεάζει τη χρησιμότητά τους, αλλά καθιστά δύσκολη την αναγνώριση ατόμων. Η γενίκευση δεδομένων, από την άλλη πλευρά, περιλαμβάνει την αντικατάσταση των συγκεκριμένων τιμών των δεδομένων με πιο γενικές τιμές, όπως η αντικατάσταση συγκεκριμένων τοποθεσιών με μεγαλύτερες περιοχές ή η συγκέντρωση δεδομένων σε μεγαλύτερες κατηγορίες.

Οι τεχνικές απόκρυψης κάλυψης μπορούν επίσης να χρησιμοποιηθούν για τη διατήρηση της ανωνυμίας στα ίχνη δεδομένων. Αυτές οι τεχνικές μπορούν να περιλαμβάνουν NN-cloak, Casper, interval cloak, Hilbert cloak, center cloak, clique cloak και space-depending cloak. Κάθε μία από αυτές τις τεχνικές έχει τα δικά της δυνατά και αδύνατα σημεία και είναι κατάλληλη για διαφορετικούς τύπους δεδομένων και σεναρίων.

Είναι σημαντικό να σημειωθεί ότι ενώ τεχνικές μπορούν να χρησιμοποιηθούν για την ενίσχυση της ιδιωτικότητας και της ανωνυμίας, δεν είναι αλάνθαστες και δεν μπορούν να εγγυηθούν πλήρη προστασία. Η χρήση αυτών των τεχνικών εγείρει επίσης ερωτήματα σχετικά με τον συμβιβασμό μεταξύ ιδιωτικότητας και ασφάλειας, καθώς και την πιθανότητα κατάχρησης.

Συμπερασματικά, τα ίχνη δεδομένων είναι μια σημαντική πτυχή της ιδιωτικής ζωής και της ανωνυμίας στην ψηφιακή εποχή. Η κατανόηση των διαφόρων τεχνικών που χρησιμοποιούνται για τη διαχείριση και τη διατήρηση της ανωνυμίας στα ίχνη δεδομένων είναι ζωτικής σημασίας για τη διασφάλιση του απορρήτου στον όλο και πιο συνδεδεμένο κόσμο μας.

3.5 CryptDB

Το CryptDB είναι ένα σύστημα βάσης δεδομένων που παρέχει ασφαλή, αποτελεσματική και επεκτάσιμη επεξεργασία ευαίσθητων δεδομένων. Αναπτύχθηκε από ερευνητές στο MIT το 2011 και έκτοτε έχει γίνει μια δημοφιλής λύση για την προστασία ευαίσθητων δεδομένων στο cloud και σε άλλα κοινόχρηστα περιβάλλοντα.

Η βασική ιδέα πίσω από το CryptDB είναι η κρυπτογράφηση ευαίσθητων δεδομένων σε μεμονωμένο επίπεδο κυψέλης και στη συνέχεια η εκτέλεση λειτουργιών βάσης δεδομένων στα κρυπτογραφημένα δεδομένα χρησιμοποιώντας SQL. Το σύστημα χρησιμοποιεί έναν συνδυασμό τεχνικών κρυπτογράφησης και τεχνικών βελτιστοποίησης βάσης δεδομένων για να διασφαλίσει ότι τα κρυπτογραφημένα δεδομένα μπορούν να υποβληθούν σε αποτελεσματική επεξεργασία, διατηρώντας παράλληλα την εμπιστευτικότητά τους.

Ένα από τα κύρια πλεονεκτήματα του CryptDB είναι η ικανότητά του να παρέχει ισχυρή προστασία για ευαίσθητα δεδομένα, ακόμη και όταν τα δεδομένα αποθηκεύονται σε κοινόχρηστο περιβάλλον. Αυτό το καθιστά ιδανική λύση για εφαρμογές που απαιτούν απόρρητο δεδομένων, όπως εφαρμογές υγειονομικής περίθαλψης ή οικονομικές εφαρμογές.

Το CryptDB λειτουργεί κρυπτογραφώντας τα δεδομένα σε μεμονωμένο επίπεδο κυψέλης και στη συνέχεια εκτελώντας λειτουργίες SQL στα κρυπτογραφημένα δεδομένα. Αυτό επιτρέπει στο σύστημα να εκτελεί λειτουργίες βάσης δεδομένων σε ευαίσθητα δεδομένα χωρίς να διακυβεύεται η εμπιστευτικότητά τους. Το σύστημα χρησιμοποιεί έναν συνδυασμό τεχνικών, συμπεριλαμβανομένης της ομομορφικής κρυπτογράφησης, για να διασφαλίσει ότι τα κρυπτογραφημένα δεδομένα μπορούν να υποβληθούν σε αποτελεσματική επεξεργασία.

Ένα άλλο βασικό χαρακτηριστικό του CryptDB είναι η ικανότητά του να παρέχει λεπτομερή έλεγχο της ασφάλειας των δεδομένων. Αυτό επιτρέπει στους διαχειριστές να καθορίσουν ποια

δεδομένα είναι κρυπτογραφημένα και ποια όχι, με βάση τις συγκεκριμένες ανάγκες της εφαρμογής τους.

Συνολικά, το CryptDB είναι ένα ισχυρό εργαλείο για την προστασία ευαίσθητων δεδομένων σε κοινόχρηστα περιβάλλοντα και είναι ιδανικό για εφαρμογές που απαιτούν απόρρητο δεδομένων, όπως εφαρμογές υγειονομικής περίθαλψης ή οικονομικές εφαρμογές. Ο συνδυασμός κρυπτογράφησης, βελτιστοποίησης βάσης δεδομένων και λεπτομερούς ελέγχου της ασφάλειας δεδομένων το καθιστούν μια πολύτιμη λύση για οργανισμούς που πρέπει να προστατεύουν ευαίσθητα δεδομένα.

Ένα από τα βασικά του ελαττώματα είναι ότι απαιτεί σημαντικές τροποποιήσεις στο σύστημα διαχείρισης της βάσης δεδομένων και στον κώδικα της εφαρμογής προκειμένου να ενσωματωθούν τα χαρακτηριστικά ασφαλείας του. Επιπλέον, το CryptDB μπορεί να εισάγει επιβάρυνση απόδοσης και καθυστέρηση λόγω της κρυπτογράφησης και της αποκρυπτογράφησης των δεδομένων. Βασίζεται επίσης σε αξιόπιστους διακομιστές, οι οποίοι μπορεί να είναι ευάλωτοι σε επιθέσεις και ενδέχεται να μην παρέχουν ισχυρή προστασία έναντι όλων των τύπων επιθέσεων, όπως αυτές που στοχεύουν στην κλοπή κλειδιών κρυπτογράφησης ή στον κίνδυνο του ίδιου του αλγόριθμου κρυπτογράφησης. Τέλος, οι εγγυήσεις ασφαλείας της CryptDB βασίζονται στην υπόθεση ότι οι κρυπτογραφικοί αλγόριθμοι που χρησιμοποιούνται είναι ασφαλείς και άθραυστοι, κάτι που μπορεί να μην συμβαίνει πάντα στην πράξη.

Το ARX μπορεί να παρέχει καλύτερη προστασία απορρήτου από το CryptDB με διάφορους τρόπους:

3.5.1 Ευελιξία

Το ARX παρέχει μια ευέλικτη και φιλική προς τον χρήστη διεπαφή που επιτρέπει στους χρήστες να ορίζουν τις απαιτήσεις απορρήτου τους και να ρυθμίζουν με ακρίβεια τη διαδικασία ανωνυμοποίησης. Το CryptDB, από την άλλη πλευρά, απαιτεί τροποποίηση του συστήματος διαχείρισης της βάσης δεδομένων και του κώδικα εφαρμογής.

3.5.2 Ακρίβεια

Το ARX παρέχει λεπτομερή έλεγχο της διαδικασίας ανωνυμοποίησης, επιτρέποντας τη διατήρηση της χρησιμότητας και της ποιότητας των δεδομένων, ενώ το CryptDB δεν παρέχει σαφή μηχανισμό για τη διατήρηση της ποιότητας των δεδομένων.

3.5.3 Απόδοση

Το ARX μπορεί να χειριστεί μεγάλα σύνολα δεδομένων αποτελεσματικά, ενώ το CryptDB μπορεί να εισάγει επιβάρυνση απόδοσης και καθυστέρηση λόγω της κρυπτογράφησης και της αποκρυπτογράφησης των δεδομένων.

3.5.4 Προστασία δεδομένων

Το ARX παρέχει υποστήριξη για μοντέλα απορρήτου, όπως το k-anonymity και το l-diversity, τα οποία προστατεύουν από διαφορετικούς τύπους επιθέσεων απορρήτου. Το CryptDB παρέχει μόνο προστασία απορρήτου βάσει κρυπτογράφησης.

Συνοπτικά, το ARX παρέχει πιο ευέλικτη, ακριβή, αποτελεσματική και αποτελεσματική προστασία απορρήτου σε σύγκριση με το CryptDB.

3.6 Σχετικές Εργασίες

3.6.1 Clique Cloak

An Efficient K-Anonymization Algorithm for Large Spatial Databases. Αυτή η εργασία παρουσιάζει έναν αποτελεσματικό και επεκτάσιμο αλγόριθμο για την ανωνυμοποίηση μεγάλων χωρικών βάσεων δεδομένων. Οι συγγραφείς προτείνουν μια νέα προσέγγιση που ονομάζεται clique cloak, η οποία χωρίζει τα δεδομένα σε επικαλυπτόμενες συστάδες και αντικαθιστά τις αρχικές τιμές με τα κεντροειδή αυτών των συστάδων. Ο αλγόριθμος αξιολογείται χρησιμοποιώντας ένα πραγματικό σύνολο δεδομένων και αποδεικνύεται αποτελεσματικός και αποδοτικός.

Η εργασία [3] παρουσιάζει ένα πλαίσιο για την προστασία του απορρήτου σε μεγάλες χωρικές βάσεις δεδομένων. Η κύρια συμβολή της εργασίας είναι η εισαγωγή του αλγόριθμου Clique Cloak, ο οποίος είναι μια μέθοδος ανωνυμοποίησης που προστατεύει το απόρρητο μετατρέποντας τα δεδομένα με τρόπο που διατηρεί τη γενική δομή των δεδομένων, ενώ συγκαλύπτει τις ευαίσθητες πληροφορίες. Ο αλγόριθμος Clique Cloak λειτουργεί διαιρώντας τα δεδομένα σε μικρές ομάδες που ονομάζονται κλίκες και στη συνέχεια γενικεύοντας τα δεδομένα σε κάθε κλίκα σε ένα κοινό επίπεδο λεπτομέρειας. Το έγγραφο εξετάζει επίσης τις ανταλλαγές μεταξύ του απορρήτου και της χρησιμότητας δεδομένων και δείχνει πώς ο αλγόριθμος Clique Cloak παρέχει μια καλή ισορροπία μεταξύ των δύο. Συνολικά, το έγγραφο παρέχει μια χρήσιμη συμβολή στον τομέα της προστασίας της ιδιωτικής ζωής για χωρικές βάσεις δεδομένων.

3.6.2 A General Framework for Privacy-Preserving Data Publishing

Αυτό το έγγραφο παρέχει μια ολοκληρωμένη επισκόπηση των τεχνικών δημοσίευσης δεδομένων που διατηρούν το απόρρητο, συμπεριλαμβανομένης της καταστολής δεδομένων, της διαταραχής δεδομένων και της γενίκευσης δεδομένων. Οι συγγραφείς περιγράφουν αρκετούς αλγόριθμους για την εκτέλεση καθεμιάς από αυτές τις τεχνικές και αξιολογούν την απόδοση κάθε αλγορίθμου χρησιμοποιώντας μια ποικιλία μετρήσεων. Το έγγραφο εξετάζει επίσης τους συμβιβασμούς μεταξύ της ιδιωτικής ζωής και της χρησιμότητας δεδομένων και παρέχει καθοδήγηση για την επιλογή της καταλληλότερης τεχνικής για ένα συγκεκριμένο σενάριο δημοσίευσης δεδομένων.

Η εργασία [4] που παρουσιάζει ένα γενικό πλαίσιο για τη διατήρηση της ιδιωτικής ζωής στη δημοσίευση δεδομένων. Οι συγγραφείς υποστηρίζουν ότι οι παραδοσιακές μέθοδοι για την προστασία του απορρήτου, όπως η καταστολή δεδομένων και η γενίκευση, είναι συχνά υπερβολικά περιοριστικές και δεν αντιμετωπίζουν πλήρως τις ανησυχίες των ατόμων σχετικά με το απόρρητο. Αντίθετα, οι συγγραφείς προτείνουν ένα νέο πλαίσιο που ενσωματώνει πολλαπλές τεχνικές διατήρησης της ιδιωτικής ζωής, συμπεριλαμβανομένης της γενίκευσης, της καταστολής και της διαταραχής, για να παρέχει πιο ευέλικτη και αποτελεσματική προστασία της ιδιωτικής ζωής. Αυτό το πλαίσιο είναι επίσης σε θέση να χειρίζεται μεγάλα και πολύπλοκα σύνολα δεδομένων, καθιστώντας το κατάλληλο για τον αυξανόμενο όγκο ευαίσθητων δεδομένων που παράγονται και μοιράζονται στη σημερινή κοινωνία. Οι συγγραφείς αποδεικνύουν την αποτελεσματικότητα του πλαισίου τους μέσω μιας σειράς πειραμάτων και δείχνουν ότι είναι σε θέση να παρέχει ισχυρή προστασία της ιδιωτικής ζωής διατηρώντας παράλληλα τη χρησιμότητα των δεδομένων για αναλυτικούς σκοπούς.

3.6.3 NN-Cloak

Ένας αποτελεσματικός και αποδοτικός αλγόριθμος K-Anonymization για δεδομένα τροχιάς. Αυτό το έγγραφο προτείνει έναν νέο αλγόριθμο k-ανωνυμοποίησης για δεδομένα τροχιάς, ο οποίος είναι ιδιαίτερα ευαίσθητος και απαιτεί προσεκτική προστασία. Ο αλγόριθμος, που ονομάζεται NN-Cloak, χρησιμοποιεί έναν συνδυασμό ομαδοποίησης και χωρικής γενίκευσης για την επίτευξη k-ανωνυμίας. Οι συγγραφείς αξιολογούν τον αλγόριθμο χρησιμοποιώντας δεδομένα τροχιάς του πραγματικού κόσμου και επιδεικνύουν την αποτελεσματικότητα και την αποδοτικότητά του.

Ο αλγόριθμος στην εργασία [1] βασίζεται στην τεχνική ομαδοποίησης k-means και επιδιώκει να αντιμετωπίσει την πρόκληση της διατήρησης της ιδιωτικής ζωής κατά τη δημοσίευση της τροχιάς Οι συγγραφείς αξιολόγησαν την αποτελεσματικότητα και την αποδοτικότητα του NN-Cloak σε δεδομένα τροχιάς του πραγματικού κόσμου και συνέκριναν τα αποτελέσματά του με αυτά άλλων υπαρχόντων αλγορίθμων. Οι συγγραφείς συζήτησαν επίσης τους περιορισμούς του αλγορίθμου και τις μελλοντικές κατευθύνσεις για τη βελτίωσή του.

3.6.4 Anonymizing Spatial Trajectory Data

Ανασκόπηση της τελευταίας τεχνολογίας. Αυτή η εργασία παρέχει μια ολοκληρωμένη ανασκόπηση της υπάρχουσας βιβλιογραφίας σχετικά με την ανωνυμοποίηση δεδομένων χωρικής τροχιάς. Οι συγγραφείς συζητούν διάφορες τεχνικές ανωνυμοποίησης, όπως η καταστολή δεδομένων, η διαταραχή δεδομένων, η γενίκευση δεδομένων και οι μέθοδοι απόκρυψης στοιχείων. Το έγγραφο παρέχει επίσης μια ολοκληρωμένη σύγκριση των διαφόρων τεχνικών, επισημαίνοντας τα δυνατά και αδύνατα σημεία τους και παρέχει καθοδήγηση για την επιλογή της καταλληλότερης τεχνικής για ένα συγκεκριμένο σενάριο δημοσίευσης δεδομένων.

Το άρθρο [2] προτείνει ένα πλαίσιο για την ανωνυμοποίηση δεδομένων χωρικής τροχιάς χρησιμοποιώντας την k-ανωνυμία και την l-diversity ως κριτήρια προστασίας της ιδιωτικής ζωής. Το πλαίσιο περιλαμβάνει έναν αλγόριθμο ανωνυμοποίησης που χρησιμοποιεί τεχνικές χωρικής και

χρονικής ομαδοποίησης για την επίτευξη k-ανωνυμίας και l-ποικιλομορφίας στα δεδομένα. Ο αλγόριθμος χρησιμοποιεί επίσης μεθόδους καταστολής και γενίκευσης δεδομένων για να διασφαλίσει ότι πληρούνται τα κριτήρια απορρήτου. Η αποτελεσματικότητα του πλαισίου αξιολογήθηκε χρησιμοποιώντας σύνολα δεδομένων πραγματικού κόσμου και τα αποτελέσματα δείχνουν ότι η προτεινόμενη μέθοδος παρέχει υψηλό επίπεδο προστασίας της ιδιωτικής ζωής διατηρώντας παράλληλα τη χρησιμότητα των δεδομένων.

3.6.5 Cloaking Location-Based Services: A Survey of Techniques and Attacks

Ένα άλλο δημοφιλές έγγραφο στον τομέα του απορρήτου χωρικών δεδομένων από τους Fan Wu, Kai Chen και Cheng-Zhong Xu. Αυτό το έγγραφο παρέχει μια περιεκτική επισκόπηση των τεχνικών που χρησιμοποιούνται στην απόκρυψη υπηρεσιών που βασίζονται σε τοποθεσία και των επιθέσεων που έχουν προταθεί για να σπάσουν αυτές τις τεχνικές. Οι συγγραφείς αξιολογούν τις διαφορετικές τεχνικές απόκρυψης με βάση τα δυνατά και αδύνατα σημεία τους και υπογραμμίζουν τις προκλήσεις που αντιμετωπίζει η ανάπτυξη αποτελεσματικών λύσεων διατήρησης της ιδιωτικής ζωής για υπηρεσίες που βασίζονται σε τοποθεσία.

Είναι το άρθρο [4] που παρέχει μια ολοκληρωμένη επισκόπηση των διαφορετικών τεχνικών και μεθόδων που χρησιμοποιούνται για την προστασία του απορρήτου στις υπηρεσίες που βασίζονται σε τοποθεσία (LBS). Οι συγγραφείς παρουσιάζουν μια ταξινόμηση των μεθόδων απόκρυψης σε τεχνικές που βασίζονται σε δεδομένα και σε τεχνικές που βασίζονται στο διάστημα και συζητούν τα δυνατά και τα αδύνατα σημεία κάθε προσέγγισης. Εξετάζουν επίσης διάφορες επιθέσεις στο απόρρητο στο LBS και αξιολογούν την ανθεκτικότητα των διαφορετικών τεχνικών απόκρυψης έναντι αυτών των επιθέσεων. Η εργασία ολοκληρώνεται επισημαίνοντας τις προκλήσεις και τα ανοιχτά προβλήματα στη δημοσίευση δεδομένων για τη διατήρηση της ιδιωτικής ζωής στο LBS και παρέχοντας πληροφορίες για μελλοντική έρευνα σε αυτόν τον τομέα.

4

Πειραματική Υλοποίηση

Στο συγκεκριμένο κεφάλαιο θα γίνει συγκριτική μελέτη ανωνυμοποίησης μιας βάσης δεδομένων, χρησιμοποιώντας το πρόγραμμα ανοιχτού κώδικα "ARX".

Λίγα λόγια για το "ARX" :

Το ARX είναι ένα ολοκληρωμένο λογισμικό ανοιχτού κώδικα για την ανωνυμοποίηση ευαίσθητων προσωπικών δεδομένων. Υποστηρίζει μια μεγάλη ποικιλία μοντέλων απορρήτου και κινδύνου, μεθόδων μετατροπής δεδομένων και μεθόδων για την ανάλυση της χρησιμότητας των δεδομένων εξόδου.

Το λογισμικό έχει χρησιμοποιηθεί σε ποικίλα πλαίσια, συμπεριλαμβανομένων εμπορικών πλατφορμών ανάλυσης μεγάλων δεδομένων, ερευνητικών έργων, κοινής χρήσης δεδομένων κλινικών δοκιμών και για εκπαιδευτικούς σκοπούς.

Το ARX είναι σε θέση να χειρίζεται μεγάλα σύνολα δεδομένων σε υλικό εμπορευμάτων και διαθέτει ένα διαισθητικό γραφικό περιβάλλον εργασίας χρήστη μεταξύ πλατφορμών.

Η βάση δεδομένων μας αποτελείται από 21 στήλες και έχει 9551 εγγραφές. Περιέχει ευαίσθητα και «υπερευαίσθητα» δεδομένα αλλά και μεγάλο όγκο από δεδομένα τα οποία θεωρούνται «κοινά»

Αρχικοποιημένο στάδιο βάσης δεδομένων :

Η βάσης μας, στο αρχικό της στάδιο και χωρίς καμία επεξεργασία, όπως είναι προφανές, δεν παρέχει καμία ασφάλεια στα δεδομένα της. Ο καθένας μπορεί να αναγνώσει και να επεξεργαστεί τις εγγραφές της.

Στην συνέχεια θα αναλύσουμε τρεις παραλλαγές της βάσης μας, όπου με διάφορες παραμέτρους θα προχωρήσουμε στην ανωνυμοποίηση.

1η παραλλαγή :

Στην πρώτη παραλλαγή (Πίνακας 4.1) παρουσιάζεται η αρχικοποιημένη κατάσταση της βάσης και έχουμε χρησιμοποιήσει αρκετά «αυστηρές» παραμέτρους.

k-Anonymity	5
l-diversity Aggregate rating	5
l-diversity City	5
l-diversity Country Code	2
l-diversity Cuisines	5
l-diversity Currency	5
l-diversity Locality	5
l-diversity Locality Verbose	5
Suppression Limit	100%
Precomputation	0%
Measure	Loss
Aggregate function	Rank
Suppression/Generalization	50/50
Weights	
Restaurant	1
Address	0,5
Longitude	0,5
Latitude	0,5
Has Table	0,5
Has Online	0,5
Is delivering	0,5
Price range	0,5
Rating colour	0,5
Rating text	0,5
Votes	0,5

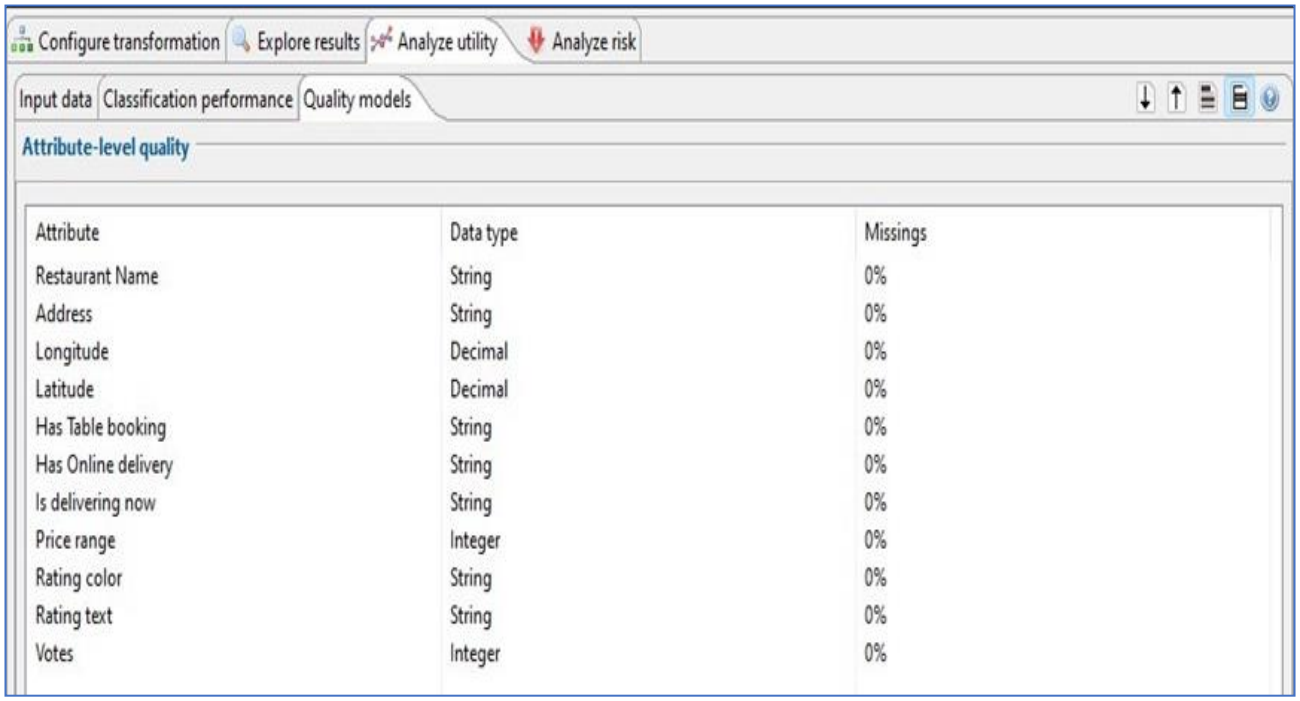
Πίνακας 4.1

Παράμετροι ανωνυμοποίησης του "ARX" στην πρώτη παραλλαγή

Τα αποτελέσματα που μας έδωσε το ARX είναι τα εξής :

Αυτή η ενότητα εμφανίζει μέτρα ποιότητας δεδομένων που λαμβάνονται για δεδομένα εξόδου χρησιμοποιώντας διάφορα μοντέλα γενικής χρήσης. Επιπλέον, παρουσιάζονται οι τύποι δεδομένων των χαρακτηριστικών καθώς και το κλάσμα των τιμών που λείπουν.

Σημείωση: Όλα τα μέτρα που παρουσιάζονται σε αυτήν την προβολή έχουν κλιμακωθεί και κανονικοποιηθεί στο εύρος [0, 1] όπου το 0 αντιπροσωπεύει ένα σύνολο δεδομένων από το οποίο έχουν αφαιρεθεί όλες οι πληροφορίες και το 1 αντιπροσωπεύει το μη τροποποιημένο σύνολο δεδομένων εισόδου. Κατά συνέπεια, οι υψηλότερες τιμές που επιστρέφονται για τα ποιοτικά μοντέλα είναι πάντα καλύτερες.

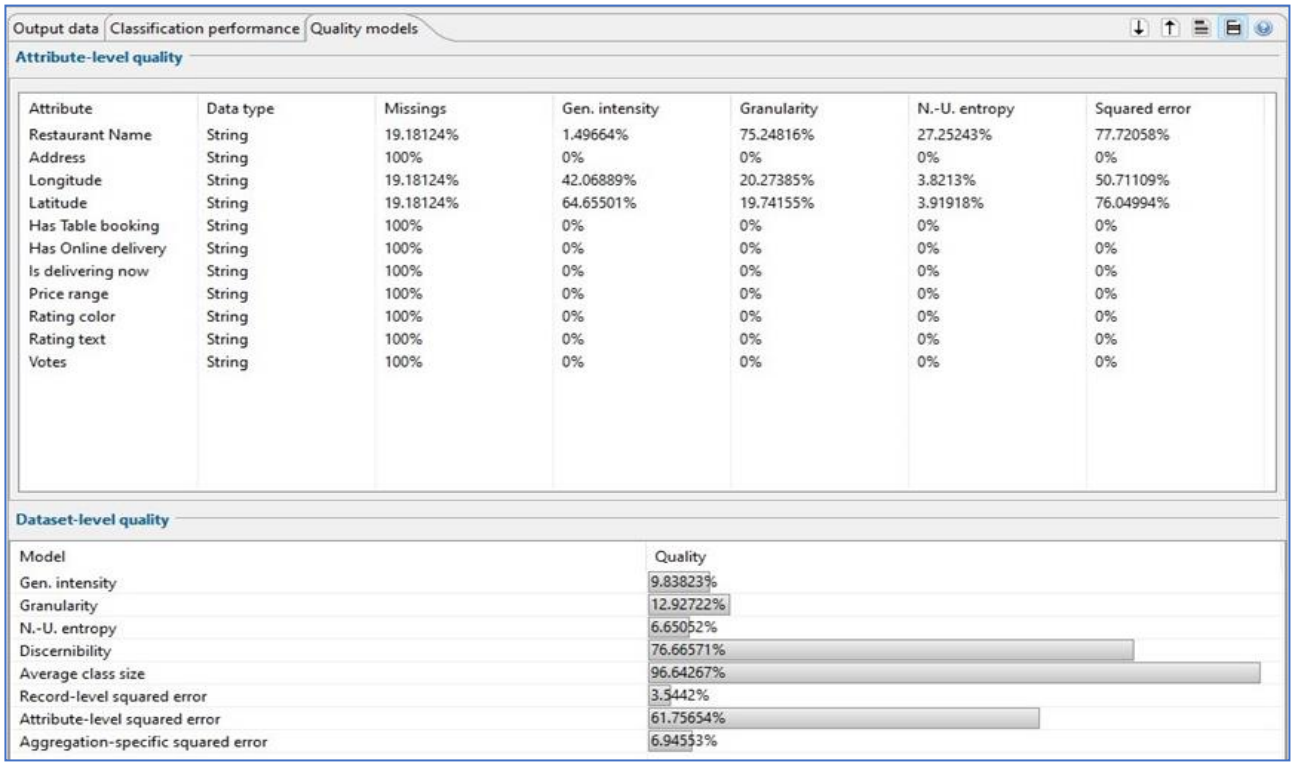


Attribute	Data type	Missings
Restaurant Name	String	0%
Address	String	0%
Longitude	Decimal	0%
Latitude	Decimal	0%
Has Table booking	String	0%
Has Online delivery	String	0%
Is delivering now	String	0%
Price range	Integer	0%
Rating color	String	0%
Rating text	String	0%
Votes	Integer	0%

Σχήμα 4.1

Αρχική κατάσταση του περιεχόμενου της βάσης

Η προβολή στο σχήμα 4.1 δείχνει την ποιότητα σε επίπεδο χαρακτηριστικών, δηλαδή εκτιμήσεις ποιότητας που σχετίζονται με μεμονωμένα οιονεί αναγνωριστικά. Η προβολή στο σχήμα 4.2 δείχνει την ποιότητα σε επίπεδο δεδομένων, δηλαδή εκτιμήσεις ποιότητας για το συνολικό σύνολο οιονεί αναγνωριστικών.



Output data | Classification performance | Quality models

Attribute-level quality

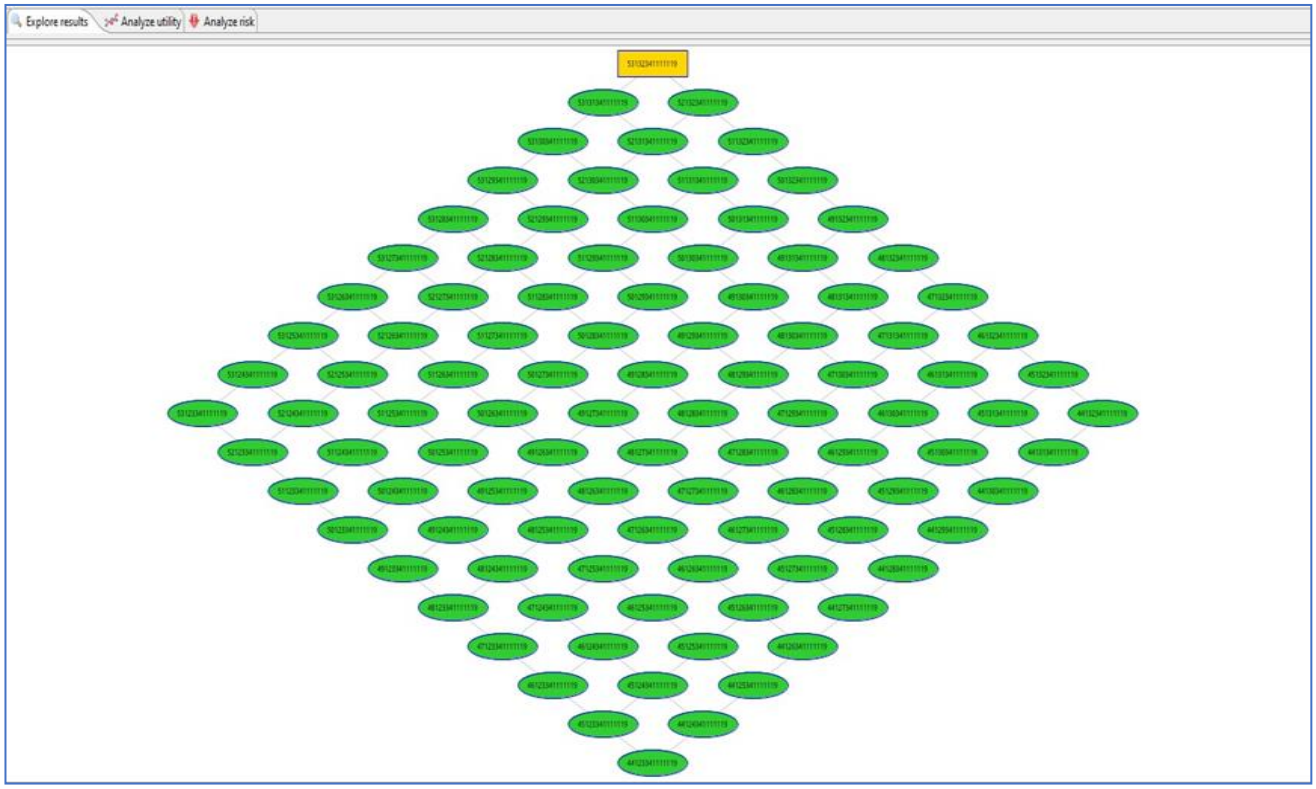
Attribute	Data type	Missings	Gen. intensity	Granularity	N.-U. entropy	Squared error
Restaurant Name	String	19.18124%	1.49664%	75.24816%	27.25243%	77.72058%
Address	String	100%	0%	0%	0%	0%
Longitude	String	19.18124%	42.06889%	20.27385%	3.8213%	50.71109%
Latitude	String	19.18124%	64.65501%	19.74155%	3.91918%	76.04994%
Has Table booking	String	100%	0%	0%	0%	0%
Has Online delivery	String	100%	0%	0%	0%	0%
Is delivering now	String	100%	0%	0%	0%	0%
Price range	String	100%	0%	0%	0%	0%
Rating color	String	100%	0%	0%	0%	0%
Rating text	String	100%	0%	0%	0%	0%
Votes	String	100%	0%	0%	0%	0%

Dataset-level quality

Model	Quality
Gen. intensity	9.838823%
Granularity	12.92722%
N.-U. entropy	6.65052%
Discernibility	76.66571%
Average class size	96.64267%
Record-level squared error	3.5442%
Attribute-level squared error	61.75654%
Aggregation-specific squared error	6.94553%

Σχήμα 4.2

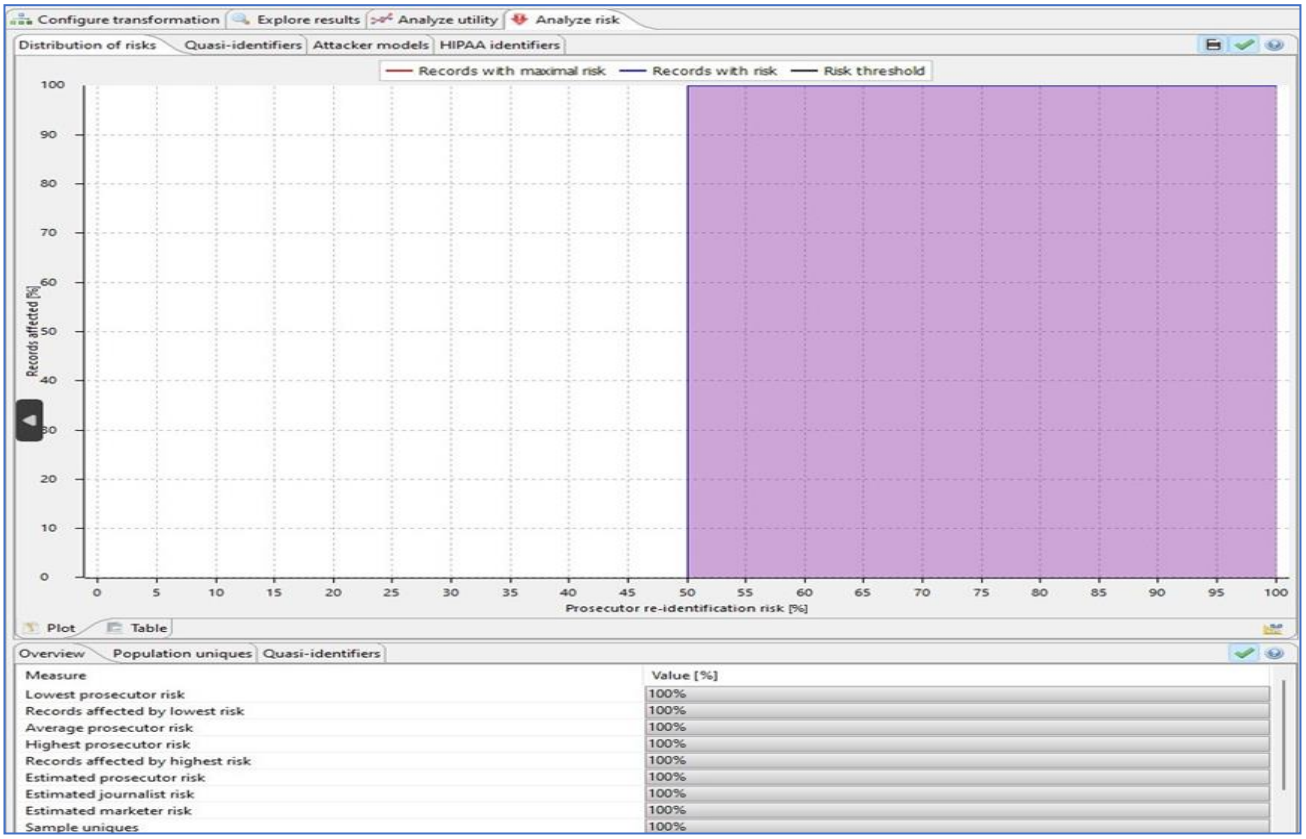
Κατάσταση του περιεχόμενου της βάσης με τις παραμέτρους εφαρμοσμένες



Σχήμα 4.3

Δέντρο ανωνυμοποίησης

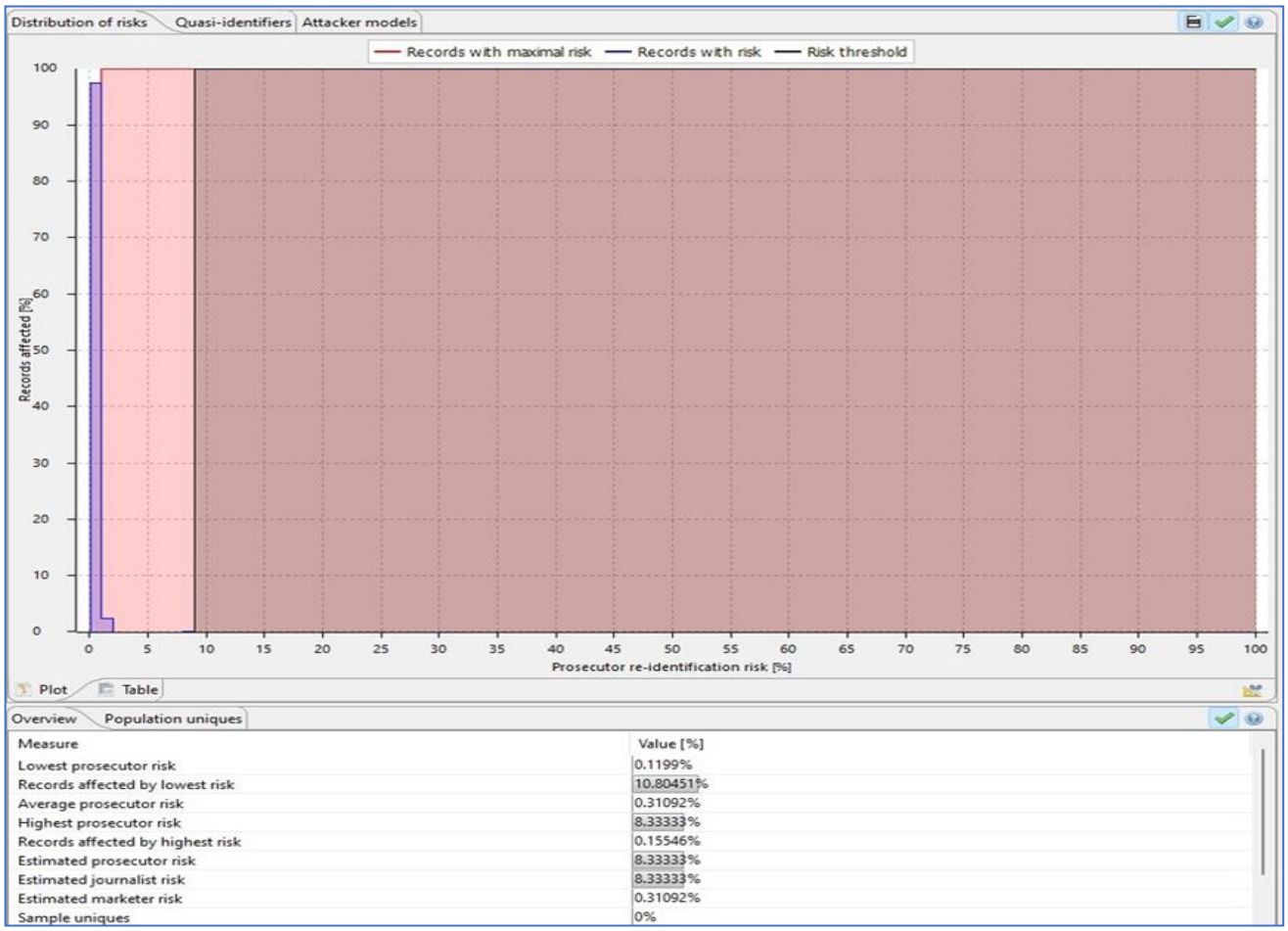
Στο σχήμα 4.3, ο χώρος της λύσης αναλύεται και οι μετασχηματισμοί μπορούν να εφαρμοστούν στο σύνολο δεδομένων εισόδου και εμφανίζεται ένα υποσύνολο του τρέχοντος χώρου λύσης. Κάθε μετασχηματισμός προσδιορίζεται από τα επίπεδα γενίκευσης που καθορίζει για τα οιονεί αναγνωριστικά στο σύνολο δεδομένων εισόδου.



Σχήμα 4.4

Εκτίμηση ρίσκου των δεδομένων της βάσης σε αρχικοποιημένη κατάσταση

Στο σχήμα 4.4, εμφανίζεται η κατανομή των κινδύνων επαναπροσδιορισμού μεταξύ των εγγραφών του συνόλου δεδομένων, όπου στην προκειμένη περίπτωση, τα δεδομένα είναι τελείως εκτεθειμένα. Η κατανομή υπολογίζεται και για τα δεδομένα εισόδου και εξόδου, είτε ως ιστόγραμμα είτε ως πίνακας.



Σχήμα 4.5

Εκτίμηση ρίσκου των δεδομένων της βάσης σε ανωνυμοποιημένη κατάσταση

Στο σχήμα 4.5, εμφανίζεται η κατανομή των κινδύνων επαναπροσδιορισμού μεταξύ των εγγραφών του συνόλου δεδομένων, όπου στην προκειμένη περίπτωση, τα δεδομένα είναι κυρίως προστατευμένα και μόνο ένα πολύ μικρό ποσοστό είναι σε κίνδυνο. Η κατανομή υπολογίζεται και για τα δεδομένα εισόδου και εξόδου, είτε ως ιστόγραμμα είτε ως πίνακας.

Σχήμα 4.6

Εκτίμηση ρίσκου των δεδομένων της βάσης σε αρχικοποιημένη κατάσταση

Στο σχήμα 4.6 εμφανίζει μια επισκόπηση πολλών μέτρων για τον επαναπροσδιορισμό των κινδύνων. Στην ανώτερη περιοχή αυτής της προοπτικής, παρέχονται εκτιμήσεις κινδύνου για τρία διαφορετικά μοντέλα επιτιθέμενων: (1) το σενάριο του εισαγγελέα, (2) το σενάριο του δημοσιογράφου και (3) το σενάριο του μάρκετινγκ. Στην συγκεκριμένη περίπτωση οι δείκτες μας δείχνουν πως ο κίνδυνος είναι στο ύψιστο επίπεδο.

Μπορούν να παρέχονται κατώφλια για τον υψηλότερο κίνδυνο από οποιαδήποτε εγγραφή, για τις εγγραφές που έχουν υψηλότερο κίνδυνο από αυτό το όριο και για το μέσο κλάσμα των εγγραφών που μπορούν να επαναπροσδιοριστούν με επιτυχία. Περισσότερες λεπτομέρειες σχετικά με τις μεθόδους μπορείτε να βρείτε στο βιβλίο Guide to the Anonymization of Personal Health Information του Khaled el Emam.

Σχήμα 4.7

Εκτίμηση ρίσκου των δεδομένων της βάσης σε ανωνυμοποιημένη κατάσταση

Στην συγκεκριμένη περίπτωση οι δείκτες μας δείχνουν πως ο κίνδυνος είναι σε ελάχιστο επίπεδο.

i. 2η παραλλαγή :

Στην δεύτερη παραλλαγή (Πίνακας 4.2) έχουμε χρησιμοποιήσει πιο ελαστικές παραμέτρους.

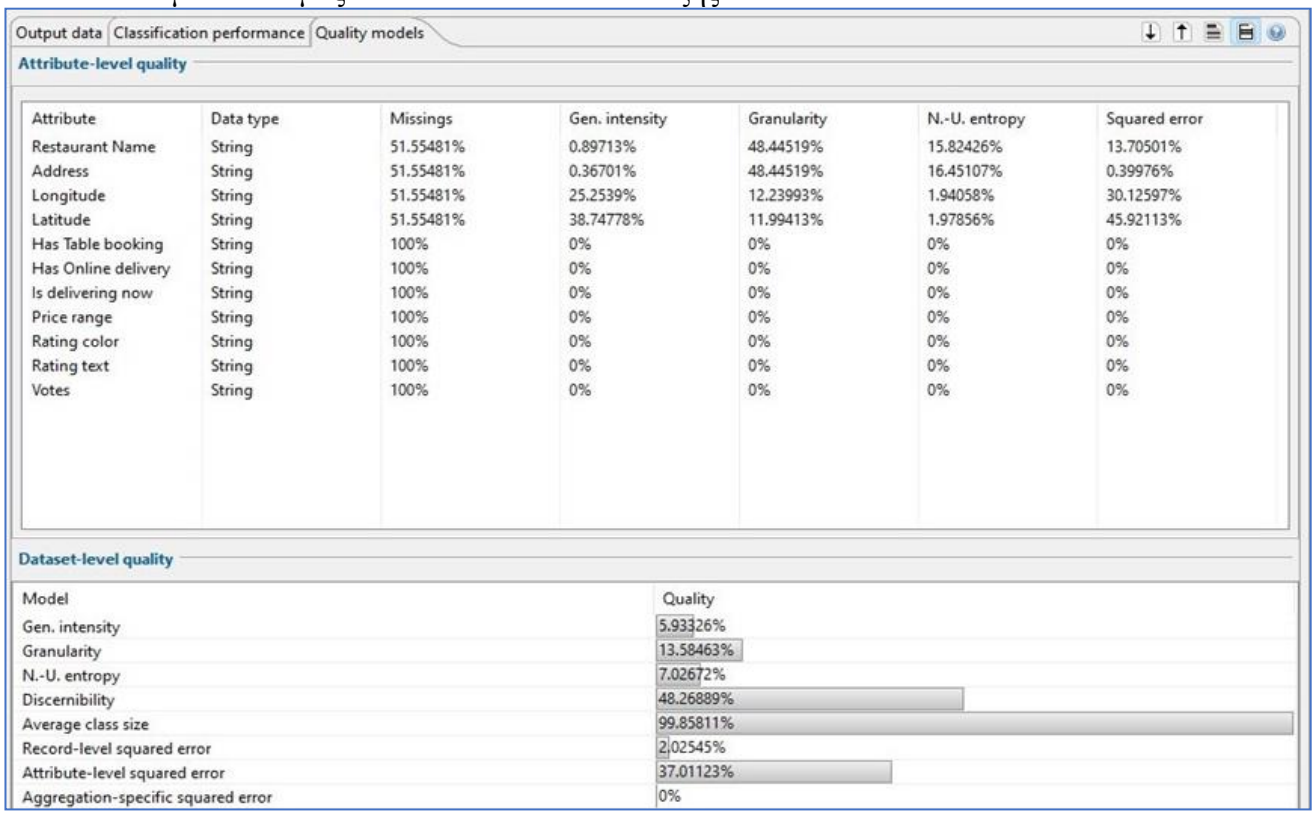
k-Anonymity	2
l-diversity Aggregate rating	2
l-diversity City	2
l-diversity Country Code	2
l-diversity Cuisines	2
l-diversity Currency	2
l-diversity Locality	2
l-diversity Locality Verbose	2
Suppression Limit	70%
Precomputation	0%
Measure	Loss

Aggregate function	Rank
Suppression/Generalization	50/50
Weights	
Restaurant	1
Address	0,7
Longitude	0,8
Latitude	0,8
Has Table	0,5
Has Online	0,5
Is delivering	0,5
Price range	0,5
Rating colour	0,5
Rating text	0,5
Votes	0,5

Πίνακας 4.2

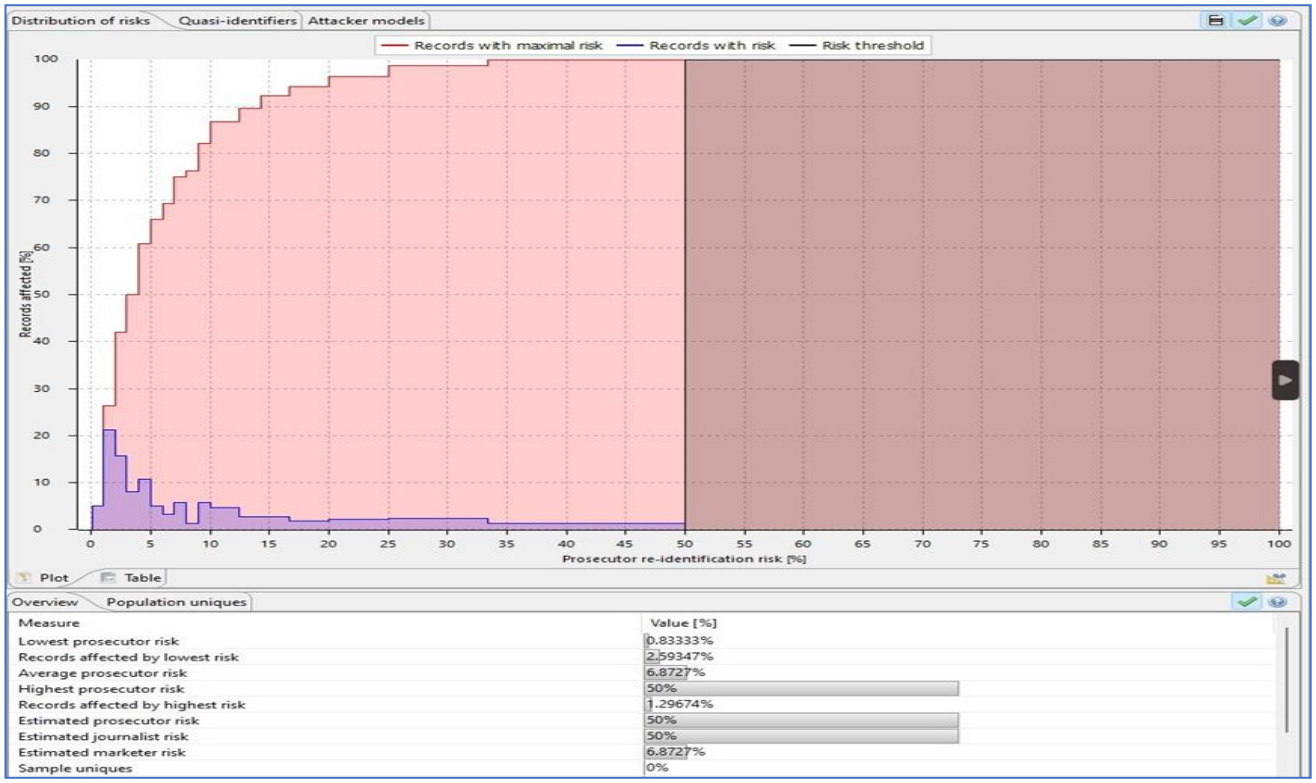
Παράμετροι ανωνυμοποίησης του "ARX" στην δεύτερη παραλλαγή

Τα αποτελέσματα που μας έδωσε το ARX είναι τα εξής :



Σχήμα 4.8

Κατάσταση του περιεχόμενου της βάσης με τις παραμέτρους εφαρμοσμένες

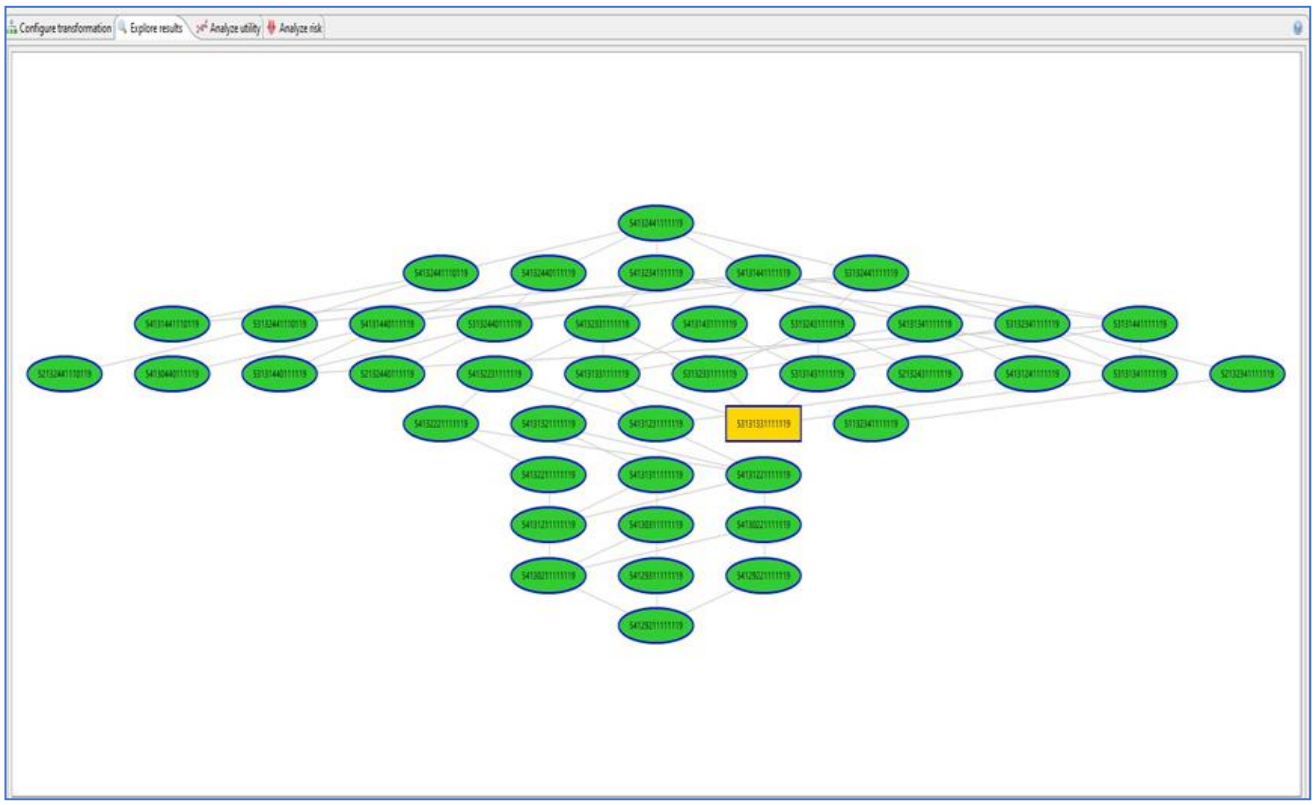


Σχήμα 4.9

Εκτίμηση ρίσκου των δεδομένων της βάσης σε ανωνυμοποιημένη κατάσταση

Σχήμα 4.10

Εκτίμηση ρίσκου των δεδομένων της βάσης σε ανωνυμοποιημένη κατάσταση



Σχήμα 4.11

Δέντρο» ανωνυμοποίησης

ii. 3η παραλλαγή :

Στην τρίτη παραλλαγή (πίνακας 4.3) έχουμε χρησιμοποιήσει ενδιάμεσες παραμέτρους.

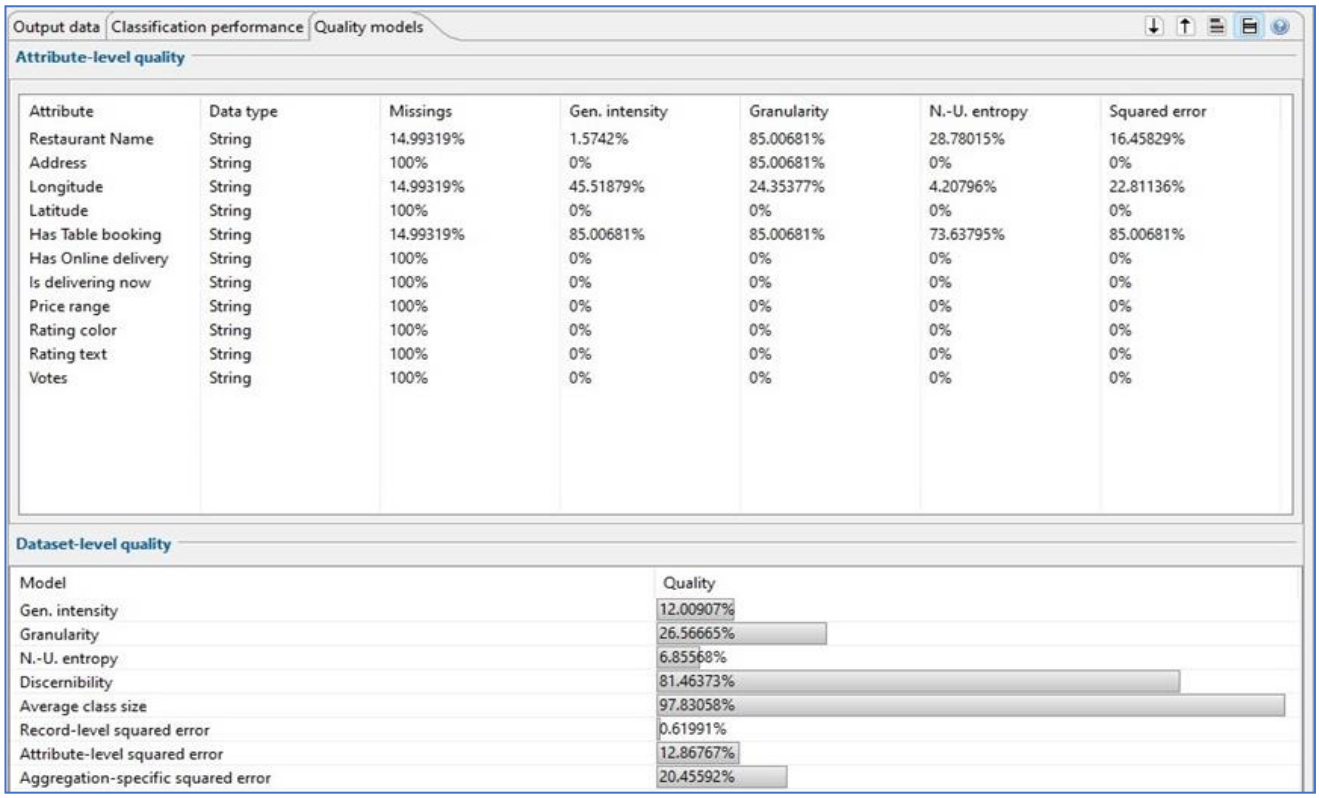
k-Anonymity	3
l-diversity Aggregate rating	3
l-diversity City	3
l-diversity Country Code	3
l-diversity Cuisines	3
l-diversity Currency	3
l-diversity Locality	3
l-diversity Locality Verbose	3
Suppression Limit	70%
Precomputation	0%
Measure	Loss
Aggregate function	Rank
Suppression/Generalization	50/50
Weights	
Restaurant	1
Address	0,7

Longitude	0,8
Latitude	0,8
Has Table	0,5
Has Online	0,5
Is delivering	0,5
Price range	0,5
Rating colour	0,5
Rating text	0,5
Votes	0,5

Πίνακας 4.3

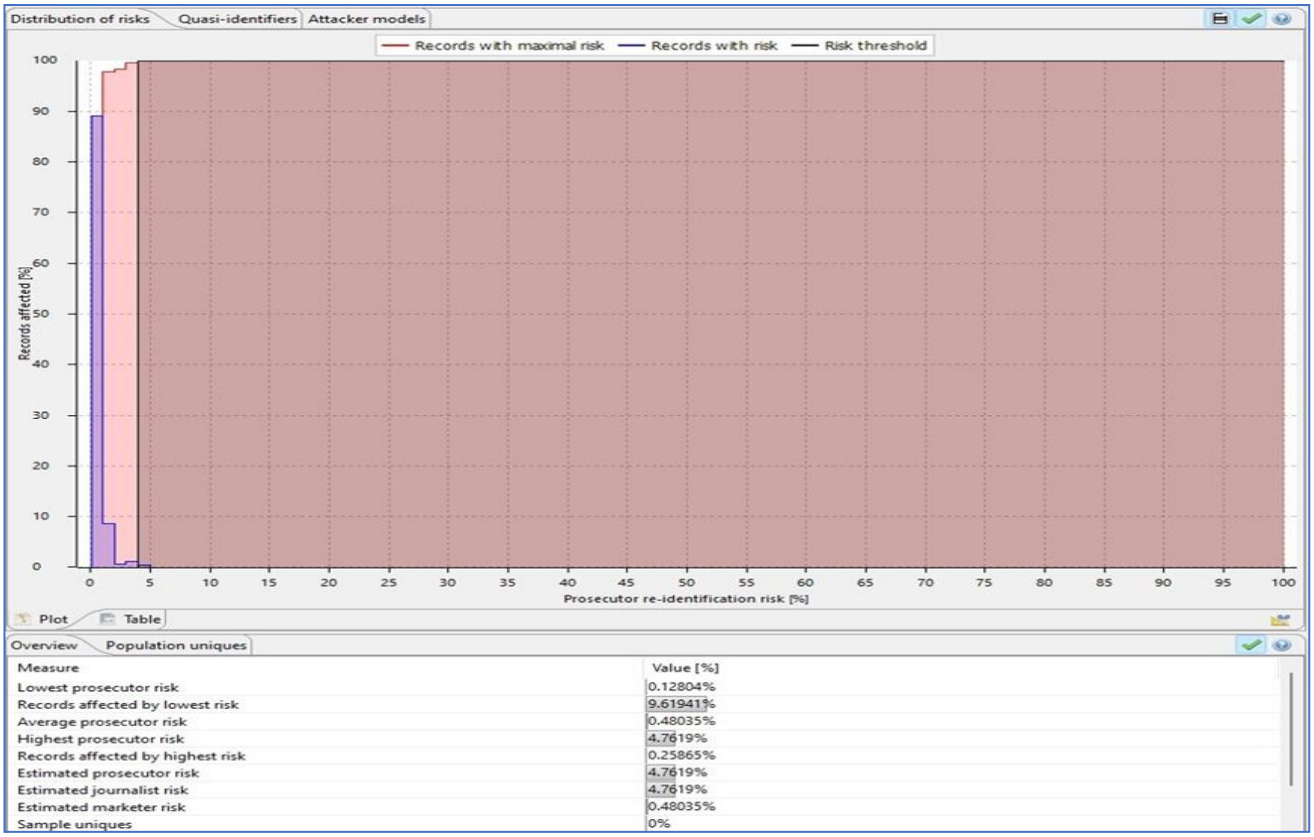
Παράμετροι ανωνυμοποίησης του "ARX" στην τρίτη παραλλαγή

Τα αποτελέσματα που μας έδωσε το ARX είναι τα εξής :



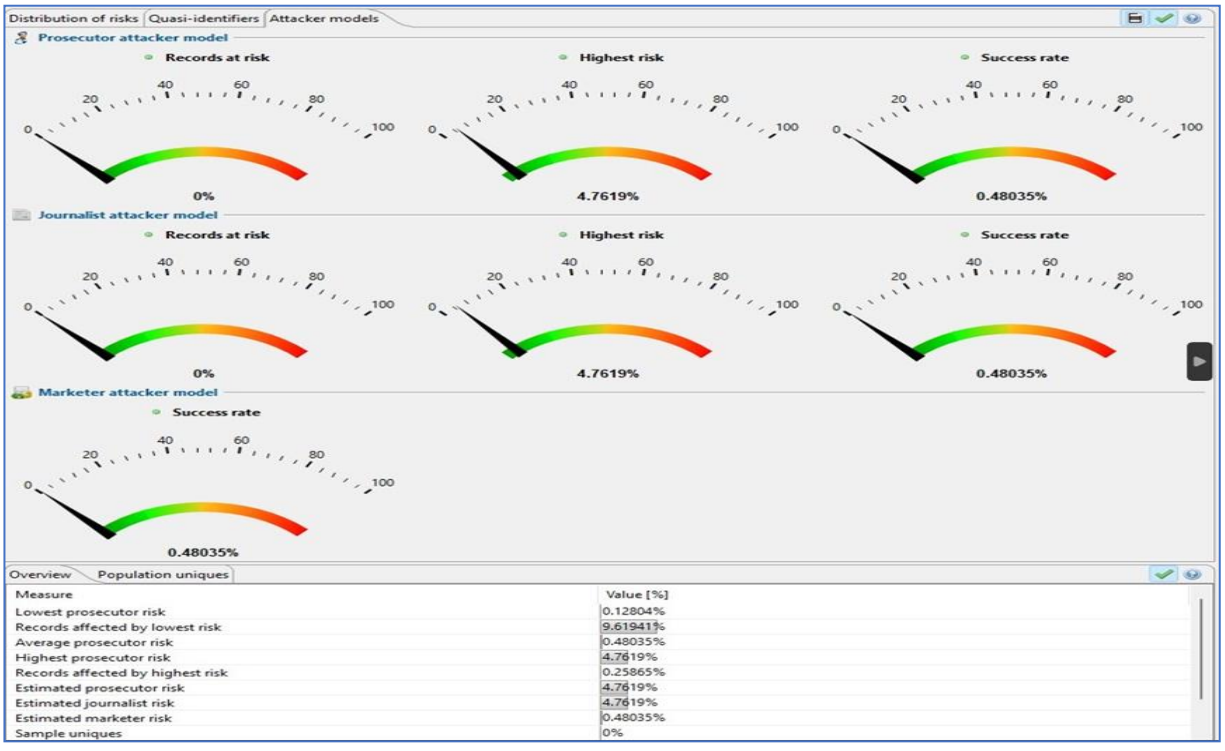
Σχήμα 4.12

Κατάσταση του περιεχόμενου της βάσης με τις παραμέτρους εφαρμοσμένες



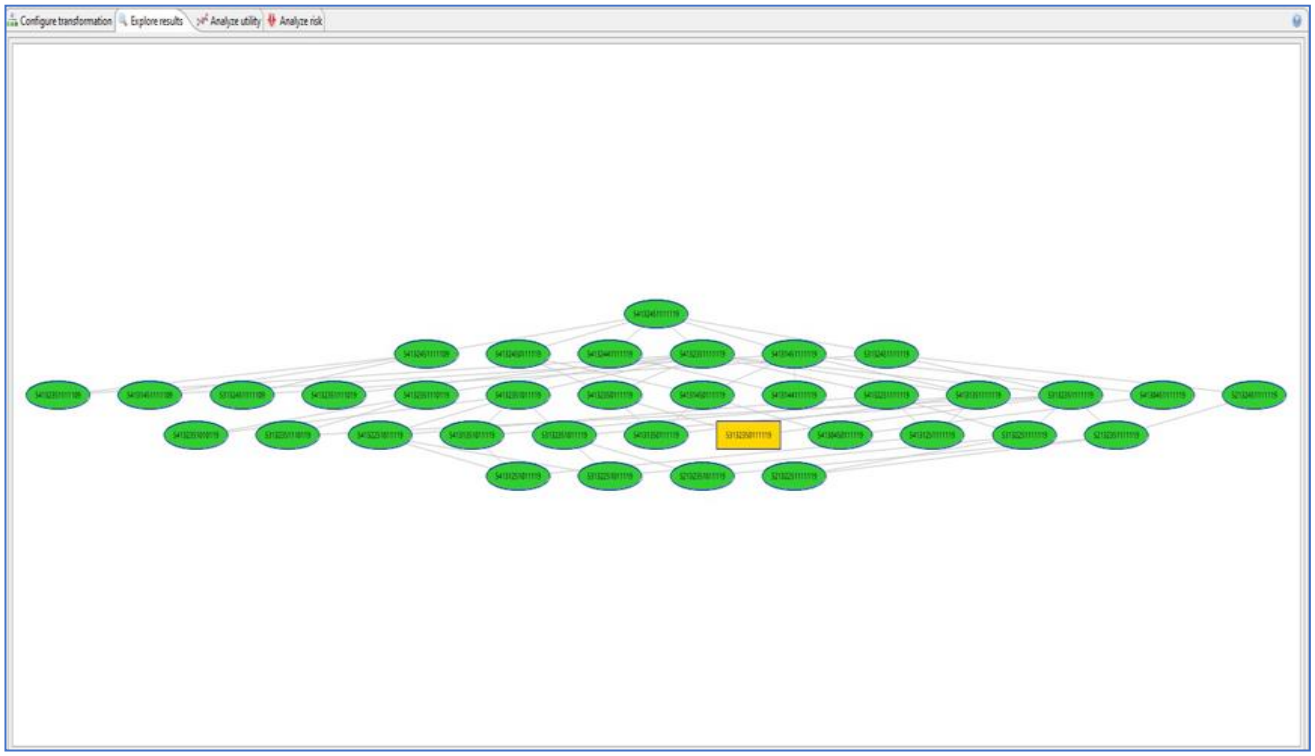
Σχήμα 4.13

Εκτίμηση ρίσκου των δεδομένων της βάσης σε ανωνυμοποιημένη κατάσταση



Σχήμα 4.14

Εκτίμηση ρίσκου των δεδομένων της βάσης σε ανωνυμοποιημένη κατάσταση



Σχήμα 4.15

«Δέντρο» ανωνυμοποίησης

4.1 Αποτελέσματα

Λαμβάνοντας υπόψιν την κατάσταση που βρίσκονται τα δεδομένα της βάσης πριν από οποιαδήποτε επεξεργασία με το ARX, καταλήγουμε πως, με την κατάλληλη παραμετροποίηση και επεξεργασία αυτής της βάσης, τα δεδομένα αυτά αλλάζουν. Από πλήρως εκτεθειμένα και αναγνώσιμα απέναντι στον καθένα, είναι εμφανές πως μετά την επεξεργασία μέσω του ARX, έχουμε να κάνουμε με δεδομένα προστατευμένα και σε πολύ μεγάλο βαθμό αποκρυμμένα.

Πιο συγκεκριμένα:

- Στην αρχικοποιημένη κατάσταση της βάσης έχουμε το 100% της βάσης εκτεθειμένη σε επιθέσεις κατήγορων, από επιθέσεις δημοσιογράφων όπως και από επιθέσεις marketing
- Στην πρώτη παραλλαγή της βάσης έχουμε το 8% της βάσης εκτεθειμένη σε επιθέσεις κατήγορων, 8% από επιθέσεις δημοσιογράφων όπως και 0,31% από επιθέσεις marketing
- Στην δεύτερη παραλλαγή της βάσης έχουμε το 50% της βάσης εκτεθειμένη σε επιθέσεις κατήγορων, 50% από επιθέσεις δημοσιογράφων όπως και 6,8% από επιθέσεις marketing
- Στην τρίτη παραλλαγή της βάσης έχουμε το 4,5% της βάσης εκτεθειμένη σε επιθέσεις κατήγορων, 4,5% από επιθέσεις δημοσιογράφων όπως και 0,4% από επιθέσεις marketing

5

Συμπεράσματα

Συμπερασματικά, το ζήτημα της ανωνυμίας στις υπηρεσίες που βασίζονται στην τοποθεσία έχει γίνει όλο και πιο σημαντικό στον σημερινό κόσμο που βασίζεται στα δεδομένα. Η χρήση δεδομένων τοποθεσίας σε διάφορες εφαρμογές έχει οδηγήσει στην πιθανότητα αποκάλυψης ευαίσθητων πληροφοριών, γεγονός που μπορεί να οδηγήσει σε σοβαρές ανησυχίες σχετικά με το απόρρητο. Αυτός είναι ο λόγος για τον οποίο έχουν αναπτυχθεί και μελετηθεί τεχνικές για την προστασία της ανωνυμίας στα δεδομένα τοποθεσίας στον τομέα της δημοσίευσης δεδομένων για τη διατήρηση του απορρήτου.

Έχουν προταθεί αρκετές τεχνικές για την αντιμετώπιση αυτού του προβλήματος, όπως η καταστολή δεδομένων, η διαταραχή δεδομένων και η γενίκευση δεδομένων. Μεταξύ αυτών, η γενίκευση δεδομένων θεωρείται η πιο αποτελεσματική λύση και έχει εφαρμοστεί σε αρκετούς αλγόριθμους διατήρησης της ιδιωτικής ζωής. Αυτοί οι αλγόριθμοι περιλαμβάνουν τους CryptDB, ARX και NN-Cloak.

Επιπλέον, έχει γίνει εκτενής έρευνα σχετικά με τη χρήση της γενίκευσης δεδομένων στην ανωνυμοποίηση δεδομένων τροχιάς, και αυτό οδήγησε στην ανάπτυξη πολλών τεχνικών απόκρυψης, όπως το Clique Cloak, το Center Cloak, το Casper Technique, το Interval Cloak και το Hilbert Cloak. Όλες αυτές οι τεχνικές στοχεύουν στην παροχή υψηλού επιπέδου ανωνυμίας διατηρώντας παράλληλα τη χρησιμότητα των δεδομένων.

Συνολικά, η έρευνα σχετικά με την ανωνυμία στις υπηρεσίες που βασίζονται στην τοποθεσία έδειξε ότι είναι δυνατή η προστασία ευαίσθητων πληροφοριών, ενώ παράλληλα επιτρέπεται χρήσιμη ανάλυση. Ενώ εξακολουθούν να υπάρχουν προκλήσεις, όπως η πιθανότητα επίθεσης και η δυσκολία εξισορρόπησης της ιδιωτικής ζωής και της χρησιμότητας, σημειώνεται πρόοδος στην εξεύρεση αποτελεσματικών λύσεων σε αυτά τα ζητήματα. Ως εκ τούτου, αναμένεται ότι θα συνεχιστούν περαιτέρω πρόοδοι στον τομέα αυτό τα επόμενα χρόνια.

Στο μέλλον, οι ερευνητές και οι προγραμματιστές θα μπορούσαν να συνεχίσουν να εργάζονται για τη βελτίωση και την επέκταση των δυνατοτήτων των εργαλείων δημοσίευσης δεδομένων που διατηρούν το απόρρητο, όπως το ARX. Ένα πιθανό έργο θα μπορούσε να είναι η διερεύνηση νέων τεχνικών ανωνυμοποίησης για σημειακά χωρικά δεδομένα που εξισορροπούν καλύτερα το απόρρητο και τη χρησιμότητα. Ένα άλλο έργο θα μπορούσε να είναι η ενίσχυση του ARX με πρόσθετα χαρακτηριστικά, όπως βελτιωμένους μηχανισμούς ασφαλείας ή ενσωμάτωση με άλλες τεχνολογίες διατήρησης της ιδιωτικής ζωής. Επιπλέον, η διερεύνηση του τρόπου εφαρμογής του ARX και παρόμοιων εργαλείων σε άλλους τύπους ευαίσθητων δεδομένων, όπως προσωπικές πληροφορίες υγείας ή οικονομικά αρχεία, θα μπορούσε να είναι ένας άλλος τομέας εστίασης. Καθώς η ζήτηση για απόρρητο και ασφάλεια στην κοινή χρήση δεδομένων συνεχίζει να αυξάνεται, πιθανότατα θα υπάρξουν πολλές συναρπαστικές ευκαιρίες για νέα έρευνα και ανάπτυξη σε αυτόν τον τομέα.

Βιβλιογραφία

1. Li, N., Li, T., Venkatesh, S., & Li, X. (2007). NN-Cloak: An Effective and Efficient K-Anonymization Algorithm for Trajectory Data. In Proceedings of the 2007 ACM SIGMOD international conference on Management of data (pp. 961-972).

2. Kalnis, P., Zhang, Y., Papadias, D., & Tao, Y. (2007). Anonymizing spatial trajectory data. In Proceedings of the 2007 ACM SIGMOD international conference on Management of data (pp. 597-608).

3. Hund, M. E., & Kifer, D. (2012). A general framework for privacy-preserving data publishing. ACM Transactions on Database Systems (TODS), 37(4), 30.

4. Papadimitriou, S., Samatas, M., & Economou, A. (2012). Cloaking location-based services: A survey of techniques and attacks. Pervasive and Mobile Computing, 8(3), 447-468.

5. Sun, Y., Ling, C. X., & Sun, J. (2013). Clique cloak: An effective and efficient location privacy protection approach in LBSs. Pervasive and Mobile Computing, 9(6), 1058-1072.

6. Hernández-Orallo, J., & Jonsson, B. (2010). CryptDB: Protecting Confidentiality with Encrypted Query Processing. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data (pp. 85-96).

7. Huang, Z., Qu, Y., He, T., & Tao, Y. (2010). Towards privacy-aware mobile location-based services. ACM Transactions on Information and System Security (TISSEC), 13(2), 11.

8. "K-Anonymity: A Model for Protecting Privacy." Ramakrishnan, Raghu and Livny, Moni. ACM Transactions on Database Systems (TODS), 31(1), 2006.

9. "Differential Privacy." Cynthia Dwork. Invited paper, 33rd International Colloquium on Automata, Languages and Programming (ICALP), 2006.

10. "Anonymization of Spatial Trajectory Data: A Systematic Review." Liu, Yun and Li, Zhenhua and Cui, Wei and Hu, Jie and Li, Xiaoming. ACM Transactions on Intelligent Systems and Technology (TIST), 8(3), 2017.

11. "Cloaking Location-Based Services: A Survey of Techniques and Attacks." Guan, Ming and Zhang, Weixin and Zhang, Dongxiao and Su, Wei. IEEE Communications Surveys & Tutorials, 15(4), 2013.

12. "Anonymizing Spatial Trajectory Data." Mokbel, Mohamed F. and Gao, Xiaofang and Shekhar, Shashi. IEEE Transactions on Knowledge and Data Engineering, 22(6), 2010.

13. "Anonymizing Spatial Trajectory Data: A Survey." Cheng, Jian and Lu, Jian and Dai, Lianxin and Wu, Weili. ACM Transactions on Spatial Algorithms and Systems (TSAS), 1(1), 2015.

14. "A General Framework for Privacy-Preserving Data Publishing." Aggarwal, Charu and Yu, Philip S. and Honavar, Vasant. ACM Transactions on Knowledge Discovery from Data (TKDD), 5(1), 2011.

15. "Data Anonymization: Techniques and Applications." Fan, Wei and Kamiran, Faisal and Samarati, Pierangela. Data Mining and Knowledge Discovery Handbook, 2nd edition, pages 191-223. Springer, 2016.